UNIVERSIDAD DE COSTA RICA
FACULTAD DE CIENCIAS ECONÓMICAS
ESCUELA DE ESTADÍSTICA

Tema

APLICACIONES PRÁCTICAS DE LA ENCUESTA DE HOGARES DE PROPÓSITOS MÚLTIPLES DE JULIO DE 2004

Seminario de graduación para optar por el grado de Licenciatura en Estadística

YADIRA MARÍA ALVARADO SALAS
ROSE MARY FONSECA GONZÁLEZ
TATIANA POCHET VINDAS

Ciudad Universitaria Rodrigo Facio
SAN JOSÉ, COSTA RICA
AGOSTO 2008

RESERVADOS TODOS LOS DERECHOS

Yadira María Alvarado Salas

Rose Mary Fonseca Rodríguez

Tatiana Pochet Vindas

DEDICATORIA

Por lo que soy y por cuanto tengo, le dedico este estudio a mi Señor Dios, Quien en su Trinidad insondable, escribe sobre renglones torcidos.

A mis padres, ancianos hoy, que han esperado con paciencia ver el día en que su hija concluyera el estudio.

A mis hermanas y hermano, que siempre vieron en mí, la capacidad que Dios me había dado para llevar a cabo con éxito, un plan iniciado treinta y cinco años atrás.

A mis sobrinos, que compartiendo sus experiencias de estudiantes y sus metas cumplidas, reanimaban en mí el ansia de proseguir.

A mis amigas y amigos que alentaron mi frágil deseo de concluir proyectos iniciados.

YADIRA

A Dios por el camino recorrido y por darme la fortaleza necesaria para salir siempre adelante, pese a las dificultades.

A la memoria de mi Mami y mi Papi (María Luisa y Juan), que me transmitieron ese espíritu de seguir siempre adelante, tras mis sueños e ilusiones.

En especial a Rebeca y Alejandro, que a su corta edad pudieron comprender el significado de que yo culminara este trabajo, que supieron entender mis ausencias, y aceptaron sacrificarse en tantas horas que les pertenecían y que contribuyen a diario a construirme como persona.

Además, al incondicional apoyo brindado por quienes con su cariño y comprensión lograron impulsarme hacia objetivos mayores: mi familia (mis hermanas y hermanos).

ROSE MARY

Con todo mi cariño dedico este proyecto a Dios, por ser mi fuente de inspiración y por guiar mis pasos.

A mi familia, por todo el tiempo y sacrificio empleado en cultivar mi educación, por su paciencia y apoyo incondicional.

TATIANA

AGRADECIMIENTO

Nuestro agradecimiento al Comité Asesor: Olman Ramírez Moreira, como director del Seminario y a los lectores Pilar Ramos Vargas y Rafael Segura Carmona, quienes compartieron reiteradamente, su valioso tiempo y sus conocimientos en la realización de estas investigaciones.

Al Instituto Nacional de Estadística y Censos (INEC), por toda la colaboración en el suministro de las bases de datos, y por brindar la seguridad de contar con todo el apoyo de la institución.

De igual manera un agradecimiento especial a Giselle Argüello Venegas, experta muestrista del INEC, por su desinteresada contribución en los temas de errores. Sus comentarios tan oportunos, favorecieron grandemente estos contenidos.

Un especial agradecimiento para Gilbert Brenes Camacho por sus enseñanzas y colaboración esclarecedora, siempre oportuna, en la técnica y en el análisis estadístico de los modelos log-lineales.

También agradecemos a todas aquellas otras personas que de alguna manera nos aportaron su granito de arena para finalizar este proyecto.

RESUMEN

La Encuesta de Hogares de Propósitos Múltiples, ejecutada anualmente por el Instituto Nacional de Estadística y Censos (INEC), provee una riqueza de información estadística a nivel nacional digna de utilizar. Efectivamente, los resultados de la encuesta de julio 2004 y el empate de las encuestas 2004-2005 se aprovecharon para desarrollar cuatro áreas de investigación de interés técnico y social.

ERRORES MUESTRALES

El método de medición tradicional o clásico, utiliza aproximaciones de series de Taylor para medir el error muestral. Este procedimiento requiere el cumplimiento del supuesto de normalidad para calcular ese error y los respectivos intervalos de confianza. Este supuesto es de difícil cumplimiento, sobre todo en variables nominales. El objetivo de esta sección, es obtener una medición de ese error, mediante una técnica alternativa como lo es el procedimiento Bootstrap, para validar la precisión de la estimación aplicada actualmente por el INEC.

Los resultados, sobre 28 indicadores analizados, demostraron que el error estándar medido a través del procedimiento Bootstrap, resultaron inferiores en más ocasiones, que el obtenido a través de aproximaciones de series de Taylor. Al definir un rango de tolerancia del -5 al 5%, mediante un coeficiente diferencial, se obtuvo que menos de la cuarta parte de los indicadores registraron diferencias fuera de ese rango, lo que permite validar la precisión de las estimaciones realizadas por el INEC.

ERRORES NO MUESTRALES

La ejecución de todo el proceso de una encuesta debe seguir normas de control de calidad ejercidos en el cuestionario, trabajo de campo, crítica, codificación, digitación, limpieza de datos y publicación. No obstante, todo ese esfuerzo no logra anular la

presencia de los errores no muestrales. El empate de las encuestas de julio 2004 y 2005 proporcionó la plataforma electrónica para detectar posibles errores, en las respuestas dadas a 12 variables susceptibles de medirlos, logrando totalizarlos. De igual manera, esta investigación logró determinar el impacto del tipo de informante en la generación de errores, utilizando una regresión logística.

Solo un 25% de los hogares coincidentes no tuvieron errores, mientras que la mitad de ellos registró uno o dos; el otro 25% obtuvo tres o más errores. Las dos variables con mayor cantidad de errores, fueron *Nivel educativo* y *Lugar de nacimiento*. El modelo logístico obtuvo el mejor ajuste con tres variables: *Tipo de informante*, *Tamaño del hogar* y *Nivel de educación*. Este modelo logró determinar que la probabilidad de cometer al menos un error es más alta cuando las respuestas se obtienen de un informante diferente al del año anterior. Análogamente, conforme aumenta el nivel de escolaridad formal del informante, como también si aumenta el número de miembros del hogar, se incrementa el riesgo de obtener al menos un error.

LA CONFORMACIÓN DE PAREJA: SU RELACIÓN CON LA EDUCACIÓN FORMAL

La palabra 'homogamia' deriva del latín, 'homo' que significa igual y 'gamia' se refiere a la unión en matrimonio. En tal sentido, en este apartado del documento se pretendió explicar la relación del nivel educativo en la conformación de las parejas y la reproducción de la estratificación social en la estructura familiar. Para ello se utilizó la técnica para el análisis de datos categóricos conocida como modelos log-lineales, dado que las variables de análisis son nominales o categóricas. Se probaron cinco modelos especificados por el autor Park (1991), a saber: modelo simétrico, modelo de la diferencia en la diagonal principal, modelo diagonal simétrico, modelo asimétrico y modelo de hipergamia – hipogamia.

El modelo que resultó explicativo para la situación de la conformación de las parejas costarricenses, una vez que se verificó que el modelo de independencia no se cumplía, fue el modelo de hipergamia-hipogamia o modelo de asociación no simétrica en la diagonal diferente con valores para BIC=2803,44 y LR^2/g.l=451,41, resultando significativos, situación que señala la hipótesis en cuanto a que el emparejamiento de la parejas se realiza siguiendo un patrón diferente en el nivel educativo de los miembros de la pareja, indistintamente de quién tenga mayor o menor nivel de instrucción (los niveles de homogamia, hipergamia o hipogamia son diferentes). Ante este escenario se examinó la diferencia en los niveles de hipergamia y de hipogamia, separadamente, los datos mostraron evidencia de favorecer el modelo de hipogamia (vp= 0,01; con el coeficiente negativo), que se refiere a que las mujeres forman pareja con hombres que tienen menor nivel educativo formal.

Respecto, de la reproducción de la estratificación social en la estructura familiar, los resultados reportaron que la bondad de ajuste del modelo original sin la variable situación de pobreza comparada con el modelo original con la variable situación de pobreza presenta una mejor bondad de ajuste, razón por la cual se definió un patrón diferente para los pobres y no pobres; señalando un mejor ajuste (BIC=5.95, LR/g.l=317.69).

LA INCORPORACIÓN DE LA MUJER UNIVERSITARIA A LA POBLACIÓN ECONÓMICAMENTE ACTIVA

La iniciativa de este proyecto surge como respuesta al cambio que se ha estado presentando en nuestro país con respecto a la mujer universitaria y su incorporación a la población económicamente activa (PEA). Para poder determinar los factores e importancia de este proyecto, se realizó un análisis sobre la situación actual, siendo fuente esencial para incluir las variables en estudio.

La técnica estadística que se empleó para la obtención de resultados fue la de regresión logística, por permitirnos utilizar a la variable respuesta como dicotómica, en la

cual muestra que las mujeres universitarias que pertenecen a la población económicamente activa representan el 65.45% del total de la base de datos a utilizar (680 observaciones), mientras que el 34.55% no están incorporadas a la PEA. Por otro lado el ajuste del modelo logístico resultó muy aceptable y satisfactorio, al igual que la submuestra que se utilizó para la validación de dicho modelo. De manera que las variables independientes incluidas dentro de este modelo reducido y que resultaron significativas son: la edad en años cumplidos, la cantidad de años de escolaridad, el logaritmo natural del ingreso mensual personal, cantidad de activos en el hogar, el logaritmo natural del ingreso mensual total del hogar y la cantidad de miembros en el hogar. Por lo que las características del entorno (hogar) y las personas afectan si la mujer universitaria se incorpora a la población económicamente activa.

El análisis de la bondad del ajuste utilizado es el adecuado y no existe evidencia estadística para desechar el modelo. Este modelo permite la clasificación correcta del 88.64% de los casos, valor bastante alto para el tipo de información que se analiza. La sensitividad es bastante elevada siendo un 94.41%, y la especificidad moderadamente alta con un 77.72%.

ÍNDICE

Página

Página

Página

INTRODUCCIÓN

Este documento es resultado del trabajo final de graduación para optar por el grado de Licenciatura en Estadística, con varios énfasis, bajo la modalidad de Seminario de Graduación. Para tal efecto, se escogió trabajar en varios temas desarrollados por cada participante en este seminario, cuya principal fuente de información es la Encuesta de Hogares de Propósitos Múltiples (EHPM) de julio 2004 [1/], elaborada por el Instituto Nacional de Estadística y Censos (INEC).

Los temas de investigación escogidos para este Seminario, son los siguientes:

- Errores de muestreo.
- Errores no muestrales.
- La conformación de pareja: su relación con la educación formal.
- La incorporación de la mujer universitaria a la población económicamente activa (PEA).

Los dos primeros temas fueron desarrollados por Yadira María Alvarado Salas; el tema de la conformación de pareja y su relación con la educación formal fue abordado por Rose Mary Fonseca González, y por último, la incorporación de la mujer universitaria a la población económicamente activa (PEA), le correspondió a Tatiana Pochet Vindas.

Los datos fueron proporcionados por el INEC, quien cedió las bases de datos de la EHPM de 2004 y el empate 2004-2005 utilizado solamente en el capítulo de *Errores no muestrales*. No obstante, en el proceso de la investigación, las estudiantes requirieron hacer arduos reprocesamientos de estas bases, con el fin de definir y obtener las variables necesarias para desarrollar sus temas.

[1/] Para el capítulo *Errores no muestrales* se utilizó además, el empate de las EHPM 2004-2005 (véase página 76).

La historia de las encuestas de hogares en Costa Rica se remonta a 1966; sin embargo, el actual programa (EHPM) inicia en 1987, de manera anual (a julio de cada año), ejecutado bajo la tutela directa de la Dirección General de Estadística y Censos del Ministerio de Economía, Industria y Comercio, hoy INEC.

La EHPM constituye una valiosa fuente de información sobre diversos aspectos relativos a los hogares y a las personas que los conforman. Tiene especial importancia el Módulo de Empleo, que investiga aspectos referentes a la fuerza de trabajo y sus características: empleo, desempleo, subempleo e ingresos. También es útil para investigar otros temas complementarios acerca de las características demográficas y socioeconómicas de los hogares y de sus residentes [INEC, 1999:1]. De esta manera, la EHPM brinda información nacional acerca de un gran espectro de variables, y además, posibilita la incorporación de nuevos módulos de investigación en áreas específicas, según sean demandadas por diferentes usuarios.

En general, el objetivo de las encuestas es brindar estimaciones de ciertas características para una determinada población en un momento dado. Específicamente, para la EHPM de Costa Rica, el INEC menciona tres objetivos:

a. Mantener un flujo continuo de estadísticas relacionadas con la fuerza de trabajo, el empleo, el desempleo, el subempleo y los ingresos, así como de otras variables socioeconómicas necesarias para el establecimiento de políticas y la formulación de planes orientados al desarrollo económico y social del país, y para la evaluación de sus efectos.

b. Proveer información periódica, sistemática y oportuna en los períodos intercensales, referente a las variables mencionadas.

c. Servir de fuente de información a instituciones gubernamentales, universitarias, o de investigación, interesadas en temas relativos a la población y el empleo, y en otros temas que se introduzcan periódicamente en la encuesta.

Cada tema de investigación en este Seminario es analizado en forma separada por las alumnas, de tal manera que se define un problema por investigar y los objetivos –general y específicos- propios de cada uno. La presentación de la información, por lo tanto, se expone en cuatro capítulos, uno por cada tema, en apartados correspondientes a la siguiente estructura:

- Introducción, que contiene una justificación, la definición del problema y los objetivos planteados.
- El marco o referente teórico, que proporciona el sustento que la teoría aporta a la investigación.
- La metodología empleada para alcanzar los objetivos señalados.
- La presentación de los resultados obtenidos.
- Las conclusiones pertinentes.

Además, la bibliografía y los anexos correspondientes se presentan en forma independiente.

En síntesis, el presente Seminario de Graduación pretende aprovechar de manera integral, la riqueza de información que proporciona la Encuesta de Hogares de Propósitos Múltiples de julio 2004 en los cuatro campos mencionados.

CAPÍTULO 1
ERRORES DE MUESTREO

1.1 INTRODUCCIÓN

La encuesta por muestreo es un método para recolectar información sobre una población humana, obtenida de una fracción de la población seleccionada [Lisinger y Warwick, 1985:15-16]. Este proceso de seleccionar solamente una parte de la población para realizar inferencias; es decir, para *"... sacar conclusiones de un gran número de acontecimientos fundándose en las observaciones de parte de los mismos"* [Siegel 1976:20], conlleva a un margen de error asociado que se puede dividir en dos campos: errores de muestreo y errores no muestrales (que no provienen del diseño muestral empleado). Los primeros dependen de la variabilidad de las características investigadas y del diseño muestral utilizado para estimarlas, mientras que los segundos provienen de la condición inherente de querer medir una característica, independientemente de si se trata de una muestra o un censo.

La Encuesta de Hogares de Propósitos Múltiples (EHPM), ejecutada por el Instituto Nacional de Estadística y Censos (INEC), posee un diseño muestral complejo [2/], que le permite brindar información acerca de una gran gama de variables en el ámbito nacional. Análogamente, las estimaciones de los errores muestrales son complejas, esto quiere decir, que en cada etapa del diseño, se requiere unir combinaciones de cálculos aritméticos y probabilísticos en fórmulas específicas.

En la actualidad, se han desarrollado programas de cómputo que facilitan no solo el procesamiento de los datos estadísticos, sino la medición del error muestral a través de dos métodos generales: el método tradicional o clásico, ejecutado a través de una aproximación

---

[2/] Se considera un diseño muestral complejo a aquel que involucra dos o más etapas de selección.

de series de Taylor y, el método de las replicaciones repetidas, en sus diferentes modalidades (grupos aleatorios, jacknife, bootstrap, entre otros). Dado que son procedimientos diferentes, conducen a estimaciones distintas, aunque la investigación empírica (Kish y Frankel 1970; Kish y Frankel 1974, mencionado en Lepkowski y Bowles [1996:3]) ha demostrado que para muchos estadísticos, las diferencias en los métodos son pequeñas.

El desconocimiento de las distribuciones de las variables e indicadores objeto de estudio, o el nivel no métrico de los datos, hace recomendable acceder a algún método alternativo libre del supuesto de algún tipo específico de distribución. El método clásico o tradicional, que utiliza aproximaciones de series de Taylor, requiere del supuesto de normalidad para el cálculo del error muestral y de los respectivos intervalos de confianza. Sin embargo, cuando los datos no concuerdan con este tipo de distribución, las estimaciones podrían estar sesgadas. Cuesta y Herrero [1999:1] admiten que *"... no son pocas las voces que ponen en duda que en las ciencias sociales el cumplimiento de estas asunciones distribucionales sea tan frecuente como parecería deducirse del generalizado uso de las técnicas paramétricas. Hinkle y Winstead justifican la persistencia en la utilización de la distribución normal en las ventajas computacionales que proporcionaba antes de la irrupción de los ordenadores"*.

En el caso de la EHPM, el tamaño de muestra empleado no deja dudas de que queda cubierta por el Teorema del Límite Central. No obstante, realizar un ejercicio alternativo sería beneficioso para validar sus resultados.

La EHPM, en su publicación ordinaria, presenta un capítulo completo acerca de los errores de muestreo para las variables más importantes, obtenidos por el método de estimación clásico (aproximación series de Taylor), como se muestra en el anexo A. No obstante, estas variables seleccionadas para ser publicadas, son categóricas, lo que no permite visualizar adecuadamente el supuesto de normalidad. Por otra parte, en el INEC no existe referencia a alguna medición de estos errores a través de métodos alternativos libres

del supuesto de normalidad, que permita validar los resultados obtenidos, por lo cual, resulta pertinente un ejercicio que refleje las diferencias entre dos tipos de procedimientos.

En la Escuela de Estadística de la Universidad de Costa Rica, se han realizado algunos ejercicios académicos sobre la técnica Bootstrap como mecanismo de medición de los errores de muestreo en la Encuesta de Hogares; sin embargo, la verificación de la magnitud de las diferencias encontradas entre esta técnica y el método clásico, nunca ha sido expuesta.

El problema planteado en este estudio, por lo tanto, es: ¿existen diferencias significativas en la medición del error muestral, empleando dos métodos diferentes (aproximación de Series de Taylor y replicaciones repetidas), para distintos indicadores? De esta manera, los resultados de la investigación, podrían ofrecer estimaciones alternativas sobre el tema; asimismo coadyuvar a una mejor y mayor divulgación del significado de los errores de muestreo a los que se expone una encuesta; determinar sus magnitudes, y sensibilizar sobre la presencia de estos errores, lo que refuerza la idea de que las estimaciones puntuales no son suficientes, y que por ello, se requiere de los intervalos de confianza; aspecto que propicia un ambiente de discusión.

Para dar respuesta a esa pregunta, se plantean diferentes objetivos aplicados sobre un conjunto de variables extraídas de la EHPM de julio de 2004.

Objetivo general

Determinar las diferencias en la estimación del error muestral, mediante los métodos clásico y de replicaciones repetidas, para validar la precisión de la estimación aplicada, actualmente, por el INEC.

Objetivos específicos

- Seleccionar un conjunto de indicadores de diferente construcción: totales, tasas, porcentajes, promedios y otros, para que sirvan de indicadores globales de interés general.
- Calcular los errores muestrales, bajo el método de aproximación Series de Taylor, para ese conjunto de indicadores, cuyos valores no necesariamente aparecen en la publicación de la Encuesta de Hogares.
- Aplicar un método alternativo para el cálculo de errores de muestreo de ese conjunto de indicadores, procedente de las replicaciones repetidas, para comparar ambos métodos.
- Contrastar los resultados obtenidos con ambos métodos, mediante un coeficiente diferencial que permita relativizar y comparar las diferencias entre los distintos indicadores.

1.2 REFERENTE TEÓRICO

1.2.1 Estadística Paramétrica y No Paramétrica

" *En el desarrollo de los métodos estadísticos modernos, las primeras técnicas de inferencia que aparecieron fueron las que hicieron buen número de suposiciones acerca de la naturaleza de la población de la que se obtuvieron los puntajes. Puesto que los valores de población son "parámetros", estas técnicas estadísticas son llamadas paramétricas. ... Tales técnicas nos conducen a conclusiones que tienen limitaciones, por ejemplo: "Si las suposiciones con respecto a la forma de la población son válidas, entonces podemos concluir que…"* " [Siegel, 1976:21].

En la primera mitad del siglo pasado, se presenció "*... el desarrollo de gran número de técnicas de inferencia que no hacen suposiciones numerosas ni severas acerca de los parámetros. Estas nuevas "distribuciones-libres" o técnicas no paramétricas, permiten llegar a conclusiones a las que hay que hacer menos reservas. Cuando usamos una de ellas, podemos decir que "independientemente de la forma de la(s) población(es), podemos sacar su conclusión que…"* " [Siegel, op cit]. Específicamente, fue en las ciencias sociales donde se necesitaron estos métodos para analizar sus variables, pues los tradicionales fueron ineficaces para analizar las de estos campos, que en su mayoría, no son métricas ni tienen una distribución conocida [Mata, 1985:xi].

El nivel de medición de las variables marca el tipo de estadística o técnica por emplear. Las variables se pueden clasificar en una de las escalas siguientes [Siegel, 1976:42-49]:

- **Nominal:** La medición se da en un nivel elemental, cuando los números u otros símbolos se usan para la clasificación de objetos, personas o características. Ejemplos: sexo, estado civil, condición de actividad, ocupación.

- **Ordinal:** Sus valores se pueden clasificar en categorías ordenadas, de tal modo que surja un rango ordenado completo. Ejemplos: nivel socioeconómico, clase social, lugar en la clase.
- **De intervalo:** Esta escala tiene todas las características de una escala ordinal y además se conoce la distancia entre dos números cualesquiera; es decir, se conoce la magnitud de los intervalos (distancias) entre todos los objetos de la escala. En esta, el punto cero y la unidad de medida son arbitrarios. Ejemplo: temperatura.
- **De razón o proporción:** Tiene todas las características de una escala de intervalo y además, tiene un punto cero real en su origen; es decir, el cero representa la ausencia de la característica que se evalúa. Ejemplos: peso, estatura.

Siegel [1976:50] afirma que los datos medidos por escalas nominales u ordinales deben analizarse por métodos no paramétricos, mientras que los datos medidos con escalas de intervalo o de razón deben analizarse por métodos paramétricos, si los supuestos estadísticos paramétricos son sostenibles.

Específicamente, en la medición del empleo es frecuente utilizar variables nominales. Algunas de ellas son o se transforman en variables dicotómicas; en otras palabras, estas variables poseen solo dos posibles valores, por ejemplo hombre y mujer; ocupado y no ocupado, pobre y no pobre. Si estas variables deben señalar la ausencia o presencia de una característica, se les asigna valores 0 y 1, y se les denomina variable binaria dicotómica [Gujarati, 2004:561] o variable dummy. Los códigos o "*... símbolos que designan a los diferentes grupos en una escala nominal pueden intercambiarse sin alterar la información esencial de la escala; debido a esto, las únicas estadísticas descriptivas admisibles son aquellas que no se alteran en este proceso: el modo, la frecuencia, el conteo, etc. En ciertas condiciones, podemos probar las hipótesis de la distribución de casos en las categorías usando la prueba estadística no paramétrica,* χ^2 *, o prueba basada en la fórmula binomial; que son apropiadas para datos nominales, pues revelan la frecuencia en las categorías; es decir, en datos enumerativos.* " [Siegel, 1976:43-44].

1.2.2 Errores presentes en el muestreo

Kish [1972:24] señala que el diseño de muestras incluye dos grandes procesos: el de selección de los elementos y el de la estimación de valores de la población, esto con el propósito de realizar inferencias estadísticas aceptables de los datos muestrales hacia datos poblacionales. Ahora bien, por inferencia estadística se entiende el derivar conclusiones o hacer predicciones con el empleo de datos obtenidos de muestras aleatorias. "*Típicamente, en la inferencia estadística los datos muestrales son empleados para obtener inferencias acerca de una o más poblaciones desde las cuales las muestras han sido extraídas*" [Sheskin, 2000:1]. Más ampliamente, Quintana [1996: 17, 49] define la inferencia como la aplicación de "*... instrumentos estadísticos para obtener conclusiones y hacer generalizaciones válidas a la población bajo estudio, a partir de la información proporcionada por una muestra seleccionada al azar de tal población* **...** *bajo conocimiento que esta generalización de resultados no es 100% segura, sino que conlleva un cierto margen de error asociado***...**".

Este error asociado proviene tanto de los errores muestrales y de los no muestrales. Kish [1975:587] divide los errores en errores variables, que se suponen son aleatorios, y en sesgos [3/], que son errores sistemáticos de las mediciones; asegurando que *"... los errores de muestreo constituyen la mayor parte de los errores variables de una encuesta, y los sesgos provienen fundamentalmente de fuentes ajenas al muestreo"*. La adición de ambos errores (muestrales y no muestrales) es conocido como el Error Total de una muestra, y aunque en la realidad se desconoce su magnitud, el conocimiento de su existencia es suficiente para aceptar con cierto margen de incertidumbre las estimaciones últimas proporcionadas por una encuesta.

[3/] Según la Real Academia Española (RAE), sesgo significa: torcido, cortado o situado oblicuamente; oblicuidad o torcimiento de una cosa hacia un lado, en el corte, o en la situación, o en el movimiento. En sentido estadístico, "*el sesgo se refiere a los errores sistemáticos que afecta a cualquier muestra que se toma en un diseño muestral concreto con el mismo error constante*", en contraposición al error variable que se supone es aleatorio [Kish, 1975:587].

El proceso de estimación tiene que ver con los cálculos de estimaciones poblacionales con base en datos muestrales. Estos datos están sujetos a errores, por cuanto se trabaja con una parte de la población y no con la totalidad de los elementos de ella. Éstos se denominan errores y sesgos de muestreo.

El proceso de selección se refiere a la obtención de una posible muestra de tamaño *n*, del conjunto posible de muestras -también de tamaño *n*- de la población, con el objetivo de obtener una estimación. Todas estas posibles muestras, asociadas a un diseño muestral determinado, con miras a obtener una estimación específica -que puede ser, por ejemplo, la estimación del total o de la media-, configuran una distribución de muestreo de un estimador. En el caso específico de la media aritmética, ésta se define como *"... la distribución teórica de todos los valores posibles del estimador* $(\bar{y})$*, cada uno con su probabilidad de materialización (P). Los valores posibles y sus probabilidades dependen del diseño de muestreo (tamaño, selección y estimación) aplicado a una población fija de características"* [Kish, 1975:31].

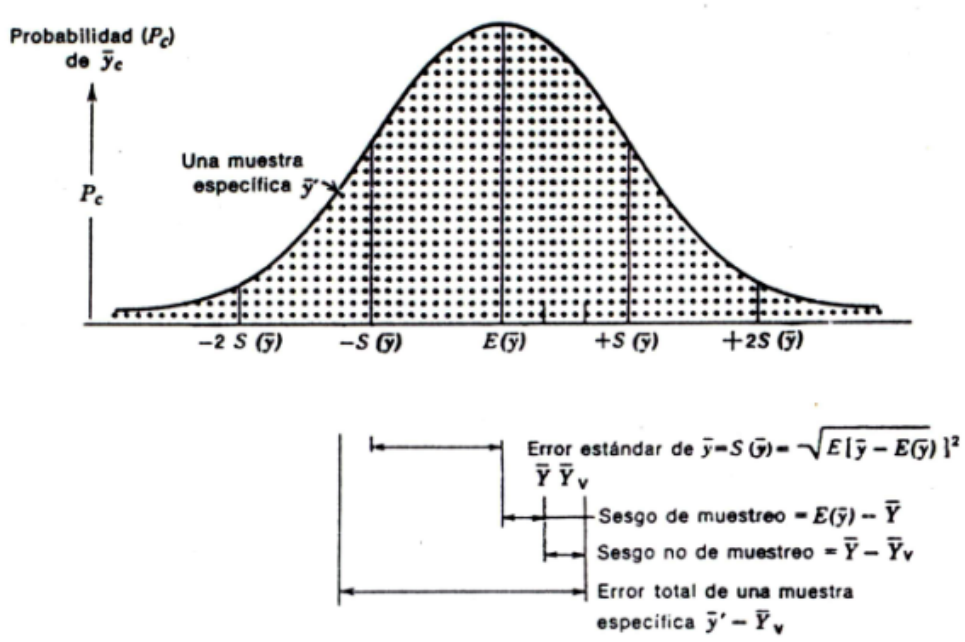
Algunos autores definen el error de muestreo como la diferencia entre el valor observado (estimado) y el valor desconocido (real) de la población sobre la que se está realizando la investigación. Así, el error de muestreo no es tanto una equivocación, sino un margen de incertidumbre [Díaz, 2004:46]. Los siguientes conceptos proporcionados por Kish [1975:29-34] y su nomenclatura, aclaran y amplían la definición, cuando lo que se quiere estimar es la media de una variable:

- La media muestral $\left(\bar{y}\right)$ se obtiene con los datos de la muestra. Ésta depende del diseño empleado y de la combinación particular de los elementos seleccionados, de tal forma que, aun tomando diferentes muestras de una misma población, del mismo tamaño y bajo el mismo diseño, producirían diferentes medias muestrales.
- La media de la población $\left(\bar{Y}\right)$ es una constante, generalmente, desconocida. Su valor se obtendría si se hiciera un censo, por lo que depende de sus N valores y,

consecuentemente, está afectada por errores no muestrales. Se estima con base en la media muestral.

- La diferencia entre ambas $\left(\bar{y} - \bar{Y} \right)$ es la desviación real y es desconocida. Esta diferencia comprende tanto errores muestrales como no muestrales.
- La media verdadera de la población $\left(\bar{Y}_v \right)$ es una constante teórica desconocida y está exenta de cualquier error.
- La diferencia entre las dos constantes $\left(\bar{Y} - \bar{Y}_v \right)$ incluye exclusivamente errores no muestrales.
- La diferencia entre la media muestral y la media verdadera $\left(\bar{y} - \bar{Y}_v \right)$ representa el error total.

De una manera muy didáctica, Kish [1975:33] grafica estos errores en el siguiente esquema.



El gráfico representa una distribución de muestreo -dado un diseño y tamaño de muestra- con los valores de las medias muestrales posibles y sus probabilidades de

ocurrencia. La media de esta distribución es el valor esperado del estimador $E(\bar{y})$, que en este caso, es la media poblacional, por lo que $E(\bar{y}) = \sum_c \left(P_c \, \bar{y}_c \right)$. A continuación los errores y sesgos mostrados en el esquema:

- **Error estándar de $\bar{y}$:** a la desviación estándar de la distribución de muestreo $S(\bar{y})$ se le denomina error estándar, y puede conceptualizarse como la raíz cuadrada de la varianza de la distribución muestral, que equivale a la raíz cuadrada de la desviación cuadrática media alrededor de la media $E(\bar{y})$.

  $$S(\bar{y}) = \sqrt{Var(\bar{y})} = \sqrt{E\left[\bar{y} - E(\bar{y})\right]^2} = \sqrt{\sum_c P_c \left[y_c - E(\bar{y}) \right]^2}$$

  En la práctica, solamente se obtienen los datos de una muestra y se conoce únicamente un solo punto de la distribución muestral; por ello, se han desarrollado diferentes fórmulas para obtener la estimación del error estándar para muchos diseños muestrales.

- **Sesgo de muestreo:** la $E(\bar{y})$ puede ser igual o no al valor poblacional $\bar{Y}$, por lo que a la diferencia $E(\bar{y}) - \bar{Y}$ se le denomina sesgo de muestreo. Un ejemplo sin este sesgo, es en el muestreo irrestricto aleatorio (MIA*)*: el diseño más simple de todos, productor de media y varianza insesgadas.

- **Sesgo no de muestreo:** surge de los procedimientos de observación imperfectos (sesgos de medición) y de otros tipos de errores en las etapas de crítica, codificación y procesamiento electrónico.

- **Error total de una muestra específica:** si se considera a $\bar{y}'$ como un valor muestral específico, $\bar{Y}$ el valor de la población y, $\bar{Y}_v$ el valor verdadero, se tiene que:

$\bar{y}' - \bar{Y}_v$ = ***Error Total de una muestra específica***

$\bar{y}' - \bar{Y}$ = ***Error de muestreo*** + ***Sesgos***

= ***Error de muestreo*** + ***Sesgo de muestreo*** + ***Sesgos no de muestreo***

$\bar{Y} - \bar{Y}_v$ = ***Sesgos no de muestreo***

Ya se mencionó que el error estándar corresponde a la raíz cuadrada de la desviación cuadrática media alrededor de la media $E(\bar{y})$. Análogo a esta medición, el error cuadrático medio consiste en la desviación cuadrática media alrededor del valor de la población $\bar{Y}$:

$$ECM(\bar{y}) = \sum_c P_c \left(\bar{y}_c - \bar{Y} \right)^2$$

$$= Var(\bar{y}) + \left[E(\bar{y}) - \bar{Y} \right]^2$$

Error Cuadrático Medio	=	*Varianza de la distribución muestral*	+	*Cuadrado del sesgo de muestreo*

El primer término es la varianza de la distribución muestral (cuadrado del error estándar), y el segundo, es el cuadrado del sesgo de muestreo que llega a anularse en diseños de muestreo insesgados, con lo cual $ECM(\bar{y}) = Var(\bar{y})$, lo que implicaría entonces que, $\sqrt{ECM(\bar{y})} = Error\ estándar$.

Dentro de este contexto, se dice que un estimador cualquiera $\hat{y}$ es [Lohr, 2000:28]:

- **insesgado**, si $E(\hat{y}) = Y$;
- **preciso**, sí $V(\hat{y}) = E\left[\hat{y} - E(\hat{y}) \right]^2$ es pequeña; y
- **exacto**, si $ECM(\hat{y}) = E\left(\hat{y} - Y \right)^2$ es pequeño.

Adicionalmente, el problema de la precisión de las estimaciones puede ser abordado también a partir de la construcción de intervalos de confianza. En forma general, podría representarse así [Cuesta y Herrero, 1999:8]:

$$P(\theta \pm Error\ Máximo) = 1 - \alpha$$

Donde: α es el nivel de confianza y,

Error máximo es el valor absoluto de la función de distribución correspondiente al percentil $\left(\frac{\alpha}{2} \right) S_e$

En síntesis, el error de muestreo ocurre en cualquier encuesta; el sesgo de muestreo se anula en diseños muestrales insesgados y, los errores (o sesgos) no de muestreo, ocurren siempre. Con esto se quiere decir que si el diseño es insesgado, la atención debe prestarse a los errores muestrales, y a los no de muestreo.

Concretamente, el error de muestreo está definido como el error ocasionado por entrevistar a una parte de la población objeto de estudio, y es un indicador de la precisión de las estimaciones realizadas [Díaz, 2004:47]. Cuanto más pequeño sea este valor, la calidad del estimador es mejor [INEI, 2001]. Su importancia radica en que la inferencia estadística se basa en estos errores; por ello, son tan importantes dentro de la teoría de encuestas por muestreo. Esta inferencia toma la forma de:

$$\bar{v} \pm t_p\, s(\bar{v})$$

donde: $\bar{v}$ = estadística no especificada

$s(\bar{v})$ = error estándar respectivo

t_p = constante escogida, imagen de una función de densidad estadística (generalmente normal o Student), asociada a la probabilidad *P*.

[Kish, 1975:35-37].

Cada diseño muestral tiene su propio cálculo de varianza muestral, y por ende, de error estándar. El diseño más simple de todos es el Muestreo Irrestricto Aleatorio (MIA), y aunque es poco utilizado en la práctica, la sencillez de sus propiedades matemáticas da el fundamento teórico requerido. Además, este diseño permite que todas las otras selecciones probabilísticas puedan verse como restricciones a ésta, y su cálculo suele ser utilizado como base para ajustarlo después, mediante el efecto del diseño de muestreo que en realidad se haya empleado [Kish, 1975:61-62]. Las fórmulas del MIA, reconocidas por Kish con el subíndice cero al lado de la variable (y_0), son las siguientes:

$$\textit{Error estándar de la media} \quad = \quad s(\bar{y}_0) = \sqrt{(1-f)\,\frac{s^2}{n}}$$

$$\textit{Error estándar del total} \quad = \quad s(N\bar{y}_0) = N\,s(\bar{y}_0) = N\sqrt{(1-f)\,\frac{s^2}{n}}$$

en donde: s^2 = *Varianza de los elementos de la muestra*

$$= \frac{1}{n-1}\sum_i^n (y_i - \bar{y})^2 = \frac{1}{n-1}\left[\sum_i^n y_i^{\,2} - \frac{y^2}{n}\right]$$

Con $y = \sum_j^n (y_i)$

= Total simple de la muestra para la variable Y_i.

f = *fracción de muestreo* = $\frac{n}{N}$

La fórmula del error estándar en un MIA es clara: pondera el promedio de la varianza insesgada de los elementos de la muestra con respecto a aquellos no incluidos en

ella, mediante la corrección para poblaciones finitas $(1-f)$ [4/]; lo mismo ocurre con la estimación del error estándar del total. Análogamente sucede en cada uno de los diferentes diseños muestrales simples [5/]: la varianza muestral media es insesgada y esta variabilidad es "reflejada" hacia el resto de la población.

El *error relativo* o *coeficiente de variación* de un valor estimado es el indicador utilizado para comparar errores estándares entre variables cuyas unidades de medida son diferentes. La EHPM [INEC, 2004:75] define el *error relativo* como el cociente del *error estándar* entre el *valor estimado*, con lo cual se obtiene más claramente el nivel de precisión de una estimación.

$$\textit{Error relativo o Coeficiente de variación} = \frac{\textit{Error estándar}}{\textit{Valor estimado}} \quad [6/]$$

La experiencia ha demostrado que existen rangos de precisión aceptables para los valores obtenidos del *error relativo*. Estimaciones con un *coeficiente de variación* de hasta un 5% son muy precisas; si el *coeficiente de variación* llega hasta un 10%, las estimaciones siguen siendo precisas; un *coeficiente de variación* con un valor de hasta 20% es aceptable; y por último, más allá de un 20%, indica que la estimación es poco confiable, y por tanto, se debe utilizar con precaución [INEC, op cit].

[4/] Para métodos de selección de igual probabilidad (mesip), f es la fracción de muestreo uniforme general para los elementos, y la corrección para poblaciones finitas es: $(1-f) = \left(1 - \frac{n}{N}\right)$.

[5/] Además del MIA, se pueden mencionar el muestreo estratificado y el muestreo de conglomerados.

[6/] Específicamente Kish [1975:72] lo menciona como el coeficiente de variación de la media ($\bar{y}$), el cual es el mismo para el total. La fórmula desarrollada por él es la siguiente:

$$CV(N\bar{y}) = \frac{N\,s(\bar{y})}{N\bar{y}} = \frac{s(\bar{y})}{\bar{y}} = CV(\bar{y})$$

En la vida real, se debe optar entre una posición extrema de simplicidad de diseño, lo cual arroja fórmulas simples, a practicidad y bajos costos para la organización del trabajo de campo. Esta última opción se logra mediante diseños complejos que combinan diferentes tipos de selecciones para lograr el objetivo deseado: buenas estimaciones al menor costo posible. Ahora bien, conforme aumenta la complejidad del diseño, se incrementa también la complejidad de los cálculos necesarios para estimar el error muestral, lo cual queda parcialmente subsanado con la utilización de paquetes estadísticos adecuados.

Existen dos familias de métodos para la estimación de los errores muestrales. Unos son analíticos, basados en la función de optimización del modelo, y otros, de simulación o remuestreo, basados en la distribución, paramétrica o no, de las variables o los parámetros del modelo [Cerviño, 2004:90]. En particular, para los diseños complejos de muestra, los métodos para la estimación de la varianza son el de aproximación de Serie de Taylor y el de replicaciones repetidas [Lepkowski y Bowles, 1996:2-3].

1.2.3 Aproximación de Serie de Taylor (Método clásico o tradicional)

Se le denomina también método de linealización, porque el teorema de Taylor permite linealizar una función suave, no lineal $h\,(t_1, t_2, \ldots, t_k)$ de los totales de la población, lo que proporciona las constantes $a_0, a_1, \ldots, a_k$ tales que:

$$h\,(t_1, \ldots, t_k) \;\approx\; a_0 + \sum_{i=1}^{k} a_i\, t_i$$

Con lo que se podría aproximar $V\left[h\,(\hat{t}_1, \ldots, \hat{t}_k)\right]$ mediante

$$V\left[\sum_{i=1}^{k} a_i\, \hat{t}_i\right] \;=\; \sum_{i=1}^{k} a_i^2\, V\left(\hat{t}_i\right) \;+\; 2\sum_{i=1}^{k}\sum_{j=i+1}^{k} a_i\, a_j\, Cov\left(\hat{t}_i, \hat{t}_j\right)$$

Estas aproximaciones realizadas con serie de Taylor, se han utilizado durante mucho tiempo para calcular las variancias aproximadas [Lohr, 2000:286].

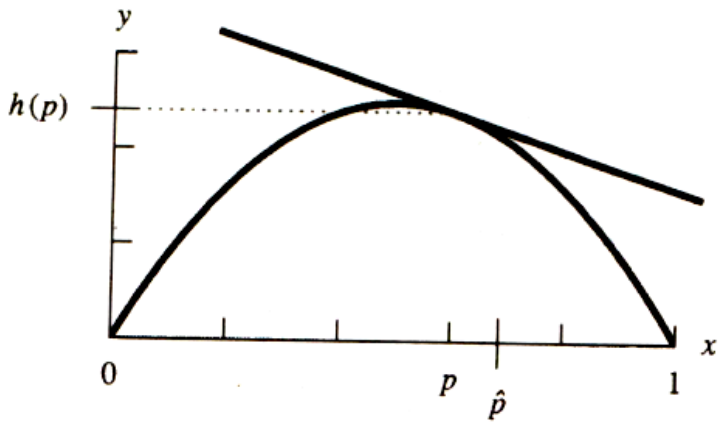
Por ejemplo, si $\theta = p(1-p)$, donde p es una proporción de la población, esta función se puede estimar mediante $\hat{\theta} = \hat{p}(1-\hat{p})$ Ahora bien, si se tiene que:

a. $\hat{p}$ es un estimador insesgado de p;

b. $V(\hat{p})$ es conocido;

c. $h(x) = x(1-x)$, de modo que $\theta = h(p)$ y $\hat{\theta} = h(\hat{p})$

En este caso, se tiene que h es una función no lineal de x, pero se puede aproximar en cualquier punto cercano a a mediante la recta tangente a la función. La pendiente de la recta tangente está dada por la derivada, como se muestra a continuación:



La versión de primer orden del teorema de Taylor, establece que si la segunda derivada de *h* es continua, entonces:

$$h(x) = h(a) + h'(a)(x-a) + \int_a^x (x-t)\,h''(t)\,dt$$

En condiciones que generalmente se satisfacen en estadística, el último término es pequeño con respecto a los dos primeros, por lo que se usa la aproximación:

$$h(\hat{p}) \approx h(p) + h'(p)(\hat{p}-p)$$
$$= p(1-p) + (1-2p)(\hat{p}-p)$$

Entonces:

$$V[h(\hat{p})] \approx (1-2p)^2 V(\hat{p}-p)$$

Y puesto que $V(\hat{p})$ es conocida, se puede calcular la variancia aproximada de $h(\hat{p})$ [Lohr, 2000:287].

1.2.4 Método de Replicaciones Repetidas

Con excepción de los grupos aleatorios, que dividen la muestra en grupos más pequeños para estimar la varianza muestral, los otros métodos que se mencionan en este apartado, son de remuestreo. Las técnicas de remuestreo construyen poblaciones hipotéticas derivadas del mismo conjunto de observaciones muestrales; es decir, la muestra original es considerada como si, en sí misma, fuese una población de la cual se pueden extraer diferentes muestras para estimar una varianza con base en esas submuestras. Algunos de estos métodos son los siguientes:

- **Grupos Aleatorios:** El diseño básico de la encuesta se copia de manera independiente *R* veces; es decir, cada vez que se extrae una muestra, sus unidades se reemplazan en la población de modo que estén disponibles para muestras posteriores. Entonces, las *R* réplicas de la muestra producen *R* estimaciones independientes de la cantidad de interés; la variabilidad entre esas estimaciones puede servir para estimar la varianza de $\hat{\theta}$ [Lohr, 2000:289-290]. Cada grupo, además de ser una submuestra aleatoria de la muestra completa, es una muestra de la población con las mismas propiedades probabilísticas, si bien de menor tamaño [Mirás, 1985:225-226].
- **Réplica repetida balanceada (RRB):** La réplica repetida balanceada (RRB) utiliza la variabilidad entre R réplicas de mitades de muestra, seleccionadas de manera equilibrada para estimar la varianza de $\hat{\theta}$. Este método se utiliza en diseños

estratificados, de tal forma que si se extraen dos unidades en cada estrato, en total se pueden formar 2^H mitades de muestra posibles, donde H = número de estratos, con la sugerencia de que se utilicen todas para la estimación de la varianza. El procedimiento solicita que se obtenga un par ordenado de observaciones para cada estrato, que identifique claramente su pertenencia a la primera o segunda observación del par, con lo que se asocian a pares de pertenencia (1, -1). Si la mitad de la muestra r contiene una de las dos observaciones de cada par, para todos los estratos, y la otra mitad contiene el complemento del par, se dice que la réplica está balanceada [Lohr, 2000:295-296].

- **Jacknife (Navaja):** El Jacknife es una técnica de remuestreo, que amplía el método de grupos aleatorios, lo que permite que las réplicas de los grupos se traslapen. En el denominado *navaja con una eliminación*, se seleccionan n muestras que contienen n-1 puntos diferentes de los n puntos de la muestra original x . [Lohr, 2000:300]
- **Técnica Bootstrap:** La técnica bootstrap trata la muestra como si fuera la población, y aplica un remuestreo para generar una estimación empírica de la distribución muestral del estadístico. Para ello, se extrae un gran número de "remuestras" de tamaño *n* de la muestra original, aleatoriamente y con reposición. Así, aunque cada remuestra tenga el mismo número de elementos que la muestra original, mediante el remuestreo con reposición, cada una podría contener algunos de los datos originales representados en ella, más de una vez, y otros no aparecerían [Cuesta y Herrero, 1999:3].

Esta última técnica se diferencia del enfoque tradicional paramétrico, en que se emplea un gran número de cálculos repetitivos para estimar la forma de la distribución muestral del estadístico, en lugar de fuertes asunciones y fórmulas analíticas [Cuesta y Herrero, 1999:2]. Precisamente, esta es la razón por la cual, para este estudio, se seleccionó dicha técnica, por lo que se explicará con mayor detalle a continuación.

1.2.5 Técnica Bootstrap (docimasia o iniciador [7/])

El "bootstrap" le permite al investigador hacer inferencias en casos donde las soluciones analíticas son inviables, y donde las asunciones sobre la distribución muestral del estadístico no se sostienen. El bootstrap no es, por tanto, un estadístico per se, más bien, es una aproximación al uso de la estadística para hacer inferencias sobre los parámetros poblacionales. La idea central es que, muchas veces, puede ser mejor extraer conclusiones sobre las características de la población, estrictamente a partir de la muestra que se maneja, que hacer asunciones, quizás, poco realistas, sobre la población [Cuesta y Herrero, op cit].

Por lo tanto, cada una de estas remuestras, probablemente será leve y aleatoriamente diferente de la muestra original. Dado que los elementos en estas varían poco, un estadístico $\hat{\theta}^*$, calculado a partir de una de ellas posiblemente asumirá un valor ligeramente diferente de los otros $\hat{\theta}^*$ y del $\hat{\theta}$ original. De aquí que se afirme que una distribución de frecuencias relativas de esos $\hat{\theta}^*$, calculada a partir de las remuestras, es una estimación de la distribución muestral de $\hat{\theta}$ [Cuesta y Herrero, 1999:3].

Se habla de *Bootstrap paramétrico* cuando se tiene algún conocimiento más allá del aportado por la muestra; por ejemplo, cuando se conoce la función de distribución de la variable objeto de estudio, pero se desconocen los parámetros que deben ser estimados. En caso contrario, cuando se extraen conclusiones sobre las características de la población, estrictamente a partir de la muestra que se maneja, se está en presencia del *Bootstrap no paramétrico* [Cuesta y Herrero, 1999:4].

[7/] Nombrado así por otros autores en Cuesta y Herrero [1999]:
Docimasia: (Del gr. δοκιμασία, de δοκιμάζειν, probar, ensayar) (Real Academia Española).
1. f. Arte de ensayar los minerales para determinar los metales que contienen y en qué proporción.
2. f. *Med.* Serie de pruebas a que se somete el pulmón del feto muerto para saber si ha respirado antes de morir.
Iniciador: (Del lat. *initiātor, -ōris*). **1.** adj. Que inicia. U. t. c. s. (Real Academia Española).

Cuando se aborda el problema de la precisión a través de los intervalos de confianza, se sabe que, en el contexto paramétrico clásico, es necesario conocer la distribución muestral del estadístico, tanto en lo que se refiere a su forma como a su error estándar. Sin embargo, si la distribución es asimétrica, un intervalo de confianza calculado de la manera clásica no estará centrado en la distribución, con lo que se corre el riesgo de que la media poblacional no se encuentre dentro de dicho intervalo, o en el mejor de los casos, que se encuentre más próximo a uno de los límites [Núñez, 2003:5].

Desde el enfoque del "bootstrapping", se han propuesto varios métodos de construcción de intervalos confidenciales alternativos al enfoque clásico: los métodos de aproximación normal y del percentil son los más intuitivos; pero además, existen otros [Cuesta y Herrera, 1999:8]:

- *Método de la aproximación normal:* Este método es similar al procedimiento paramétrico de construcción de intervalos. Si es posible asumir que el estadístico se distribuye según la curva normal, pero el cálculo del error típico resulta analíticamente difícil, o no existe fórmula para su cálculo, entonces se puede emplear la distribución muestral "bootstrap" para estimar el error típico e insertarlo en la correspondiente expresión del intervalo de confianza.

  El punto débil de este método es que renuncia a su esencia de técnica no-paramétrica, pues se basa en un fuerte supuesto: la de normalidad de la distribución, de tal forma que cuando esta asunción sea violada, los resultados obtenidos no serán mejores que con el método tradicional [Cuesta y Herrero, 1999:8-10].
- *Método percentil:* consiste en obtener los percentiles respectivos para producir el intervalo deseado. Así, si se desea un intervalo de confianza del 90%, se deben conocer los percentiles $P_{0,05}$ y $P_{0,95}$ de la distribución "bootstrap" [Núñez, 2003:13]. Este método conserva la esencia no-paramétrica del enfoque "bootstrap" y libera al usuario de las asunciones de la estadística paramétrica. "El número de remuestras necesario para llevar a cabo la construcción de intervalos confidenciales por este método (y en general por los métodos bootstrap) se fija entre 1000 y 2000 (Efron,

1988), para intentar garantizar una mejor estimación de las colas de la distribución" [Cuesta y Herrero, 1999:10].

- *Método percentil corregido por el sesgo*: Realiza un ajuste al método percentil para aproximar la distribución normal [Núñez, 2003:13].
- *Método percentil corregido por el sesgo acelerado:* Efectúa otra adecuación sobre el método anterior, con el fin de tomar en cuenta la asimetría de la distribución bootstrap [Núñez, op cit].
- *Método bootstrap – t:* Obtiene los percentiles respectivos al α y al $(1-\alpha)$ en la distribución "bootstrap". A estos percentiles se les multiplica por el error de muestreo, lo cual forma la amplitud del intervalo [Núñez, op cit].

Para estimar un estadístico bajo el método bootstrap, se deben seguir algunos pasos básicos. Éstos son así definidos por Núñez [2003:8-9]:

- Se construye una distribución de probabilidad empírica $\hat{F}(x)$ a partir de la muestra, con una probabilidad de $\frac{1}{n}$ para cada punto, x_1 , x_2 , ..., x_n.
- A partir de ella, se extrae una muestra aleatoria simple de tamaño n con reposición; esta es una "remuestra" x_b^*.
- A partir de ésta se calcula el estadístico $\hat{\theta}_1^*$.
- Se realizan B repeticiones de los pasos anteriores, con lo que se obtiene un conjunto de B estadísticos, y se realizan B estimaciones independientes de θ:

$$\hat{\theta}_1^*,\ \hat{\theta}_2^*,\ \cdots,\ \hat{\theta}_B^*$$

- Con esos B estadísticos resultantes del remuestreo, se construye una nueva distribución de probabilidad, donde cada estimador $\hat{\theta}_b^*$ tendrá una probabilidad de $\frac{1}{B}$ en dicha distribución.

La magnitud de *B*, en la práctica, depende de las pruebas que se les aplica a los datos. En general, B debería estar entre 50 y 200 para estimar el error típico de $\hat{\theta}$, y al menos de 1000 para estimar intervalos de confianza alrededor de $\hat{\theta}$.

Esta distribución es la estimación bootstrap de la distribución muestral de $\hat{\theta}$, $\hat{F}^*\left(\hat{\theta}^*\right)$, y puede usarse para inferir sobre θ.

El estimador bootstrap del parámetro θ se define como:

$$\hat{\sigma}^*_{(.)} = \frac{\sum_{b=1}^{B}\left(\hat{\theta}_b *\right)}{B}$$

Es decir, es la media de los valores del estadístico calculados en las B remuestras bootstrap. Por su parte, la desviación estándar de la muestra se expresa por:

$$\hat{\sigma}^*_B = \left(\frac{\sum_{b=1}^{B}\left[\hat{\theta}^*_b - \hat{\theta}^*_{(.)}\right]}{B-1} \right)^{\frac{1}{2}}$$

Esta fórmula se asemeja a la expresión matemática para la desviación estándar del muestreo irrestricto aleatorio. Por lo anterior, se puede decir que con la técnica bootstrap se obtiene una muestra irrestricta de B estadísticos, a partir de los cuales se obtiene la estimación de la distribución poblacional [Núñez, 2003:11].

1.2.6 Ventajas y desventajas de los Métodos

Cada uno de los métodos descritos anteriormente, posee ventajas y desventajas. A continuación, se muestran las más sobresalientes de ellas.

Método	Ventajas	Desventajas
Aproximación serie de Taylor	Si se conocen las derivadas parciales, casi siempre la linealización proporciona una estimación de la varianza para una estadística, aplicable en diseños generales de muestreo [Lohr, 2000:289]. Existe un gran desarrollo de software especializado en este método [Lohr, op cit].	Esta aproximación clásica supone distribución normal [Cuesta y Herrero, 1999:1]. No todas las estadísticas se pueden expresar como una función suave de los totales de la población, como la mediana [Lohr, 2000:289]. Para funciones complejas que impliquen actualización de pesos, el método puede ser difícil de aplicar, con lo cual, se necesitaría otra fórmula de varianza para cada estadística no lineal estimada, que podría requerir programación especial [Lohr, op cti].

Método	Ventajas	Desventajas
Replicaciones repetidas		
a. Grupos aleatorios [Lohr, 2000:293]	Para estimar la varianza, no se requiere de software especializado, además de que se pueden estimar varianzas de percentiles y de funciones no suaves.	Se requiere de por lo menos 10 grupos aleatorios para obtener una estimación estable de la varianza, y no inflar el intervalo de confianza con la distribución *t* en lugar de la distribución normal. Estos grupos aleatorios, en diseños complejos, pueden ser difíciles de formar, ya que cada grupo debe tener la misma estructura que la muestra completa.
b. Réplica repetida balanceada (RRB) [Lohr, 2000:299]	Proporciona una estimación de la varianza asintóticamente equivalente a la que corresponde a métodos de linealización para funciones suaves de los totales de la población, y para los cuantiles.	Requiere de un diseño de dos unidades primarias por estrato, aunque en la práctica, se extiende a otros diseños de muestreo mediante esquemas de equilibrio más complejos. Este método estima la varianza con reemplazo y puede sobreestimar la varianza si los N_h (cantidad de unidades primarias en el estrato *h*) son pequeños.
c. Jacknife [Lohr, 2000:301-302]	Es de utilidad general; se utiliza el mismo procedimiento para estimar la varianza de cada estadística donde se pueda usar la navaja.	Tiene un pobre desempeño para estimar las varianza de algunas estadísticas, como los cuantiles en un MIA.

Método	Ventajas	Desventajas
d. Bootstrap	Está libre de supuestos acerca de las distribuciones de las variables, ya que su esencia es no paramétrica [Cuesta y Herrero, 1999:10]. La estimación del error estándar es posible para todos los estimadores, y no requiere de cálculos teóricos [Cuesta y Herrero, 1999:5]. Los intervalos con bootstrap son efectivos capturando la asimetría [Efron y Tibshirani 1986, mencionado en Núñez, 2003:13]. Propone varios métodos de construcción de intervalos de confianza alternativos al enfoque clásico [Cuesta y Herrero, 1999:8]. Sirve para funciones no suaves (como los cuantiles) en diseños generales de muestreo [Lohr, 2000:303-304].	Requiere más cálculos que las técnicas de *RRB* o del Jacknife, pues generalmente, B es un número muy grande [Lohr, 2000:303-304]. Existe menos desarrollo de trabajo teórico [Lohr, op cit].

1.2.7 Situación actual

En la publicación de la EHPM de 1995, el INEC expone por primera vez los errores de muestreo correspondientes a esa encuesta. Éstos se presentan de una forma sencilla, para tres grupos de indicadores:

- *condición de actividad* (*población total*, *fuerza de trabajo*, *ocupados* y *desocupados*);
- *tasas* (*bruta* y *neta de participación*, *de ocupación* y *desempleo abierto*); y
- *hogares pobres* y *no pobres* e *ingresos promedio* y *per cápita del hogar*.

A partir de la publicación de 1996, la información suministrada se vuelve más rica, aunque se suprimen detalles en los hogares, esto es:

- la *condición de actividad* se desglosa por *regiones de planificación*;
- las *tasas* se presentan por *zona* y se desagregan por *rama de actividad*, *grupo* y *categoría ocupacionales*; y
- aunque los errores de los *hogares pobres* y *no pobres* permanecen para los años siguientes, se eliminan definitivamente los errores de los *ingresos del hogar*.

Desde 1999 la información publicada sufre nuevamente ampliaciones en el sentido de que se presentan como:

- Estimaciones de variabilidad de la *Población total* según *Condición de actividad* y *Región de planificación*.
- Estimaciones de variabilidad de la *Población ocupada* según *Rama de actividad*, *Grupo ocupacional* y *Categoría ocupacional*. Para la encuesta del 2005 se agrega el *Sector institucional*.
- Estimaciones de variabilidad de la *Tasa de participación* (*bruta* y *neta*), *de Ocupación* y *de Desempleo abierto*, por *Región* (*Central* y *Resto de regiones*) y *Zona*.
- Estimaciones de variabilidad del *Porcentaje de hogares*, según *Condición de pobreza* y *Zona*.

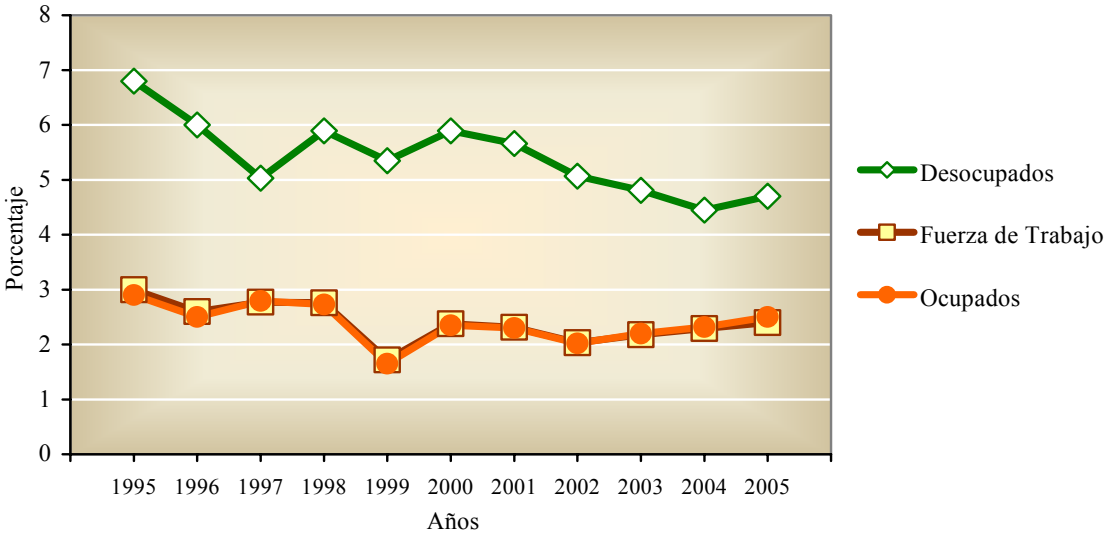
En los primeros años de la publicación, estos errores se presentan en un anexo. Para los últimos años, y siempre en la última sección, son incorporados como un capítulo adicional y presentados junto con otras medidas afines; esto es: el *valor estimado* (como punto de referencia), el *coeficiente de variación* (*error relativo*) y los *intervalos de confianza al 95%.* Antes de su presentación, y de manera pedagógica, se incorporan las definiciones respectivas.

Lo clásico, en todas las encuestas de hogares, es que el cálculo de los errores se fundamente en el método de serie de Taylor, y el INEC no es la excepción. Particularmente, esta institución los calcula a través del paquete IMPS, por varias razones:

- la principal, porque es un paquete especialmente diseñado para encuestas dirigidas a los hogares, con muestras complejas; y
- porque es un programa ampliamente utilizado y promocionado por el Bureau of Census de los Estados Unidos; es gratuito y está disponible en la página de Internet de esa Oficina de Censos, por lo que se facilita su actualización.

El contraste de la precisión entre diferentes variables, se logra a través de los *errores relativos* o *coeficientes de variación*. Con base en los rangos de precisión aceptables, se puede afirmar que en los últimos once años de ejecución de las EHPM, esos errores muestran estimaciones precisas (más de 5 y hasta 10%) para la variable *Desocupados*, lo que manifiesta una importante disminución en los últimos tres años, por lo que sus estimaciones son muy precisas (hasta 5%). Por su parte, la *Fuerza de Trabajo* y los *Ocupados* también obtienen estimaciones muy precisas a lo largo de todo el período, lo que muestra una consistencia temporal en los datos.

Gráfico 1.1
COSTA RICA: Errores Relativos (%) de los principales indicadores de la Condición de Actividad de la EHPM. 1995-2005



FUENTE: Instituto Nacional de Estadística y Censos. Encuesta de Hogares de Propósitos Múltiples. 1995 – 2005 (a julio de cada año).

1.3 METODOLOGÍA

La observación parcial de la población, desde un contexto de muestreo aleatorio, produce estimaciones lejanas, cercanas o iguales a su valor real. Cuán lejano o cercano se esté de ese valor real, depende de la elección particular de la muestra seleccionada. Como ya se ha mencionado, tal variabilidad se denomina error de muestreo, por lo que obtener una adecuada medición de él es uno de los objetivos principales de toda encuesta probabilística. La oportunidad brindada por la capacidad tecnológica, permite medir este error a partir de diferentes procedimientos y comparar los resultados.

Este trabajo muestra los resultados obtenidos del error estándar en la Encuesta de Hogares de Propósitos Múltiples (EHPM) de julio 2004, a través de dos métodos diferentes: aproximación de Serie de Taylor y Bootstrap. Respecto de estas comparaciones, algunos autores opinan que, aunque producen estimaciones distintas, estas no deberían alejarse demasiado en muchos estimadores [Lepkowski y Bowles, 1996:3]. La presentación de esos cálculos aparece en el siguiente orden:

- Se calcula la aproximación de la Serie de Taylor, medición clásica para este tipo de error, con el programa SPSS; que posee el mismo algoritmo utilizado por el INEC; y
- A través del método de replicaciones repetidas, denominado Bootstrap, calculado con el paquete Stata, que en síntesis, ejecuta una serie de selecciones de B muestras con reemplazo, del tamaño original, a partir de los mismos datos de la muestra.

Para el procesamiento de los datos, se modificaron (recodificar) algunas variables proporcionadas en la base de datos de la encuesta referida. Se crearon variables dicotómicas, más específicamente variables dummy [8/], con el objetivo de agilizar los procesamientos; los valores 0 y 1 indicaron la pertenencia a un grupo específico de interés en el procesamiento de los datos, como: la *Fuerza de trabajo*, *Ocupados*, *Desocupados*,

[8/] Variable dicotómica binaria, es decir, con valores 0 y 1.

Inactivos y *Menores de 12 años*. En el tema de la pobreza, la variable dicotómica indica la pertenencia o no a los grupos de *Pobreza total*, *Extrema pobreza*, *No satisfacen necesidades básicas* y *No pobre*.

Otras variables fueron recodificadas para sintetizar su desagregación original, como ocurrió con:

- *Condición de actividad* (*conda_re*): Ocupado, Desocupado, Inactivo; y
- *Nivel de educación* (*Educa_re*): cantidad de años acumulados en la educación formal (1, 2, … , 20).

En la definición del diseño muestral complejo [INEC, 1999:10], se creó la variable *"Estrato"* que corresponde efectivamente con el concepto estadístico que responde al mismo nombre. En la EHPM éste se forma por el cruce de las variables:

- *Nivel socioeconómico*, para los niveles *bajo*, *medio bajo*, *medio*, *medio alto*, *alto* y *rural*; con códigos del 1 al 6 en el orden mencionado; y
- *Región de planificación*, correspondientes a *Región Central, Chorotega*, *Pacífico Central*, *Brunca*, *Huetar Atlántico* y *Huetar Norte*, con códigos del 1 al 6, en ese mismo orden.

Para crear esta variable (*Estrato*), se unieron los dígitos de ambas, lo que produjo 36 estratos diferentes, aunque no todos quedaron representados en la muestra.

La selección de las variables e indicadores para realizar las comparaciones entre ambos métodos, tomó como criterios básicos distintas construcciones y su importancia para los usuarios de las encuestas. En otras palabras, se incluyeron totales, promedios, porcentajes y otros tipos de indicadores descritos en la publicación de la EHPM, pero que además, fueran indicadores de alto interés para los usuarios. Otro elemento por considerar, fue que para cada procesamiento se requería de mucho tiempo, por lo que tendrían que

seleccionarse pocos, pero de interés global. Con base en los criterios anteriores, se seleccionaron los siguientes indicadores:

Tipo	Indicador
TOTALES	**Población total**
	Fuerza de trabajo
	Ocupado
	Desocupado
	Inactivos (de 12 años y más)
	Menores de 12 años
	Total jefes de hogar
	Jefes en fuerza de trabajo
	Jefes ocupados
	Jefes desocupados
	Jefes inactivos (de 12 años y más)
TASAS	Tasa bruta de participación
	Tasa neta de participación
	Tasa de ocupación
	Tasa de desempleo abierto
PORCENTAJES	% Hogares no pobres
	% Hogares pobres
	% Hogares en extrema pobreza
	% Hogares no satisfacen lo básico
PROMEDIOS	Ingreso promedio ocupación principal
	Ingreso promedio del hogar
	Ingreso promedio percápita del hogar
	Nivel educativo promedio por ocupado
	Horas promedio trabajadas

Tipo	Indicador
OTROS INDICADORES	Miembros por hogar Fuerza de trabajo por hogar Ocupado por hogar % Hogares con jefatura femenina

De estos 28 indicadores, los 11 que se agrupan dentro de los totales provienen de variables nominales (pertenecen o no al grupo respectivo). Las Tasas y Porcentajes se sustentan en variables dicotómicas, y los Promedios, en variables continuas y discretas; no obstante, todas ellas, por ser resultado de un cociente, son variables continuas, aunque de única observación en este ejercicio. Las variables restantes son discretas, y por lo tanto, susceptibles de graficarles una distribución; ellas son *Miembros por hogar*, *Fuerza de trabajo por Hogar* y *Ocupados por hogar*, y sus distribuciones apuntan a que son asimétricas positivas (véase anexo B).

1.3.1 Método aproximación Serie de Taylor

El cálculo del error muestral bajo el método tradicional (aproximación Serie de Taylor), se realizó con el paquete SPSS, y requirió de una definición preliminar del diseño complejo [SPSS 2006:1] que sería utilizado para todos los cálculos que se hicieran del error muestral para las variables deseadas. Para la construcción de este diseño, se usó el módulo de *Muestras Complejas* del paquete anteriormente mencionado, y con base en él, se desarrolló el cálculo, tanto de las estimaciones poblacionales, como de los errores muestrales asociados a ellas.

La construcción del diseño particular de la EHPM, se logró a través de la *Preparación para el análisis (Prepare for Analysis)* de la opción *Muestras complejas (Complex Samples)* en el menú *Analizar (Analyze)* del SPSS. Aquí se especificó el nombre

de las variables que asumen los roles de estrato (*estrato*), conglomerado (*consecu* [9/]) y factor de expansión (*factor*). Una vez especificado este diseño, se continuó trabajando este mismo módulo para obtener las estadísticas deseadas, como: la estimación poblacional, el tamaño de la muestra asociada, el error estándar y el coeficiente de variación. A partir de aquí se pudo iniciar el cálculo de los errores bajo este diseño (las salidas obtenidas se sintetizan en el anexo A).

1.3.2 Técnica Bootstrap

El comando Bootstrap del paquete Stata fue utilizado para calcular los errores muestrales de las mismas variables que se elaboraron para el algoritmo aproximaciones de Serie de Taylor. Una característica de este comando, es que la duración del procesamiento de los datos puede tardar más de una hora en cada uno, ya que la orden especificada es que sean 1000 repeticiones para conseguir la estimación del error y la de los intervalos de confianza. No obstante, y posiblemente debido a limitaciones en la capacidad de procesamiento del computador utilizado, las replicaciones producidas superaron las 400, por encima de la requerida para la estimación de errores muestrales (50), según algunos autores [Núñez, 2003:10], pero inferiores a las recomendadas para lograr estimaciones confiables de los intervalos de confianza (1000).

Si bien los comandos por utilizar son repetitivos, éstos se realizan sobre archivos de datos parciales de acuerdo con la población requerida para obtener la estimación; esto con el fin de agilizar el proceso sin afectar su calidad. Así, la base de datos total se fragmentó (a semejanza del *Select* en el *SPSS*) para obtener Jefes de hogar, Ocupados, Ocupados con horas conocidas, Hogares con ingreso conocido, etc.

En cada uno de esos archivos de datos, se corrieron los comandos necesarios para calcular los errores de las mismas variables e indicadores mencionados en el apartado

[9/] Esta variable representa al segmento.

anterior, a fin de lograr los contrastes deseados. La secuencia de las instrucciones, con el ejemplo de *Población total* por *Condición de actividad* ejecutado en el Stata, es la siguiente:

Explicación	Comando
Reserva de 800 megas de memoria ram para el procesamiento.	set mem 800m
Apertura de la base deseada	use "C:\Yadi\Bases\EHPM2004_Nivel.dta", clear
Definición del mismo diseño muestral empleado en el SPSS, que debe ejecutarse cada vez que se abre una nueva base [Stata 2003:70-72].	svyset [pweight=factor], strata(estrato) psu(consecu)
Comando bootstrap para: • Población total (*todos*), • Fuerza de trabajo (*fuerza*) • y Población total por Condición de actividad (*todos, by (conda_re)*). La expresión *svytotal* asegura la estimación de un conteo total a partir del diseño muestral definido con anterioridad. */	bootstrap `"svytotal todos"' _b, reps(1000) strata(estrato) cluster(consecu) bca
	bootstrap `"svytotal fuerza"' _b, reps(1000) strata(estrato) cluster(consecu) bca
	bootstrap `"svytotal todos, by(conda_re)"' _b, reps(1000) strata(estrato) cluster(consecu) bca
Cierre de la base	clear

*/ Además se le solicita al comando Bootstrap lo siguiente:

- todos los parámetros: *_b*;
- con 1000 repeticiones: *reps(1000)*;
- bajo el diseño muestral establecido: *strata(estrato) cluster(consecu)*;
- despliegue de todos los intervalos de confianza: normales, percentiles, sesgo corregido y, sesgo corregido y acelerado: *bca*.

1.3.3 Comparación de ambos métodos

Los *valores estimados*, obtenidos a través de ambos procedimientos, proporcionan igual resultado, dado que se basan en el mismo diseño muestral. Por ejemplo, la *estimación* de la *Población total* para Costa Rica, es la misma que se obtuvo tras calcular los errores *Serie de Taylor* (SPSS), y los errores *Bootstrap* (Stata): la cifra fue de 4.178.766 personas. Así ocurre en todas las demás variables; aunque, claro está, con diferentes resultados en sus *errores estándares*.

Un indicador de comparación entre los dos tipos de cálculos es la relación entre los *coeficientes de variación*, puesto que como se dijo, éste elimina el efecto de unidades de medición diferente. Sin embargo, como las *estimaciones* de las variables respectivas son idénticas (los denominadores), lo que se relacionaría en realidad, serían los *valores absolutos* de los *errores estándares*:

$$\frac{CVB_X}{CVT_X} = \frac{Coeficiente\ Variación\ Bootstrap\ \ (VariableX)}{Coeficiente\ Variación\ Serie\ Taylor\ \ (VariableX)}$$

$$= \frac{\dfrac{Error\ estándar\ Bootstrap\ \ (Variable\ X)}{Estimación\ Bootstrap\ \ (Variable\ X)}}{\dfrac{Error\ estándar\ Serie\ Taylor\ \ (Variable\ X)}{Estimación\ Serie\ Taylor\ \ (Variable\ X)}}$$

Puesto que :

$$\textit{Estimación Bootstrap}\ (\ \textit{Variable X}\ ) = \textit{Estimación Serie Taylor}\ (\ \textit{Variable X}\ )\ ,$$

entonces :

$$\frac{CVB_X}{CVT_X} = \frac{\textit{Error estándar Bootstrap}\ \textit{(Variable X)}}{\textit{Error estándar Serie Taylor}\ \textit{(Variable X)}}$$

La relación entre ambos errores para la misma variable, elimina también, el efecto de las unidades de medida, por lo que con esta, se alcanza el objetivo de comparar ambos errores. Si a esta relación se le resta una unidad, y el resultado de ello es multiplicado por cien, se obtendría un coeficiente diferencial entre ambos métodos de medición del error estándar.

$$\textbf{\textit{Coeficiente diferencial}} = \left(\frac{\textit{Error estándar Bootstrap}_{(X)}}{\textit{Error estándar Serie Taylor}_{(X)}} - 1 \right) * 100$$

No existiría diferencia alguna entre el error estándar obtenido por el método Bootstrap y el método Series de Taylor, si el resultado de este indicador es igual a cero. Las diferencias relativas positivas, significarían que el error estándar Bootstrap resultó mayor que el calculado con el método Serie de Taylor. En caso de ser negativas, sucedería todo lo contrario: Bootstrap resultó inferior que el Serie de Taylor.

No hay que perder de vista, para efectos del análisis, que cada cálculo obtenido con el método Bootstrap, es uno de los tantos que se pudo obtener. Esto es así, por cuanto éste, resulta de un conjunto particular de selecciones de muestras aleatorias del mismo tamaño al de la original.

1.4 MAGNITUD DE LOS ERRORES DE MUESTREO EN LA EHPM

"El hecho de que las variables de interés para el estudio se midan solo en una parte de la población significa que los valores finales que se obtienen referidos a toda la población son estimaciones más o menos próximas a los valores reales y varían en función de la muestra concreta que se haya seleccionado, variabilidad que se mide por el error de muestreo o error estándar" [INE, 2004:29].

Los *errores estándares* fueron medidos por dos procedimientos diferentes: *series de Taylor* (método clásico) y *Bootstrap*. Para compararlos se construyó un indicador denominado *coeficiente diferencial*, que permitió verificar que en la mayoría de las variables utilizadas las diferencias fueron pequeñas, y que en las variables de razón continuas, el procedimiento clásico sobreestima estos errores.

La medición de los *errores estándares* se realizó en la EHPM de 2004 y, si bien es cierto, la técnica *Bootstrap* puede aplicarse a todas las variables, en esta investigación se seleccionaron 28 indicadores, clasificados en diferentes tipos, a saber: totales, tasas, porcentajes, promedios y otros. De ellas, 11 son variables nominales, y las restantes pertenecen a la escala de razón, aunque solo las tres de *ingresos* son continuas observables en la muestra.

Los *errores estándares*, así como los *coeficientes de variación*, son presentados desde los dos procedimientos estudiados: *series de Taylor* y *Bootstrap*. Es importante recordar que el *coeficiente de variación*, utilizado en este contexto, es la relación entre el *error estándar* y su respectiva *estimación*, presentado en términos porcentuales. Definido de esta manera, el *coeficiente de variación* permite la comparación de variabilidades entre diferentes variables, cuyas unidades de medición son diferentes.

Las diferencias entre las estimaciones obtenidas en el SPSS con el método series de Taylor y las publicadas por el INEC, son bastante pequeñas, y éstas se deben a que, aunque

se utiliza el mismo algoritmo, el INEC lo calcula a través de otro programa estadístico: el IMPS de la Oficina de Censos (Boreau of Census) de los Estados Unidos.

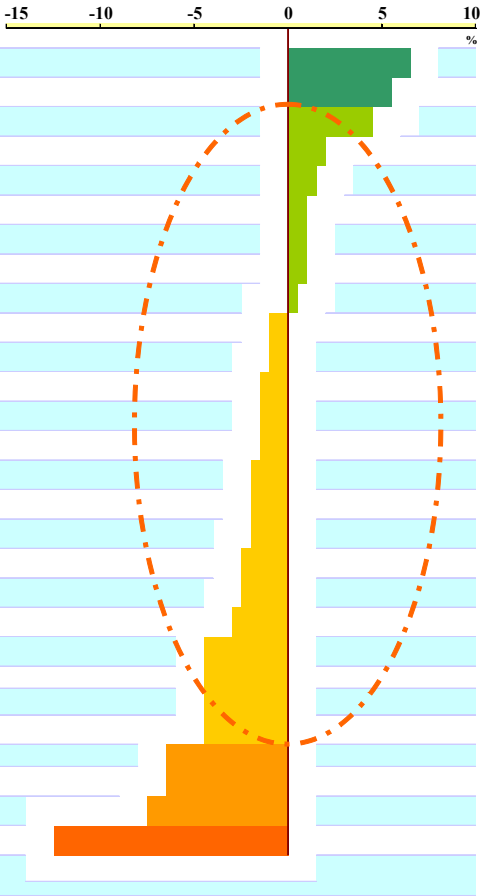
El resumen de los resultados obtenidos con ambos procedimientos (serie de Taylor y Bootstrap), se presenta a continuación:

Cuadro 1.1
COSTA RICA: Comparación de los errores estándares de la Encuesta de Hogares de Propósitos Múltiples
SEGÚN: Diferentes variables e indicadores
POR: Método de estimación
Julio 2004

Variable o Indicador	Valor estimado	Error estándar		Coeficiente diferencial [Bootstrap / Serie de Taylor - 1] * 100
		Serie Taylor	Bootstrap	Diagrama -15 -10 -5 0 5 10 %
TOTALES				
Población total	**4.178.755**	**94.416**	**90.359**	**-4,30 %**
Fuerza de trabajo	1.768.759	40.246	41.071	2,05 %
Ocupado	1.653.879	38.153	38.476	0,85 %
Desocupado	114.880	5.098	5.421	6,33 %
Inactivos (de 12 años y más)	1.481.721	33.813	35.404	4,70 %
Menores de 12 años	928.275	28.390	28.756	1,29 %
Total jefes de hogar	**1.095.345**	**22.672**	**22.372**	**-1,32 %**
Jefes en fuerza de trabajo	851.357	20.605	19.269	-6,48 %
Jefes ocupados	823.751	19.712	19.328	-1,95 %
Jefes desocupados	27.606	2.286	2.214	-3,14 %
Jefes inactivos (de 12 años y más)	243.988	6.687	6.559	-1,90 %
TASAS				
Tasa bruta de participación	42,33 %	0,33 %	0,33 %	1,09 %
Tasa neta de participación	54,42 %	0,38 %	0,37 %	-1,44 %
Tasa de ocupación	50,88 %	0,39 %	0,39 %	0,87 %
Tasa de desempleo abierto	6,49 %	0,26 %	0,25 %	-2,57 %
PORCENTAJES				
% Hogares no pobres	78,28 %	0,70 %	0,69 %	-1,26 %
% Hogares pobres	21,72 %	0,70 %	0,66 %	-4,58 %
% Hogares en extrema pobreza	5,62 %	0,33 %	0,35 %	5,35 %
% Hogares no satisfacen necesidades bá	16,11 %	0,55 %	0,54 %	-1,01 %
PROMEDIOS				
Ingreso promedio ocupación principal	159.674,92	3.636,68	3.393,54	-6,69 %
Ingreso promedio del hogar	272.304,25	6.965,17	6.087,10	-12,61 %
Ingreso promedio percápita del hogar	83.480,42	2.346,58	2.175,94	-7,27 %
Nivel educativo promedio por ocupado	9,62	0,091	0,087	-4,31 %
Horas promedio trabajadas	46,29	0,191	0,187	-2,23 %
OTROS INDICADORES				
Miembros por hogar	3,82	0,02	0,02	0,63 %
Fuerza de trabajo por hogar	1,61	0,01	0,01	-0,89 %
Ocupado por hogar	1,51	0,01	0,01	-4,34 %
% Hogares con jefatura femenina	26,70 %	0,55 %	0,54 %	-2,39 %

Tras ordenar según la magnitud del coeficiente diferencial, se evidencia que los datos extremos son pocos. En el cuadro siguiente, los cambios de colores en el diagrama señalan diferencias en el coeficiente a intervalos de 5%.

Variable o Indicador	Coeficiente diferencial $\left[\frac{\text{Bootstrap}}{\text{Serie de Taylor}} - 1\right] * 100$	Diagrama de los Coeficientes
Desocupados	6,33	
% Hogares en extrema pobreza	5,35	
Inactivos (de 12 años y más)	4,70	
Fuerza de trabajo	2,05	
Menores de 12 años	1,29	
Tasa bruta de participación	1,09	
Tasa de ocupación	0,87	
Ocupados	0,85	
Miembros por hogar	0,63	
Fuerza de trabajo por hogar	-0,89	
% Hogares no satisfacen lo básico	-1,01	
% Hogares no pobres	-1,26	
Total Jefes de Hogar	-1,32	
Tasa neta de participación	-1,44	
Jefes inactivos	-1,90	
Jefes ocupados	-1,95	
Horas promedio trabajadas	-2,23	
% Hogares con jefatura femenina	-2,39	
Tasa de desempleo abierto	-2,57	
Jefes desocupados	-3,14	
Población Total	-4,30	
Nivel educativo promedio por ocupado	-4,31	
Ocupados por hogar	-4,34	
% Hogares pobres	-4,58	
Jefes en fuerza de trabajo	-6,48	
Ingreso promedio ocupación principal	-6,69	
Ingreso promedio percápita del hogar	-7,27	
Ingreso promedio del hogar	-12,61	



El diagrama muestra que son más los coeficientes negativos que los positivos. Esto quiere decir que los *errores estándares* estimados bajo el procedimiento clásico (series de Taylor) resultan mayores al *Bootstrap*, la mayoría de las veces.

Puesto que más de la mitad de las variables obtuvieron diferencias menores al 5% (en valor absoluto), señalado en el gráfico dentro de la elipse, se podría considerar esta cifra (positiva y negativa) como un tope de tolerancia en las diferencias encontradas. Definido así, se puede decir que menos de la cuarta parte de ellas (21,4%) presentan diferencias importantes. Las variables que componen esta minoría son:

Indicador	Error muestral clásico respecto al Bootstrap mayor a [-5, 5] (en valor absoluto)
Número de desocupados	Inferior
Porcentaje de hogares en extrema pobreza	Inferior
Jefes en fuerza de trabajo	Superior
Ingreso promedio en la ocupación principal	Superior
Ingreso promedio percápita del hogar	Superior
Ingreso promedio del hogar	Superior

Las variables que resultaron con errores muestrales estimados bajo el método clásico, superiores al Bootstrap, y que además obtuvieran las magnitudes más altas en términos absolutos, fueron todos los *Ingresos promedios* y los *jefes en fuerza de trabajo*.

En el otro extremo, una de las variables de mayor variabilidad histórica es la cantidad de desocupados, que precisamente obtiene el error estándar Bootstrap más alto respecto del clásico.

Las tres variables discretas [10/] ubicaron sus resultados dentro del rango de tolerancia definido anteriormente (de -5% a 5%). Era de esperar para estas variables, que su discrepancia fuera mayor, debido a su asimetría en la distribución.

[10/] *Miembros por hogar*, *Fuerza de trabajo por hogar*, y *Ocupado por hogar*.

El hecho de que 22 de los 28 indicadores o variables obtuvieran coeficientes inferiores al 5% (en valor absoluto) concuerda con lo señalado por Kish y Frankel (1970 y 1974) (mencionado en Lepkowski y Bowles [1996:3]) en cuanto a que, la aproximación de la Serie de Taylor y los métodos de replicación repetida, aunque no producen estimaciones idénticas del error de muestreo, la investigación empírica ha demostrado que las diferencias en los métodos para muchos estadísticos son pequeñas. La teoría no señala explícitamente la magnitud de una divergencia que conlleve a preocuparse por la existencia de distribuciones no normales. No obstante, el uso adicional de un procedimiento alternativo, como el Bootstrap, para la obtención de los errores muestrales, solventaría cualquier duda al respecto, y enriquecería el análisis respectivo.

Los resultados obtenidos de los intervalos de confianza [11/] no arrojan información suficiente para relacionar la magnitud de este coeficiente diferencial con el traslape de los intervalos de confianza correspondientes; no obstante, se reconoce que si el error cambia, también lo hacen los intervalos. Aun con el incumplimiento del supuesto de normalidad, el método clásico construye intervalos de confianza simétricos alrededor de la estimación del valor de la variable. El método Bootstrap, por el contrario, calcula intervalos que incorporan los sesgos producidos por la distorsión de la distribución asimétrica. Por ello, la prudencia indica que sería recomendable construir intervalos de confianza a través de ambos métodos, para que el usuario, conocedor de estos detalles, juzgue su precisión.

Como ya se mencionó, el número de remuestras necesario para llevar a cabo la construcción de intervalos de confianza se fija entre 1000 y 2000, para intentar garantizar una mejor estimación de las colas de la distribución [Cuesta y Herrero 1999:10]. Puesto que no se logró alcanzar al menos 1000 replicaciones, no se pudo investigar al respecto.

[11/] La orden de ejecución del comando fue de realizar 1000 repeticiones, pero solo se obtuvieron un poco más de 400, debido a limitaciones en el procesador del computador utilizado.

1.5 REFLEXIONES FINALES

Se obtuvieron más coeficientes diferenciales negativos que positivos. Esto quiere decir que el error estándar, medido a través del procedimiento Bootstrap, resultó inferior en más ocasiones, que el obtenido a través de aproximaciones de series de Taylor.

Al definir un rango de tolerancia del -5% al 5%, se obtuvo que menos de la cuarta parte de las variables (21,4%) registraron coeficientes diferenciales importantes; es decir, diferencias fuera de este rango.

Los indicadores cuyas mediciones desde el método clásico, resultaron más altos en términos absolutos, fueron los correspondientes a las variables de *Ingresos promedios* y los *Jefes en fuerza de trabajo*. Por el contrario, las variables cuyo error muestral Bootstrap resultó más alto, fueron la cantidad de *Desocupados* y el *Porcentaje de hogares en extrema pobreza*.

Las diferencias entre ambos métodos fueron, por lo tanto, mayoritariamente pequeñas, tal como lo señalan Kish y Frankel [12/]: Concretamente, de los 28 indicadores analizados, 22 obtuvieron coeficientes diferenciales inferiores al 5%, otros cuatro menos del 8%, y un caso extremo 12,6%, que corresponde precisamente al *ingreso promedio del hogar*. Estas pequeñas magnitudes en los coeficientes diferenciales de los indicadores mencionados como tales, permiten validar la precisión de las estimaciones realizadas por el INEC.

Aunque el Teorema del Límite Central justifica el uso del método clásico (aproximaciones de series de Taylor) para el cálculo de los errores estándares en este tipo de encuestas, debido a su gran tamaño, debe resaltarse el impacto del incumplimiento del supuesto de normalidad en la construcción de los intervalos de confianza. Por ejemplo, las

[12/] 1970 y 1974, mencionado en Lepkowski y Borles [1996].

distribuciones asimétricas de al menos las variables discretas [13/], que se pueden observar en el Anexo B, podrían conducir a mediciones de intervalos de confianza sesgados, si éstos son calculados por el método tradicional, ya que este método produce intervalos simétricos alrededor de la estimación del valor de la variable, cuando en la realidad, deberían desviarse también, hacia la asimetría de la distribución.

Es en estos casos, se justifica el uso de un procedimiento alternativo como el Bootstrap, para la estimación de los errores muestrales y sus respectivos intervalos de confianza. La falta de capacidad en el procesador utilizado, no permitió calcular los límites de confianza Bootstrap a partir de las especificaciones recomendadas de 1000 o más repeticiones, lo que impidió comparar con mayor detalle entre los límites obtenidos por ambos métodos.

Como el remuestreo utilizado por la técnica Bootstrap proporciona muestras específicas de un gran universo de ellas, es necesario realizar muchas de estas réplicas, y duplicar los diferentes ejercicios aquí planteados para obtener resultados concluyentes.

Como recomendación al INEC, se le sugiere que en la publicación de la EHPM, denominada "Principales resultados de la EHPM", el capítulo correspondiente a los *errores muestrales* sea trasladado luego de Aspectos Metodológicos. Aunque se ha realizado un esfuerzo continuo para divulgar en mayor grado, lo concerniente a estos errores, es pertinente cambiar la estructura de la publicación en el sentido señalado, por cuanto es de esperar que en las primeras páginas el usuario indague sobre lo relevante de la encuesta. Colocar los errores como uno de los primeros temas, es proporcionarle al usuario la información necesaria para "formarlo" en el tema de la precisión de los datos.

[13/] *Miembros por hogar*, *Fuerza de trabajo por hogar*, y *Ocupado por hogar*.

BIBLIOGRAFÍA

Cerviño López Santiago, 2004. **Estudio de la incertidumbre asociada a los métodos de evaluación de las poblaciones de peces**. Tesis doctoral, Universidad de Vigo. España. www.iim.csic.es/pesquerias/ficheros/Tesis%20Santiago%20Cervino.pdf

Cuesta Marcelino y Herrero Francisco J., 1999. **Introducción al Bootstrap**. Departamento de Psicología, Universidad de Oviedo. www.psico.uniovi.es/Dpto_Psicologia/metodos/tutor.9/Welcome.html Consultado el 22 de julio de 2007.

Díaz Vidal de Rada, 2004. **Problemas de representatividad en las encuestas con muestreos probabilísticas**. Universidad Pública de Navarra. España. www.ddd.uab.es/pub/papers/02102862n74p45.pdf

Gujarati Damodar N., 2004. **Econometría**. McGraw-Hill Interamericana, México.

INE, 2004 (Instituto Nacional de Estadística). **Buenas prácticas en la elaboración de Estadísticas Oficiales**. Instituto Nacional de Estadística, Madrid, España.

INEC, 1999 (Instituto Nacional de Estadística y Censos). **Diseño de la Muestra 1999, Encuesta de Hogares de Propósitos Múltiples (EHPM).** San José, Costa Rica.

INEC, 2004 (Instituto Nacional de Estadística y Censos). **Encuesta de Hogares de Propósitos Múltiples julio 2004. Principales Resultados**. Instituto Nacional de Estadística y Censos, San José, Costa Rica.

INEI, 2001 (Instituto Nacional de Estadística e Informática). **Guía para la Evaluación de Indicadores Sociales de las Encuestas de Hogares**. Dirección Técnica de Demografía a Indicadores Sociales (DES). Lima, Perú.

Kish Leslie, 1975. **Muestreo de Encuestas**. Editorial Trillas, México.

Lepkowski Jim and Bowles Judy, 1996. **Software para el cálculo de errores de muestreo en encuestas complejas**. The Survey Statistician, Nº 35, University of Michigan, Estados Unidos de América.

Lisinger Charles A., Warwick Donald P., 1985. **La encuesta por muestreo: Teoría y práctica**. Editorial Continental S.A., México.

Lohr Sharon L., 2000. **Muestreo: diseño y análisis**. Internacional Thomson Editores, México.

Mata Greenwood Adriana, 1985. **Análisis de variancia paramétrico y no paramétrico: una aplicación para determinar el efecto del informante en la Encuesta de Hogares.** Memoria de Seminario para optar al título de Licenciada en Estadística. Universidad de Costa Rica, Costa Rica.

Mirás Julio, 1985. **Elementos de muestreo para poblaciones finitas.** Instituto Nacional de Estadística, Madrid, España.

Núñez Sosa Olmer, 2003. **Método de "Bootstrap"**. Maestría de Estadística, Universidad de Costa Rica. Documento inédito. Costa Rica.

Quintana Ruiz Carlos, 1996. **Elementos de Inferencia Estadística.** Editorial Universidad de Costa Rica, San José, Costa Rica.

Sheskin David J., 2000. **Handbook of Parametric and Nonparametric Statistical Procedures.** Second Edition, Chapman & Hall/CRC. United States of America.

Siegel Sidney, 1976. **Estadística no paramétrica aplicada a las ciencias de la conducta .** Tercera reimpresión. Editorial Trillas. México.

SPSS, 2006 (Statistical Product and Service Solutions). **Muestras Complejas de SPSS 15.0**. SPSS Inc, Chicago, EEUU.

Stata, 2003. **Stata Survey Data. Reference Manual. Release 8**. Stata Press Publication, Stata Corporation, College Statio, Texas, United States of America.

ANEXOS

A. Cuadro resumen

Cuadro A 1.1
COSTA RICA: Diferentes estimaciones provenientes de la EHPM: comparaciones de las estimaciones de variabilidad
SEGÚN: Tipo de indicador
POR: Método de estimación
Julio 2004

Indicador	Estimación	Tamaño de la muestra	Aproximación series de Taylor (SPSS)		Replicaciones (Bootstrap-STATA)			Relación Error estándar: Bootstrap / Serie de Taylor	Diferencia relativa [Bootstrap / Serie de Taylor - 1] * 100
			Error estándar	Coeficiente de variación	Error estándar	Coeficiente de variación	Replicaciones		
TOTALES									
Población Total	4.178.755	43.779	94.416	2,26 %	90.359	2,16 %	449	0,957	-4,30 %
Fuerza de trabajo	1.768.759	17.929	40.246	2,28 %	41.071	2,32 %	454	1,021	2,05 %
Ocupado	1.653.879	16.784	38.153	2,31 %	38.476	2,33 %	450	1,008	0,85 %
Desocupado	114.880	1.145	5.098	4,44 %	5.421	4,72 %	450	1,063	6,33 %
Inactivos (de 12 años y más)	1.481.721	15.690	33.813	2,28 %	35.404	2,39 %	450	1,047	4,70 %
Menores de 12 años	928.275	10.160	28.390	3,06 %	28.756	3,10 %	450	1,013	1,29 %
Total Jefes de Hogar	1.095.345	11.366	22.672	2,07 %	22.372	2,04 %	408	0,987	-1,32 %
Fuerza de trabajo	851.357	8.873	20.605	2,42 %	19.269	2,26 %	461	0,935	-6,48 %
Ocupado	823.751	8.590	19.712	2,39 %	19.328	2,35 %	431	0,980	-1,95 %
Desocupado	27.606	283	2.286	8,28 %	2.214	8,02 %	431	0,969	-3,14 %
Inactivos (de 12 años y más)	243.988	2.493	6.687	2,74 %	6.559	2,69 %	431	0,981	-1,90 %
TASAS									
Tasa bruta de participación	42,33 %	43.779	0,33 %	0,78 %	0,33 %	0,79 %	455	1,011	1,09 %
Tasa neta de participación	54,42 %	33.619	0,38 %	0,70 %	0,37 %	0,69 %	453	0,986	-1,44 %
Tasa de ocupación	50,88 %	33.619	0,39 %	0,77 %	0,39 %	0,77 %	440	1,009	0,87 %
Tasa de desempleo abierto	6,49 %	17.929	0,26 %	3,94 %	0,25 %	3,84 %	451	0,974	-2,57 %

... continuación Cuadro A 1.1

Indicador	Estima-ción	Tamaño de la muestra	Aproximación serie de Taylor (SPSS)		Replicaciones (Bootstrap-STATA)			Relación Error estándar: Bootstrap / Serie de Taylor	Diferencia relativa [Bootstrap / Serie de Taylor - 1] * 100
			Error estándar	Coeficiente de variación	Error estándar	Coeficiente de variación	Replica-ciones		
PORCENTAJES									
% Hogares no pobres	78,28 %	10.041	0,70 %	0,89 %	0,69 %	0,88 %	457	0,987	-1,26 %
% Hogares pobres	21,72 %	10.041	0,70 %	3,20 %	0,66 %	3,05 %	460	0,954	-4,58 %
% Hogares en extrema pobreza	5,62 %	10.041	0,33 %	5,88 %	0,35 %	6,19 %	420	1,054	5,35 %
% Hogares que no satisfacen necesidades básicas	16,11 %	10.041	0,55 %	3,39 %	0,54 %	3,36 %	435	0,990	-1,01 %
PROMEDIOS									
Ingreso promedio ocupación principal	159.674,92	14.887	3.636,68	2,28 %	3.393,54	2,13 %	418	0,933	-6,69 %
Ingreso promedio del hogar	272.304,25	10.041	6.965,17	2,56 %	6.087,10	2,24 %	450	0,874	-12,61 %
Ingreso percápita del hogar	83.480,42	10.041	2.346,58	2,81 %	2.175,94	2,61 %	460	0,927	-7,27 %
Nivel educativo promedio por ocupado	9,620	16.675	0,091	0,94 %	0,087	0,90 %	461	0,957	-4,31%
Horas promedio trabajadas	46,294	16.564	0,191	0,41 %	0,187	0,40 %	433	0,978	-2,23 %
INDICADORES									
Miembros por hogar	3,82	11.366	0,02	0,61 %	0,02	0,61 %	446	1,006	0,63 %
Fuerza de trabajo por hogar	1,61	11.366	0,01	0,84 %	0,01	0,83 %	448	0,991	-0,89 %
Ocupado por hogar	1,51	11.366	0,01	0,90 %	0,01	0,86 %	453	0,957	-4,34 %
Porcentaje de hogares con jefatura femenina	26,70 %	11.366	0,55 %	2,05 %	0,54 %	2,00 %	455	0,976	-2,39 %

B. Estadísticas e histogramas de algunas variables

Cuadro A 1.2
COSTA RICA: Distribución de los hogares
POR: Número de miembros
SEGÚN: Variable de interés
Julio 2004

Cantidad de miembros por hogar	Frecuencia de hogares		
	Miembros por hogar	Fuerza de trabajo por hogar	Ocupados por hogar
Total	**11.366**	**11.366**	**11.366**
0	...	1.128	1.341
1	931	5.011	5.282
2	1.748	3.471	3.266
3	2.377	1.250	1.075
4	2.628	350	294
5	1.928	127	89
6	944	22	15
7	443	5	2
8	197	...	...
9	91	...	...
10	40	...	2
11	14	...	...
12	9	2	...
13	6	...	...
14	3	...	...
15	1	...	...
16	2	...	...
20	4	...	...

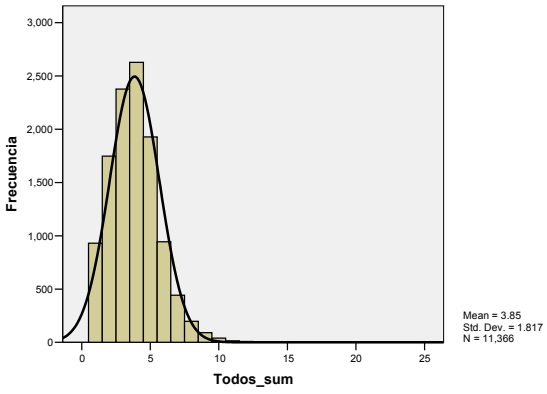
Fuente: Instituto Nacional de Estadística y Censos, Encuesta de Hogares de Propósitos Múltiples, julio 2004.

Cuadro A 1.3
COSTA RICA: Estadísticas
POR: Variable de interés
Julio 2004

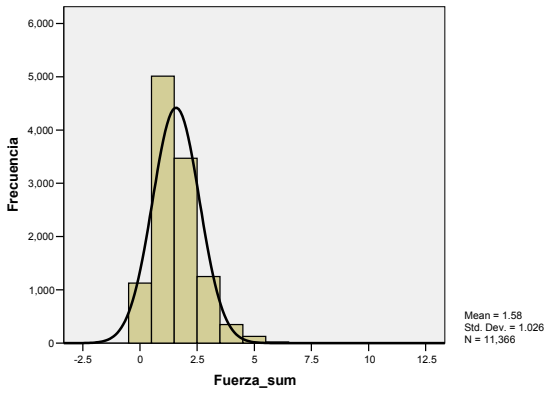
Estadísticas		Todos los miembros por hogar	Fuerza de trabajo por hogar	Ocupados por hogar
N	Válidos	11.366	11.366	11.366
	Perdidos	0	0	0
Media		3,85	1,58	1,48
Mediana		4,00	1,00	1,00
Moda		4,00	1,00	1,00
Desviación estándar		1,817	1,026	0,984
Variancia		3,303	1,053	0,968
Asimetría		0,986	1,1	1,012
Error estándar de la asimetría		0,023	0,023	0,023
Curtosis		3,543	3,243	2,354
Error estándar de la curtosis		0,046	0,046	0,046
Rango		19	12	10
Mínimo		1	0	0
Máximo		20	12	10
Percentiles	10	2	1	0
	20	2	1	1
	30	3	1	1
	40	3	1	1
	50	4	1	1
	60	4	2	2
	70	5	2	2
	80	5	2	2
	90	6	3	3

Fuente: Idem cuadro A 1.2.

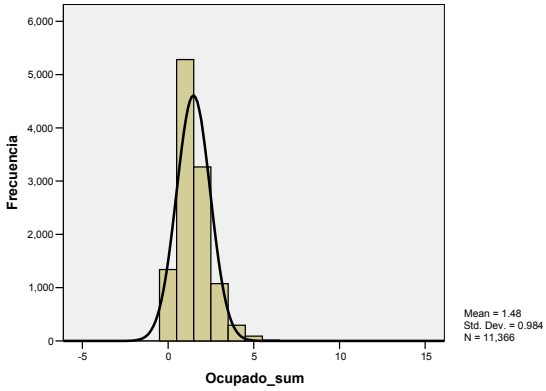
Histograma: Miembros por hogar



Histograma: Fuerza de trabajo por hogar



Histograma: Ocupados por hogar



C. Sintaxis de los Errores de Muestreo

i. Cálculo del error estándar bajo el método Bootstrap en Stata

Totales

```
set mem 800m

use "C:\Yadi\Bases\EH2004_Nivel.dta", clear
svyset [pweight=factor], strata(estrato) psu(consecu)
bootstrap `"svytotal todos"' _b, reps(1000) strata(estrato) cluster(consecu) bca
bootstrap `"svytotal fuerza"' _b, reps(1000) strata(estrato) cluster(consecu) bca
bootstrap `"svytotal todos, by(conda_re)"' _b, reps(1000) strata(estrato) cluster(consecu) bca

use "C:\Yadi\Bases\Jefes2004.dta", clear
svyset [pweight=factor], strata(estrato) psu(consecu)
bootstrap `"svytotal todos"' _b, reps(1000) strata(estrato) cluster(consecu) bca
bootstrap `"svytotal fuerza"' _b, reps(1000) strata(estrato) cluster(consecu) bca
bootstrap `"svytotal todos, by(conda_re)"' _b, reps(1000) strata(estrato) cluster(consecu) bca
```

Tasas

```
use "C:\Yadi\Bases\EH2004_Nivel.dta", clear
svyset [pweight=factor], strata(estrato) psu(consecu)
bootstrap `"svyratio fuerza todos"' _b, reps(1000) strata(estrato) cluster(consecu) bca
bootstrap `"svyratio fuerza de12_yms"' _b, reps(1000) strata(estrato) cluster(consecu) bca
bootstrap `"svyratio ocupado de12_yms"' _b, reps(1000) strata(estrato) cluster(consecu) bca
bootstrap `"svyratio desocup fuerza"' _b, reps(1000) strata(estrato) cluster(consecu) bca
```

Porcentajes

```
use "C:\Yadi\Bases\Hog_ingreso_conocido2004.dta", clear
svyset [pweight=factor], strata(estrato) psu(consecu)
bootstrap `"svyratio pobre_total todos"' _b, reps(1000) strata(estrato) cluster(consecu) bca
bootstrap `"svyratio ext_pobreza todos"' _b, reps(1000) strata(estrato) cluster(consecu) bca
bootstrap `"svyratio no_satisfacen_neces todos"' _b, reps(1000) strata(estrato) luster(consecu) bca
bootstrap `"svyratio no_pobre todos"' _b, reps(1000) strata(estrato) cluster(consecu) bca
```

Promedios

```
use "C:\Yadi\Bases\Ocu_IngresoPrinc2004.dta", clear
svyset [pweight=factor], strata(estrato) psu(consecu)
bootstrap `"svymean ingprinci"' _b, reps(1000) strata(estrato) cluster(consecu) bca

insheet using "C:\Yadi\Bases\Hog_ingreso_sinServDom_tab.txt", tab
svyset [pweight=factor], strata(estrato) psu(consecu)
bootstrap `"svymean  ingtothog"' _b, reps(1000) strata(estrato) cluster(consecu) bca

use "C:\Yadi\Bases\Hog_ingreso_percapita2004.dta", clear
svyset [pweight=factor], strata(estrato) psu(consecu)
bootstrap `"svymean ingperca"' _b, reps(1000) strata(estrato) cluster(consecu) bca
```

```
use "C:\Yadi\Bases\Educacion2004.dta", clear
svyset [pweight=factor], strata(estrato) psu(consecu)
bootstrap `"svymean educa_re"' _b, reps(1000) strata(estrato) cluster(consecu) bca

use "C:\Yadi\Bases\Horas2004.dta", clear
svyset [pweight=factor], strata(estrato) psu(consecu)
bootstrap `"svymean c22c"' _b, reps(1000) strata(estrato) cluster(consecu) bca
```

Indicadores

```
use "C:\Yadi\Bases\Hog_Condact2004.dta", clear
svyset [pweight=factor], strata(estrato) psu(consecu)
bootstrap `"svymean todos_sum"' _b, reps(1000) strata(estrato) cluster(consecu) bca
bootstrap `"svymean fuerza_sum"' _b, reps(1000) strata(estrato) cluster(consecu) bca
bootstrap `"svymean ocupado_sum"' _b, reps(1000) strata(estrato) cluster(consecu) bca

insheet using "C:\Yadi\Bases\Hog_Condact2004_tab.txt", tab
svyset [pweight=factor], strata(estrato) psu(consecu)
bootstrap `"svymean  jefe_mujer"' _b, reps(1000) strata(estrato) cluster(consecu) bca
```

ii. Cálculo del error estándar bajo el método tradicional en SPSS

```
* Preparación del plan de análisis, indispensable
     para el cálculo de errores muestrales *
* Analysis Preparation Wizard.
CSPLAN ANALYSIS
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /PLANVARS ANALYSISWEIGHT=factor
 /PRINT PLAN
 /DESIGN
     STRATA= Estrato
     CLUSTER= consecu
 /ESTIMATOR TYPE=WR.
```

Totales

```
*  Población total en  EH2004_Nivel.sav  *
* Complex Samples Frequencies.
CSTABULATE
 /PLAN FILE =
   'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /TABLES VARIABLES = Conda_re  Todos
   Fuerza  Ocupado  Desocup
Inactivo  Menor12años  De12_ymás
 /CELLS  POPSIZE
 /STATISTICS  SE CV CIN(95) COUNT DEFF
 /MISSING SCOPE = TABLE CLASSMISSING
= EXCLUDE.

*  Jefes de Hogar en  Je_Jefes2004.sav  *
* Complex Samples Frequencies.
CSTABULATE
 /PLAN FILE =
   'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /TABLES VARIABLES = Todos Fuerza
   Ocupado Desocup Inactivo Menor12años
   De12_ymás
 /CELLS  POPSIZE
 /STATISTICS  SE CV CIN(95) COUNT DEFF
 /MISSING SCOPE = TABLE CLASSMISSING
   = EXCLUDE.
```

Tasas

```
* Tasa Bruta de Participación *
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
   'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = Fuerza
   DENOMINATOR = Todos
```

```
 /STATISTICS SE CV COUNT POPSIZE DEFF
     CIN (95)
 /MISSING SCOPE = ANALYSIS
     CLASSMISSING = EXCLUDE.
```

**** Tasa Neta de Ocupación ****

```
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = Fuerza
    DENOMINATOR = De12_ymás
 /STATISTICS SE CV COUNT POPSIZE   DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.
```

**** Tasa de Ocupación ****

```
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = Ocupado
    DENOMINATOR = De12_ymás
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.
```

**** Tasa de Desempleo Abierto ****

```
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = Desocup
    DENOMINATOR = Fuerza
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.
```

Porcentajes

**** Con filtro: Hogar_ing_noServDom = 1 (FILTER) ****

```
USE ALL.
COMPUTE
   filter_$=(Hogar_ing_noServDom = 1).
VARIABLE LABEL
    filter_$ 'Hogar_ing_noServDom = 1
    (FILTER)'.
VALUE LABELS
    filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
```

**** Hogares pobres / Total de hogares ****

```
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = Pobre_total
    DENOMINATOR = Hogar_ing_noServDom
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.
```

**** Hogares extrema pobreza / Total de hogres ****

```
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = Ext_pobreza
    DENOMINATOR = Hogar_ing_noServDom
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.
```

**** Hogares no satisfacen necesidades / Total de hogares ****

```
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = No_satisfacen_neces
    DENOMINATOR = Hogar_ing_noServDom
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.
```

**** Hogares no pobres / Total de hogares ****

```
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = No_pobre
    DENOMINATOR = Hogar_ing_noServDom
```

```
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.
```

Promedios

```
* Ingreso promedio de la ocupación principal *
* Con filtro:  Ocupado = 1
      & (ingprinci > 0 & ingprinci < 99999999)
      (FILTER) *
USE ALL.
COMPUTE filter_$=(Ocupado = 1
    & (ingprinci > 0 & ingprinci < 99999999)).
VARIABLE LABEL filter_$ 'Ocupado = 1
    & (ingprinci > 0 & ingprinci < 99999999)
    (FILTER)'.
VALUE LABELS
    filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = ingprinci
    DENOMINATOR = Ocupado
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.

* Ingreso del hogar *
* Con filtro:  Hogar_ing_noServDom = 1
    & (ingtothog > 0 & ingtothog < 99999999)
    (FILTER) *
USE ALL.
COMPUTE
    filter_$=(Hogar_ing_noServDom = 1
      & (ingtothog > 0 & ingtothog < 99999999)).
VARIABLE LABEL
    filter_$ 'Hogar_ing_noServDom = 1
      & (ingtothog > 0 & ingtothog < 99999999)
      (FILTER)'.
VALUE LABELS
    filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
* Ingreso promedio del hogar *
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = ingtothog
    DENOMINATOR = Hogar_ing_noServDom
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.

* Ingreso percápita del hogar *
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = ingperca
    DENOMINATOR = Hogar_ing_noServDom
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.

* Nivel educativo promedio por ocupado *
* Con filtro:  Ocupado = 1
   & Educa_re  < 99 (FILTER) *
USE ALL.
COMPUTE filter_$=(Ocupado = 1
    & Educa_re < 99).
VARIABLE LABEL
    filter_$ 'Ocupado = 1 & Educa_re < 99
    (FILTER)'.
VALUE LABELS
    filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = Educa_re
    DENOMINATOR = Ocupado
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.
```

```
* Horas promedio trabajadas por ocupado *
* Con filtro:  Ocupado = 1
    & (c22c > 0 & c22c < 99)  (FILTER) *
USE ALL.
COMPUTE
    filter_$=(Ocupado =  1
    & (c22c > 0 & c22c < 99)).
VARIABLE LABEL
    filter_$ 'Ocupado =  1
     & (c22c > 0 & c22c < 99) (FILTER)'.
VALUE LABELS
    filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
   'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = c22c
   DENOMINATOR = Ocupado
 /STATISTICS SE CV COUNT POPSIZE DEFF
   CIN (95)
 /MISSING SCOPE = ANALYSIS
   CLASSMISSING = EXCLUDE.
```

Indicadores

```
* Miembros por hogar *
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = Todos_sum
    DENOMINATOR = Contador
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.
```

```
*  Fuerza de trabajo por hogar *
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = Fuerza_sum
    DENOMINATOR = Contador
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.
```

```
* Ocupado por hogar *
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = Ocupado_sum
    DENOMINATOR = Contador
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.
```

```
*Porcentaje de hogares con jefatura femenina *
COMPUTE Jefe_mujer = 0 .
EXECUTE .
IF (b04 = 2) Jefe_mujer = 1 .
EXECUTE .
* Complex Samples Ratios.
CSDESCRIPTIVES
 /PLAN FILE =
    'C:\Yadi\Diseño\Diseño_EHPM2004.csaplan'
 /RATIO NUMERATOR = Jefe_mujer
    DENOMINATOR = Contador
 /STATISTICS SE CV COUNT POPSIZE DEFF
    CIN (95)
 /MISSING SCOPE = ANALYSIS
    CLASSMISSING = EXCLUDE.
```

CAPÍTULO 2
ERRORES NO MUESTRALES

2.1 INTRODUCCIÓN

Los errores no muestrales [14/] son todos los errores de estimación que no resultan del diseño muestral operacionalizado. No necesariamente son pequeños, y contrario a los errores de muestreo, son difíciles de estimar, por lo que se desconoce su totalidad. Principalmente, están ligados a los controles de calidad ejercidos en la elaboración del cuestionario, el trabajo de campo, crítica, codificación, digitación, limpieza de datos y publicación.

La ejecución de todo el proceso de la encuesta debe seguir normas de control de calidad, pero esto no exime de la presencia de dichos errores. Implementar algún tipo de estimación de los errores no muestrales en algunas de las facetas, aún no medidas, conduciría a obtener mejores estimaciones de la calidad de la información.

El INEC se ha esforzado por minimizar una serie de errores no muestrales en las etapas de campo, crítica, codificación y limpieza. Entre ellos, se puede mencionar la supervisión en todas las etapas y la implementación de un listado de inconsistencias para la limpieza, además de un análisis exhaustivo de las frecuencias de todas las variables.

Como la EHPM tiene la característica de sustituir un 25% de la muestra cada año, existe un porcentaje alto de la muestra de viviendas y personas que empatan de un año a otro [15/]. Por ello, el INEC ha vinculado electrónicamente un juego de encuestas, de tal forma que pudo empatar un grupo de viviendas entre encuestas consecutivas desde el 2000 hasta el 2005, lo que permite determinar, para cada par de años, una muestra de viviendas

[14/] Denominados así en el manual de la ONU [1983].
[15/] Ese porcentaje no llega al 75% debido a que el uso de las viviendas cambian de rama de actividad (de hogar a comercio, a industria, etc.), no responden, o cambian los residentes habituales de las viviendas.

tipo panel. Mediante un procedimiento de programación, se pudo identificar cuáles de esas viviendas representaban los mismos núcleos familiares (hogares) de la entrevista realizada un año antes. De esta manera, fue posible obtener una de esas muestras para los últimos dos años del período (2004-2005) que interesa a esta investigación.

Es posible estimar algunos potenciales errores no muestrales con esta información, tras responder a la pregunta: ¿hay consistencia en la información obtenida, entre 2004 y 2005, de los hogares y sus residentes? Esto con la certeza de que las diferencias observadas podrían provenir principalmente de tres fuentes: del entrevistado, del entrevistador y del propio procedimiento de empate.

Los objetivos planteados seguidamente abarcan un conjunto de variables extraídas de las encuestas 2004 y 2005.

Objetivo general

Estimar la magnitud del error no muestral en un grupo de variables seleccionadas, en mediciones sucesivas entre 2004 y 2005, mediante técnicas de regresión.

Objetivos específicos

- Totalizar los errores no muestrales en un grupo seleccionado de variables, mediante la generación de variables de error.
- Determinar el impacto del tipo de informante en la generación de errores, mediante la creación de variables específicas.
- Focalizar la atención sobre algunas variables de control que podrían incidir en la presencia de errores, mediante la aplicación de regresiones.

2.2 REFERENTE TEÓRICO

Los errores no muestrales son más difíciles de controlar y no se pueden atribuir a la variabilidad entre las muestras, por lo que pueden presentarse tanto en encuestas como en censos. Se pueden definir, en principio, como una categoría residual; o sea, como todos los errores de estimación que no son resultado del muestreo, y que no necesariamente son pequeños o carentes de importancia. En numerosas encuestas, la contribución del error no muestral al error total, excede a la del error muestral. Éstos tienen diferentes fuentes, pueden producirse en cualquier etapa de la encuesta y a menudo, su detección y control es sumamente difícil [ONU, 1983:5].

Los errores no muestrales pueden clasificarse desde diferentes perspectivas. Un enfoque, los clasifica de acuerdo con la etapa de la encuesta en la que se producen, esto es, en sus tres fases principales, como a continuación se desarrolla.

2.2.1 Diseño y preparación de la encuesta

Esta etapa abarca aspectos como: la planificación general, la selección de tópicos por incluir en la encuesta, el diseño y la estructura de éstas, la selección de la muestra, decisiones sobre la recolección de los datos, el desarrollo de los cuestionarios, las pruebas preliminares y la selección y capacitación de los entrevistadores. En cualquiera de estas actividades los errores pueden surgir si no se siguen controles adecuados. Algunos ejemplos de estos errores son:

- Longitud y complejidad inadecuadas del cuestionario.
- Conceptos claves sin una clara definición.
- Ausencia de un esbozo aproximado de las tabulaciones.
- Falta de claridad acerca de los grupos que comprenden la población objetivo. A este error se le denomina sesgo de selección, y ocurre cuando alguna parte de la población objetivo no se incluye en la población muestreada.

- Indefinición de las cargas de trabajo de cada enumerador y supervisor.
- Calidad cuestionable de los mapas con que se cuenta.
- Falta de definición de las unidades de muestreo final, así como las reglas de asociación que vinculan las unidades de observación con las primeras.
- Tiempo insuficiente destinado a las actividades de esta etapa.
- Escasa capacitación de entrevistadores, así como de los aspectos logísticos.
- Falta de claridad acerca de las normas que deben seguir quienes responden a los cuestionarios.
- Carencia de manuales.

2.2.2 Recolección de datos

Esta etapa se refiere a la organización y ejecución del trabajo de campo para la captura de datos con el cuestionario correspondiente. Aunque puede pensarse que la fuente del error ocurre solamente durante la recolección de los datos, frecuentemente es el resultado de decisiones o selecciones deficientes en la fase de diseño y preparación de la encuesta. Específicamente, en esta etapa los errores pueden clasificarse como: errores de cobertura, falta de respuesta y errores de respuesta.

2.2.2.1 Errores de cobertura [ONU, 1983:25-36]

La falta de cobertura se produce cuando se excluyen algunas unidades de observación, ya sea en forma directa o implícitamente en el marco de muestreo operacional. Este es un caso especial, aunque común, de errores resultantes de defectos del marco muestral. Puesto que las unidades excluidas tienen una probabilidad cero de ser seleccionadas para integrar la muestra, en realidad quedan eliminadas de los resultados de la encuesta. Estos errores deben distinguirse de la exclusión deliberada y explícita de sectores de la población objetivo definida. La falta de cobertura se refiere al error negativo

de exclusión de ciertas unidades en el marco utilizado para la selección de la muestra, tal como lo muestra el siguiente diagrama.

Diagrama: Errores de cobertura

Así como ciertas unidades pueden carecer de representación en el marco, otras pueden aparecer en él más de una vez, lo que implica una posibilidad mayor que la prevista de ser seleccionadas para integrar la muestra. *A este tipo de error se le puede llamar exceso de cobertura o duplicación. Aun cuando el marco esté completo y no tenga duplicación, la falta de cobertura y el exceso de ella pueden producirse por el empleo de reglas de asociación defectuosas* [16/]*, o por una aplicación incorrecta de reglas técnicamente adecuadas* [ONU, 1983:27].

En marcos de encuestas de hogares correspondientes a diseños polietápicos, es necesario que las unidades de selección en cada etapa no se superpongan (sin traslapes) y sean especialmente identificables sobre el terreno. Los errores de cobertura surgen porque se viola algunas de estas condiciones, o varias de ellas.

---

[16]/ "*En las encuestas de hogares las operaciones de muestreo comprenden típicamente una serie de "etapas", que pasan digamos, de unidades de área mayores a segmentos menores y, finalmente, a viviendas. En cada etapa es necesario vincular las unidades comprendidas por medio de reglas de asociación; por ejemplo, que especifiquen qué viviendas se incluyen en un segmento seleccionado"* [ONU,:1983:25].

Otros errores que pueden ocurrir serían: la aplicación incorrecta de procedimientos de muestreo, falta de reglas claras de asociación entre unidades de muestreo y unidades de observación (viviendas, hogares y personas) y de su correcta aplicación.

2.2.2.2 Falta de respuesta [ONU, 1983:57-69]

La falta de respuesta surge cuando en los hogares y otras unidades de observación seleccionadas para la encuesta no se logra que proporcionen todos los datos que debían recogerse. Esta falta de obtención de resultados completos de todas las unidades seleccionadas, puede originarse en varias fuentes distintas, según la situación en que se realiza la encuesta. Por ejemplo, en una encuesta de hogares los entrevistadores pueden no ubicar uno de los hogares seleccionados, o el respondiente puede no encontrarse en el hogar cuando llega el entrevistador; o quizá, no esté dispuesto a responder o no pueda hacerlo; también, los cuestionarios llenos pueden extraviarse.

La falta de respuesta puede ser total o parcial. La total se produce cuando no se recoge ningún dato de una unidad de la muestra. La parcial o de rubros ocurre cuando una unidad no proporciona algunos puntos específicos de información o se niega a hacerlo. La falta de respuesta total es relativamente fácil de definir en principio, y muy perceptible en la práctica. Las fuentes más frecuentes de este tipo de error son: falta de acceso a las unidades de la muestra, falta de contacto o cooperación de los informantes y exceso de carga de preguntas.

2.2.2.3 Errores de respuesta

Este tipo de error se produce cuando se obtiene información, pero ésta es incorrecta. Se supone que cada individuo cubierto por una encuesta específica, tiene un valor cierto individual, totalmente independiente de la misma encuesta, de la redacción y forma de la pregunta y de quién haga la pregunta. Según este supuesto, el objetivo de la encuesta es

tratar de acercarse a ese valor; sin embargo, el éxito en dicho esfuerzo va a depender de varios elementos: naturaleza y forma de la pregunta, y cómo se plantea ésta [Kalton y Moser, 1975:378].

A diferencia de los *errores de muestreo*, los *errores de respuesta* no aplican solamente a encuestas por muestreo, sino que pueden surgir en toda la población.

a. Tipos de errores de respuesta

Aunque la clasificación no es exhaustiva, los errores de respuesta pueden clasificarse de la siguiente manera [Kalton y Moser, 1975:379]:

- **Errores no sistemáticos del entrevistador:** un entrevistador, debido a su negligencia, podría realizar un mal registro de las respuestas, lo que provocaría errores. Estos no son sistemáticos, porque en una ocasión podría sub-registrar y en la otra sobre-registrar, y al final, el error neto podría ser pequeño.
- **Sesgos del entrevistador:** el entrevistador podría realizar la pregunta ó interpretar la respuesta dada, de acuerdo con sus propias creencias o percepciones, lo que influencia las declaraciones dadas. Para este entrevistador en particular, tales errores podrían no cancelarse unos a otros, lo que provocaría una acumulación de ellos. Por lo tanto, el resultado de sus entrevistas estaría sesgado, de tal manera que si él fuera un entrevistador en un infinito número de encuestas en esta población, el promedio (valor esperado) de sus resultados podrían diferir del valor promedio cierto de la población. La diferencia entre ambos valores constituiría su sesgo neto.
- **Errores sistemáticos del informante:** un informante podría negarse sistemáticamente a dar información veraz sobre algún aspecto investigado en la encuesta. Por ejemplo, si se le preguntara ¿cuántas veces por semana asiste a un bar?, él podría dar, sistemáticamente, una cifra que no corresponde con la realidad.

Existe diferencia entre una variación de respuesta y un sesgo de respuesta. Para reconocerla, hay que suponer que una encuesta puede tratarse como si fuera conceptualmente un ensayo, repetible un gran número de veces; y en los cuales, se emplea el mismo procedimiento de encuesta, bajo las mismas condiciones esenciales, pensados como una relación en el mismo instante y sin ningún tipo de influencia de un ensayo sobre otro. Estos supuestos generarían un modelo en el cual un individuo puede dar iguales o diferentes respuestas en diferentes ensayos. De acuerdo con este esquema, una encuesta en ejecución sería considerada como la realización de un ensayo particular de un conjunto, e involucraría una colección de respuestas de una muestra seleccionada de individuos. La variación de respuesta se consideraría como el cambio en los resultados de la encuesta en repetidas pruebas.

$$y_{it} = u'_i + d_{it}$$

donde: y_{it} Valor observado de la variable para el individuo i en el ensayo t

u'_i promedio de los y_{it} sobre todos los ensayos

d_{it} desviación de respuesta del valor observado para el individuo i en el ensayo t de este valor promedio

Para clarificar la distinción entre variación y sesgo de respuesta, se toma como ejemplo un censo completo para medir el promedio de edad de una población. En ensayos conceptualmente repetibles, un individuo podría declarar su edad con diferentes cifras y el cambio sería su variación de respuesta individual. El promedio de sus respuestas diferentes podría ser o no, su edad verdadera. La diferencia entre su promedio y su edad verdadera, sería su sesgo de respuesta individual. Los promedios de los sesgos y variaciones de respuesta individuales, sobre la población total, son los sesgos y variaciones de respuesta, respectivamente.

La distinción entre sesgo de respuesta y variación de respuesta, es importante porque:

- Los efectos del sesgo permanecen constantes para cualquier tamaño de muestra, mientras que el efecto de la variación disminuye conforme el tamaño de la muestra aumenta, aunque no necesariamente de manera proporcional.
- El sesgo y la variación surten diferentes efectos en distintos estadísticos: por ejemplo, en la media aritmética, basada en una muestra grande, el sesgo es probablemente de mayor importancia. Contrariamente, para una diferencia entre dos promedios, podría esperarse, razonablemente, que los sesgos se cancelen, lo que convierte la variación en un problema más serio.
- Con diseños muestrales apropiados, las variaciones pueden ser medidas a través de la misma muestra; mientras que los sesgos pueden ser evaluados solamente por referencia de datos de algunas fuentes externas.

b. Fuentes de los errores de respuesta **[Kalton y Moser, 1975:385-388]**

- **Opiniones de los entrevistadores**

Si las opiniones de los entrevistadores fueran evidentes por la forma en que formulan la pregunta, algunos informantes deberían indudablemente tender a estar de acuerdo o en desacuerdo, con ellos. Igualmente, si las opiniones de los entrevistadores influenciaran su forma de interpretar y codificar respuestas dudosas o, si parafrasean las respuestas al iniciar las preguntas, podría resultar en un sesgo.

- **Errores debidos a las características de los entrevistadores**

Las características personales de un entrevistador, su sexo, edad, educación, y clase social, podrían influenciar las respuestas obtenidas, tanto por la impresión que provoque en el informante, como por la manera en que realice las preguntas, o quizás porque estos podrían dar respuestas con mejor o peor disposición ante los diferentes tipos de entrevistadores.

De igual manera, las expectativas de los entrevistadores pueden ser también fuente de error. En este caso, se señalan tres tipos de errores:

- **Estructura de actitudes esperadas:** el entrevistador, quizás inconscientemente, espera de las personas consistencia en sus actitudes, por lo que podría, luego de la entrevista, interpretar las respuestas dadas a la luz de esas expectativas.
- **Roles esperados:** durante la entrevista, el entrevistador adquiere una impresión del tipo de persona a la que está entrevistando (edad, clase social, ocupación, ingreso, personalidad, etc.) y, posteriormente, ante dudas o respuestas ambiguas, él podría interpretarlas con base en lo que él esperaría de esta persona.
- **Expectativas probabilísticas:** un entrevistador esperaría una cierta distribución de opiniones o características de todos sus informantes. Por ejemplo, el entrevistador podría pensar que cerca de la mitad del total de la muestra debería estar ocupada, y si su porción de entrevistas no cumple con este porcentaje, podría interpretar que sus respuestas tienen errores y, en un caso extremo, distorsionar la información en ese sentido.

- **Errores debidos a los informantes**

  El informante podría dar una respuesta diferente a la correcta, ya sea:
  - porque le falta conocimiento para darla;
  - porque su memoria le falle;
  - porque no entiende bien la pregunta; o
  - porque, conciente o inconcientemente, él no desea dar la respuesta correcta, como sucede cuando:
    - un informante exagera la frecuencia con la que visita la iglesia para aparentar ser muy respetable ante los ojos del entrevistador;
    - o bien, él podría dar una falsa imagen de su opinión política, a favor de aquél a quien el entrevistador favorece;

- o también, él podría declarar sus sentimientos a favor de la opinión del entrevistador, porque desea concluir la entrevista, etc.

En particular, para la EHPM este tipo de error presenta una dimensión más amplia, por cuanto generalmente, una sola persona, la gran mayoría de las veces miembro del hogar, es quien responde por sus propios datos y por los de los otros. Se reconoce entonces que, al existir un informante que proporciona datos de otros, la fuente de error se deba a un conocimiento no certero de la información requerida acerca de los otros residentes del hogar.

Todos estos resultados son un riesgo latente: lo que las personas dicen y cómo lo dicen, varía de acuerdo con las circunstancias y con los interlocutores.

2.2.3 Procesamiento y análisis de datos

Esta etapa comprende actividades tales como: la recepción de los cuestionarios, la crítica, codificación y verificación de las respuestas, la detección de inconsistencias, la entrada de los datos al sistema (errores de digitación), el procesamiento electrónico de los datos y la estimación y tabulación de ellos, así como la difusión de los resultados.

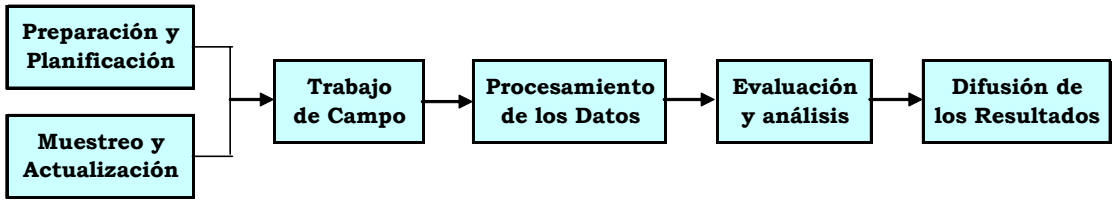
Algunos ejemplos de errores podrían ser:

- Recepción desordenada y sin registro de los cuestionarios.
- Identificación incorrecta de los cuestionarios.
- Capacitación inadecuada de los codificadores.
- Carencia de manuales para codificadores.
- Falta de un plan de control y verificación de calidad del trabajo de los codificadores.
- Software inadecuado para el ingreso de los datos.
- Deficiente digitación de los datos.
- Falta de estándares en procedimientos de imputación.
- Omisión de la revisión de las frecuencias de las variables.

2.2.4 Situación actual

Los errores no muestrales son difíciles de medir y detectar, lo que exige la aplicación de diversos controles. Para minimizarlos, el INEC ha establecido en cada etapa de la encuesta diferentes procedimientos, con el objetivo de mejorar la calidad de los datos ofrecidos a los usuarios.

Las actividades de la EHPM se pueden resumir en cinco etapas [INEC, 2005a:198]:



En cada una de estas fases se han establecido controles para minimizar los posibles errores que no provienen del diseño muestral.

2.2.4.1 Preparación, planificación, muestreo y actualización

El personal involucrado en las etapas de preparación, planificación, muestreo y actualización es altamente calificado, y con una amplia experiencia. Las revisiones son constantes, y con frecuencia, las dudas no solo se discuten al interior de la oficina encargada, sino que se acude a expertos nacionales e internacionales.

2.2.4.2 Trabajo de campo

En el trabajo de campo se capacita internamente a los supervisores y entrevistadores participantes en la encuesta. El trabajo se organiza en dos etapas [INEC, op cit]: la primera en el Gran Área Metropolitana, y la segunda, en el resto del país.

Durante la recopilación de datos, el trabajo de los enumeradores es supervisado continuamente por los supervisores de equipo, y ocasionalmente, por la supervisión general. Además, el supervisor de equipo tiene a su cargo la actualización de los mapas y del Registro de Edificios y Viviendas, lo que permitirá una mejor labor para la próxima encuesta.

Paralelamente, se reentrevista a un determinado número de hogares, cuyo cuestionario incluye un conjunto más reducido de preguntas. El objetivo consiste en detectar diferencias en las respuestas que servirán tanto para orientar el trabajo de campo, como para tener una visión general de la calidad de la recolección de los datos.

Para la EHPM de julio 2004, se desarrolló por primera vez el indicador "*Índice de Consistencia Global*". Este indicador sirve para comparar la calidad general de las distintas características evaluadas, mediante su definición para una característica determinada. El índice consiste en sumar todas las coincidencias entre entrevista y re-entrevista observadas, expresadas como porcentaje del total de las reentrevistas para esa característica en particular.

$$ICG\,(C) \quad = \quad \frac{\sum_i n_{ii}}{n} \; * \; 100$$

Los indicadores que se obtienen para cada característica de estudio, permiten realizar comparaciones entre ellas. Por ejemplo, en 2004, el *ICG* para la *Rama de Actividad* y la *Ocupación* fue superior al 80%; para la *Categoría Ocupacional* fue del 75%, mientras que para las preguntas referentes a *Ingreso*, osciló alrededor del 50% [Alvarado 2004:10-11].

2.2.4.3 Procesamiento de datos [INEC, 2005b:200-201]

Dentro de la calendarización del INEC, el procesamiento de datos es una de las etapas más extensas, por cuanto comprende varias subetapas:

Crítica y codificación: La recepción de los cartapacios con el material de trabajo es registrada cuidadosamente. De este material se extraen las entrevistas que quedaron "pendientes" para una posible visita. El trabajo de crítica y codificación se organiza por "lotes de trabajo" en estricto orden geográfico.

Para el módulo básico de la EHPM [17/], se siguen fases de revisión por grupos de variables. Posteriormente, se codifican y se pasa a verificar las etapas anteriores. Al mismo tiempo, se procura obtener información faltante por medio de llamadas telefónicas y el envío de cuestionarios sin respuesta, a través de equipos de "rescate de entrevistas pendientes".

No existe un procedimiento para imputar datos faltantes o erróneos, salvo los criterios que se utilizan para criticar y codificar los cuestionarios. Estos se basan en asignar códigos especiales prefijados a ciertas situaciones, incluyendo el código de ignorado. No obstante, en muchos casos, se realiza un tipo de "imputación deductiva"; es decir, se estudia el resto de la información y se deduce cuál es el dato más idóneo para completar la respuesta faltante.

Validación y consistencia de los datos: El proceso de detección de inconsistencias se realiza electrónicamente con el software CS-PRO, previamente programado para ello. Este paquete genera listados de posibles inconsistencias que son revisados por ocho grupos de trabajo [18/]. Dichos grupos tienen a su cargo también la limpieza de los datos; esto es, la

---

[17/] Es común que la EHPM incluya módulos adicionales al básico de empleo que siempre se realiza, por lo que las fases se incrementan conforme se incluyan módulos no básicos.

[18/] Para que este proceso concluya en cuatro semanas, tal y como está programado, se requiere que la carga de trabajo se distribuya en ocho grupos, conformados cada uno por dos supervisores del personal de planta.

corroboración directa en el cuestionario, las correcciones e indagaciones pertinentes, tanto en el instrumento como en la base de datos.

Revisión de las distribuciones de frecuencias: Una vez que se tiene la base de datos, se procesa una serie de tabulados con las frecuencias de todas las variables. Esto permite revisar que todas las cifras sean consistentes con el resto de la información. En términos generales, se trata de listar todos aquellos casos que no concuerden con la lógica del cuestionario, o bien, con situaciones normales, para detectar la fuente del error y realizar la respectiva corrección.

2.2.4.4 Evaluación y análisis

Una vez que se obtienen las primeras estimaciones, se verifican las estructuras de empleo con los datos de cuentas nacionales y con encuestas anteriores. La revisión de los datos se efectúa comparando cifras globales y observando qué tan coherente es la tendencia histórica. En aquellos casos en los que se presentan "saltos bruscos" o fuera de lo normal, se estudian las regiones con transformaciones más significativas, para canalizar las investigaciones hacia posibles respuestas del comportamiento de las cifras. Algunas fuentes que ayudan a entender mejor la situación, provienen de entes externos como el Banco Central de Costa Rica, Instituto de Investigaciones en Ciencias Económicas, Municipalidades, Caja Costarricense de Seguro Social, etc.

2.3 METODOLOGÍA

La amplia gama de fuentes y ocurrencias de los errores no muestrales, hacen que la detección de ellos sea difícil "per se". Las Naciones Unidas ha mencionado que en muchas encuestas, la contribución del error no muestral al error total, excede a la del error muestral [ONU, 1983:5], por lo que el esfuerzo de detección de ellos implicaría definir un diseño experimental que escapa al objetivo de este estudio. Esta medición de los errores no muestrales se torna en exceso difícil, si además se agrega la intención de medirlos sobre una encuesta ya realizada. No obstante, este estudio se concentró en estimar errores no muestrales relativos a la comparación de ciertas variables en dos EHPM consecutivas (2004-2005), referidas a personas y hogares provenientes de viviendas panel, a sabiendas de que los errores detectados podrían provenir de tres fuentes: entrevistado, entrevistador, y el propio procedimiento de empate.

Como ya se ha mencionado, el INEC realizó, para varias encuestas, un ejercicio de empate de viviendas sobre las tres cuartas partes de la muestra que se mantiene invariable de una encuesta a otra. Para lograrlo, una vez establecidos los segmentos sin cambios importantes, se identificaron las viviendas que permanecieron seleccionadas durante los tres años de encuesta [19/], y a partir de ahí se aplicó un procedimiento electrónico con ciertos criterios de empate, para determinar si se mantenían las mismas personas durante los tres años. *Según estos criterios, alrededor de un 20% de viviendas y un 16% de personas empataron cada año* [INEC, 2005b:14]. Para el presente trabajo, se solicitó únicamente el empate [20/] 2004-2005, conformado por 3.178 [21/] hogares que comprendían 11.061 registros de personas. Las variables propias de personas (*Sexo*, *Edad*, etc.), estuvieron asociadas a variables del hogar (*Tenencia de vivienda*, *Material de paredes*, etc.).

[19/] Una vivienda permanece en la muestra durante cuatro años consecutivos, antes de ser sustituida por otra.

[20/] Empatar: sinónimo de empalmar (en este contexto), que significa ligar o combinar planes, ideas, acciones, etc. (Real Academia Española www.rae.es).

[21/] Se eliminaron los hogares empatados que tuvieron ignorados en las variables que sirven de primer ingreso a los modelos de regresión empleados más adelante.

La medición de los errores no muestrales se llevó a cabo mediante la observación de diferencias en las respuestas obtenidas en esos dos años consecutivos. Para conseguirlas, se trabajó exclusivamente con las secciones *V* (*Vivienda y servicios*) y *B* (*Características sociodemográficas*) de la EHPM, las cuales proporcionan algunas variables que pueden permanecer estables en el transcurso de un año, o que, aun sufriendo cambios, podrían predecirse.

La información proveniente de la base de datos empatada, permitía reconocer si el informante del primer año (2004) era el mismo del siguiente (2005), con lo cual se creó una nueva variable denominada "*Tipo_informante*", con dos posibles respuestas: *Mismo informante* y *Diferente informante*. Este informante fue la persona que suministró los datos sobre las características del hogar, sobre sí mismo y sobre los otros miembros del grupo familiar en ambos años. Particularmente, cuando esta persona brindó la información sobre sí misma –en ambos años-, pasó a ser "*Mismo_autoinformante*". Esta es una variable creada separadamente para diferenciar al individuo que respondió acerca de las variables de índole personal.

Por ejemplo, en un hogar conformado por dos miembros, José y María, al comparar ambos años respecto de quién dio la información, se podrían obtener diferentes tipos de informantes, tal y como se aprecia en el diagrama siguiente:

Para un mismo hogar con dos miembros: María y José				
	La información sobre	En la encuesta **2004** la información la proporciona	En la encuesta **2005** la información la proporciona	Tipo de informante para ambos años
Situación 1	José	María	María	Mismo Informante, pero no Mismo Autoinformante
	María	María	María	Mismo Autoinformante
Situación 2	José	María	José	Diferente informante
	María	María	José	Diferente informante
Situación 3	José	José	María	Diferente informante
	María	José	María	Diferente informante

Esta nueva variable "*Tipo_informante*", es utilizada posteriormente en los modelos de regresión. En el caso de las variables de *Vivienda*, solo se pueden diferenciar por *Mismo*

Informante o *Diferente Informante*, ya que la información se refiere al estado de la vivienda y puede ser suministrada por cualquier miembro de hogar, quien podría ser el mismo del siguiente año.

Para cada persona que logró empatarse dentro de la base de datos, independientemente del tipo de informante, se registraron las coincidencias -y las no coincidencias- en ambos años, de algunas características básicas o importantes de la vivienda y de los miembros del hogar. Para conseguirlo, se seleccionaron, en cada una de las dos secciones de la encuesta (*Vivienda* y *Sociodemográfica*), las variables que podrían indicar estas inconsistencias en las respuestas dadas, para lo que se denominó con cero (0) la no inconsistencia y con uno (1) la inconsistencia o error; posteriormente, se agregó como una suma de errores. Así, se escogieron 12 variables dicotómicas: cinco referentes a características de personas y siete sobre características de las viviendas.

Con esas variables se contrastaron los rangos de variación aceptables entre los dos años. Por ejemplo, al considerar la variable *Edad*, se supondría que ésta debería reportarse como con un año más para cada persona en una encuesta posterior. Sin embargo, podría ocurrir que una persona cumpliera 20 años el 14 de julio y fuera entrevistada el 12 de ese mes, cuando aún tenía 19 años; al año siguiente, podría haber sido visitada el 17 de julio, cuando tuviera 21 años cumplidos, por lo que entre una encuesta y otra, esa misma persona tendría una diferencia de dos años. De esta manera, para cada variable seleccionada, y para todos los registros, se construyeron los siguientes criterios (donde 1 significa error -inconsistencia- y 0 significa ausencia de él):

Criterios para identificar errores en la Sección Sociodemográfica		
Variable	**Valores**	**Criterios**
Sexo	0	Si *Sexo* en 2005 = *Sexo* en 2004
	1	Si *Sexo* en 2005 es diferente al *Sexo* en 2004
Edad	0	Si *Edad* en 2005 = *Edad* en 2004 ó Si *Edad* en 2005 = *Edad* en 2004 + 1 ó Si *Edad* en 2005 = *Edad* en 2004 + 2 ó Si alguno de los dos datos tiene código de ignorado
	1	Todos los demás casos
Lugar de nacimiento	0	Si *Lugar de nacimiento* en 2005 = *Lugar de nacimiento* en 2004 ó Si alguno de los dos datos tiene código de ignorado
	1	Todos los demás casos
Estado conyugal	1	Si *Estado conyugal* en 2005 = *Soltero* y en el 2004 tiene otro estado conyugal
	0	Si alguno de los dos datos tiene código de ignorado ó Todos los demás casos
Nivel de educación	0	Si *Nivel educativo* en 2005 = *Nivel educativo* en 2004 ó Si *Nivel educativo* en 2005 avanzó un nivel respecto al del 2004 ó Si hubo traslados entre *Secundaria académica* y *Secundaria técnica*, al mismo nivel o con avance de un año ó Si en el 2004 era menor a 5 años ó Si alguno de los dos datos tiene código de ignorado
	1	Todos los demás casos

Los criterios utilizados para definir un error en cada variable seleccionada de la sección Vivienda y Servicios no son del todo incuestionables. Si bien es cierto que para cada uno de los criterios seleccionados y presentados a continuación, se podría encontrar una situación extrema que afectara la vivienda y en la cual no se provocara un error, también es cierto el hecho de que, la probabilidad de que esas situaciones extremas (no

pago de casa, incendio, cambio voluntario de piso, contaminación de acueducto, etc.) se produzcan, en uno o algunos de los casos registrados, es muy baja [22]/. Teniendo en cuenta esta limitación y en aras de realizar un ejercicio académico más exhaustivo, se presentan a continuación la definición de los criterios para identificar errores en esas variables.

Criterios para identificar errores en la Sección Vivienda y Servicios		
Variable	**Valores**	**Criterios**
Tenencia de Vivienda	1	Si en 2004 es *Propia totalmente pagada* ó *Propia pagando a plazos* y en el 2005 es *Alquilada* ó *Precario* u *Otra*
	0	Si alguno de los dos datos tiene código de ignorado ó Todos los demás casos
Material predominante en Paredes Exteriores	1	Si en 2004 es *Block o ladrillo* ó *Zócalo* ó *Prefabricado* y en 2005 es *Madera* ó *Zinc* u *Otro* ó *Material de desecho*
	0	Si alguno de los dos datos tiene código de ignorado ó Todos los demás casos
Material predominante en el Piso	1	Si en 2004 es *Mosaico, cerámica, terrazo* ó *Cemento (lujado o no)* y en 2005 es *Madera* u *Otro* ó *No tiene (piso de tierra)*
	0	Si alguno de los dos datos tiene código de ignorado ó Todos los demás casos
Abastecimiento de Agua	1	Si en 2004 es por *Tubería dentro de la vivienda* y en el 2005 es cualquier otro tipo de abastecimiento
	0	Si alguno de los dos datos tiene código de ignorado ó Todos los demás casos
Procedencia del Agua	1	Si en 2004 es *Un acueducto de AyA* ó *Un acueducto rural* ó *Un acueducto municipal* ó *Una empresa o cooperativa* y en el 2005 es *Un pozo* ó *Un río, quebrada o naciente* ó *Lluvia u otro*
	0	Si alguno de los dos datos tiene código de ignorado ó Todos los demás casos

[22]/ Estos criterios y sus limitaciones fueron realizados consensualmente por técnicos especializados del INEC.

Criterios para identificar errores en la Sección Vivienda y Servicios		
Variable	**Valores**	**Criterios**
Tenencia de Servicio Sanitario	1	Si en 2004 es *Conectado a alcantarilla o cloaca* ó *Conectado a tanque séptico* y en el 2005 es *De pozo negro o letrina* ó *Con otro sistema* ó *No tiene*
	0	Si alguno de los dos datos tiene código de ignorado ó Todos los demás casos
Tenencia de Luz Eléctrica	1	Si en 2004 es del *ICE* ó *CNFL* ó *ESPH* ó *JASEC* ó *CoopeSantos* ó *CoopeLesca* ó *CoopeGuanacaste* ó *CoopeAlfaro* ó *Planta privada* y en el 2005 es *De otra fuente de energía* ó *No hay luz eléctrica*
	0	Si alguno de los dos datos tiene código de ignorado ó Todos los demás casos

El paso siguiente fue sumar errores por registro para las variables personales, para cada hogar: *Error Total.* Ésta cuantificó las situaciones consideradas como error para esas 12 variables, de tal forma que se le asignó valor cero en caso de que no se detectara ningún error en ninguna variable y, valores más lejanos que cero, conforme se agregan errores. Esta variable de Error Total se utilizará posteriormente para realizar una regresión lineal.

Hay que recalcar que los errores obtenidos de esta manera, se refieren a un total de errores por vivienda-hogar y no por persona. Si se hiciera de otra manera, los errores aumentarían tantas veces como número de miembros de hogar hubiese.

Posteriormente, esta variable de *Error Total* se transforma en variable dicotómica, denominada *Error dicotómico*, para establecer la ausencia o presencia de error, con la finalidad de aplicar una regresión logística que permita determinar la probabilidad de ocurrencia de alguno de estos. Este modelo logístico se expresa así:

$$Prob(\ evento\ ) = \frac{1}{1 + e^{-Z}}$$

$$donde \quad Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

El modelo anterior se utiliza para predecir la probabilidad *(y)* de que la variable dependiente, en este caso *"Error dicotómico"*, presente uno de dos valores posibles (1 = error ó 0 = no error) en función de los que adoptan el conjunto de variables independientes [Jovell, 2006:15]. Para ingresar al modelo, el grupo inicial seleccionado como posibles variables independientes fueron:

Relacionadas con el informante en el 2005	**Relacionadas con la vivienda en el 2005**
Tipo de informante	*Región*
Edad	*Zona*
Sexo	*Tamaño del hogar*
Relación de parentesco	*Nivel socioeconómico*
Condición de actividad	*Nivel de pobreza*
Nivel de educación	*Ln Ingreso per cápita del hogar */*
Estado conyugal	

*/ Logaritmo neperiano del Ingreso per cápita del hogar.

No obstante, para las variables categóricas (*Tipo de informante, Relación de parentesco, Sexo, Condición de actividad, Estado conyugal, Región, Zona, Nivel socioeconómico y Nivel de pobreza*), se realizó una selección previa de acuerdo con los criterios brindados por Hosmer y Lemeshow [2000:92], que consiste en aplicar la prueba de Chi-cuadrado de Pearson, puesto que es asintóticamente equivalente a la prueba Chi-

cuadrado de la razón de verosimilitud, por lo que puede ser usada para la selección de las variables.

La tabla siguiente muestra los resultados de estas pruebas, obteniéndose que las variables categóricas que resultaron significativas, fueron: *Tipo de informante*, *Relación de parentesco*, *Estado conyugal*, *Zona* y *Nivel socioeconómico*.

Variables	Error dicotómico			Chi-cuadrado de Pearson	gl	Significancia asintótica (2 colas)	
	No error	Error	Total				
Tipo de informante							
0 = Diferente informante	236	1.063	1.299	53,103	1	0,000	*
1 = Mismo informante	555	1.324	1.879				
Total	791	2.387	3.178				
Sexo							
0 = Hombres	262	800	1.062	0,041	1	0,839	
1 = Mujeres	529	1.587	2.116				
Total	791	2.387	3.178				
Relación de parentesco							
0 = Otro	1	25	26	29,593	3	0,000	*
1 = Jefe(a)	427	1.093	1.520				
2 = Esposo(a)	313	1.001	1.314				
3 = Hijo(a)	50	268	318				
Total	791	2.387	3.178				
Condición de actividad							
0 = Inactivo	408	1.148	1.556	3,786	2	0,151	
1 = Ocupado	359	1.177	1.536				
2 = Desocupado	24	62	86				
Total	791	2.387	3.178				
Estado conyugal							
0 = Soltero	47	115	162	19,591	5	0,001	*
1 = Unión libre	38	144	182				
2 = Casado	156	466	622				
3 = Divorciado	29	81	110				
4 = Separado	81	155	236				
5 = Viudo	76	132	208				
Total	427	1.093	1.520				
Región							
0 = Huetar Norte	38	125	163	3,293	5	0,655	
1 = Central	379	1.208	1.587				
2 = Chorotega	85	250	335				
3 = Pacífico Central	83	250	333				
4 = Brunca	110	305	415				
5 = Huetar Atlántica	96	249	345				
Total	791	2.387	3.178				
Zona							
0 = Rural	503	1.321	1.824	16,532	1	0,000	*
1 = Urbano	288	1.066	1.354				
Total	791	2.387	3.178				
Nivel socioeconómico							
0 = Nivel bajo	33	136	169	20,369	5	0,001	*
1 = Nivel medio bajo	89	316	405				
2 = Nivel medio	157	550	707				
3 = Nivel medio alto	5	36	41				
4 = Nivel alto	4	28	32				
5 = Nivel rural	503	1.321	1.824				
Total	791	2.387	3.178				
Nivel de pobreza							
0 = No pobre	584	1.800	2.384	2,737	2	0,255	
1 = Extrema pobreza	47	162	209				
2 = No satisfacen necesida	160	425	585				
Total	791	2.387	3.178				

Para las variables continuas, lo más deseable es incluir un análisis univariado al modelo de regresión logístico para obtener el coeficiente y el error estándar estimados, la prueba de razón de verosimilitud para la significancia del coeficiente y el estadístico Wald univariado [Hosmer y Lemeshow, op cit] con el fin de determinar cuáles son significativas.

Variables en la ecuación	B	E.E.	Wald	gl	Sig.	Exp(B)
Edad	-0,005	0,003	3,950	1	0,047	0,995
Constante	1,331	0,122	119,040	1	0,000	3,786
Nivel de educación	0,033	0,010	11,173	1	0,001	1,033
Constante	0,840	0,088	91,330	1	0,000	2,316
Tamaño del hogar	0,203	0,026	61,761	1	0,000	1,226
Constante	0,342	0,102	11,209	1	0,001	1,408
Ln Ingreso percápita del hogar	0,135	0,044	9,341	1	0,002	1,145
Constante	-0,364	0,481	0,573	1	0,449	0,695

Con base en los procedimientos anteriores, el modelo de regresión logístico quedó constituido por las siguientes variables independientes:

Relacionadas con el informante en el 2005	**Relacionadas con el hogar en el 2005**
Tipo de informante	*Zona*
Edad	*Tamaño del hogar*
Relación de parentesco	*Nivel socioeconómico*
Nivel de educación	*Ln Ingreso percápita del hogar*
Estado conyugal	

Para ejecutar la regresión logística se empleó el procedimiento *"por pasos"* del *SPSS*, ya que éste es uno de los métodos recomendados cuando existe un elevado número

de posibles variables independientes, sin que exista una teoría o un trabajo previo que oriente la elección de las variables relevantes [Pardo y Ruiz, 2002:385]. Este procedimiento permite eliminar variables, a partir de criterios particulares, que llevan a formar, consecutivamente, nuevos modelos, hasta llegar al que consiga el mejor ajuste. En este caso, se formaron tres, a los que se les denomina *pasos*. Las opciones de este procedimiento, permite identificar y definir las variables categóricas incluidas en la lista de variables independientes (*Tipo de informante*, *Relación de parentesco*, *Estado conyugal*, *Zona*, y *Nivel socioeconómico*) y decidir qué tratamiento recibirán en el análisis.

El método de selección *"por pasos"* hacia delante, ejecutado con el estadístico *razón de verosimilitud* (*RV*), parte del modelo nulo, e incorpora las variables cuyo *RV* es significativo. Con este método, se eliminan por turno cada una de las variables del modelo, y se evalúa si la variable eliminada hace o no perder ajuste. El estadístico *RV* contrasta la hipótesis nula de que la variable eliminada tiene un coeficiente igual a 0. El valor de *RV* para una variable, se obtiene de la división del valor de *RV* para el modelo que excluye esa variable, entre el valor de *RV* para el modelo con esa variable [Pardo y Ruiz, 2002:666].

La salida del paquete *SPSS* proporciona no solo el valor de los coeficientes *B* de la ecuación de la regresión, sino que además, proporciona los correspondientes datos de *odds* o *ventaja* [23/]. El *odds* o la *ventaja*, se define como el cociente de la probabilidad de que ocurra el evento a la probabilidad de que éste no ocurra [Hernández, 1988:162]; es decir, la probabilidad de que se dé uno de los sucesos, dividido por su probabilidad complementaria [Pardo y Ruiz, 2002:660]. Simbólicamente, el *odds* sería:

[23/] Traducido así en Pardo y Ruiz [2002:661].

$$\text{odds} = \text{ventaja}$$

$$= \frac{\text{Prob (evento)}}{\text{Prob (no evento)}}$$

$$= \frac{1 \Big/ \left(1 + e^{-(B_0 + B_1 X_1 + \ldots + B_p X_p)} \right)}{1 - 1 \Big/ \left(1 + e^{-(B_0 + B_1 X_1 + \ldots + B_p X_p)} \right)}$$

$$= e^{(B_0 + B_1 X_1 + \ldots + B_p X_p)}$$

Claramente se nota que una *ventaja* se puede expresar en términos de potencias del número *e*, lo que conduce a que la interpretación de los coeficientes obtenidos en el modelo, se realice razonando en términos de cambios en los logaritmos, lo que resulta poco intuitivo. Lo preferible, es interpretar directamente el cambio en las *ventajas* y no en los logaritmos de las ventajas, con lo cual, se llega al concepto de *razón de ventajas* o *razón de odds* (*odds ratio*), que es una medida de la magnitud de la asociación entre dos variables [Jovell, 2006:29].

La interpretación de la *razón de las ventajas* se puede resumir de la siguiente forma [Pardo y Ruiz, 2002:662]:

- La *razón de las ventajas* vale 1 (y su correspondiente coeficiente de regresión vale cero) cuando la variable independiente no produce ningún efecto sobre la *ventaja* de un suceso.
- La *razón de las ventajas* es mayor que 1 (y su correspondiente coeficiente de regresión es mayor que 0) cuando un aumento en la variable independiente lleva asociado un aumento de la *ventaja* del suceso. Puesto que el punto de comparación es 1, si el valor de *Exp(B)* (*razón de odds*), es mayor que 1, la

diferencia *Exp(B)* – 1 significa la cantidad de veces que la *ventaja* de error supera a la *ventaja* de no error [Pardo y Ruiz, 2002:665].

- La *razón de las ventajas* es menor que 1 (y su correspondiente coeficiente de regresión es menor que 0) cuando un aumento en la variable independiente conlleva una disminución de la *ventaja* del suceso. Como de nuevo, el valor de comparación es 1, si el valor de *Exp(B)* (*razón de odds*), es menor que 1, la diferencia [1 – Exp (B)] significa una reducción proporcional en la *ventaja* de error ante la *ventaja* de no error [Pardo y Ruiz, 2002:666].

Los resultados de analizar los datos con esta regresión, se interpretan a partir de los coeficientes y sus *razones de odds*. Aunque este estimador es usualmente el de mayor interés en una regresión logística, debido a su fácil interpretación, su estimación tiende a una distribución sesgada. Este sesgo obedece a posibles valores en el rango entre 0 e infinito, con valores nulos iguales a 1. No obstante, en teoría, para muestras suficientemente grandes, la distribución de la estimación de la *razón de odds* es normal [Hosmer y Lemeshow,:2000:52]. En el caso de esta investigación, el tamaño de la muestra de hogares se considera suficientemente grande (3.178).

Se aplicó un ejercicio similar al anterior, pero con regresión lineal, puesto que se tenía la cantidad de errores encontrados en cada hogar dentro de la variable *Error_total*. Esta regresión fue aplicada sobre la misma población de hogares que la anterior; es decir, sobre los 3.178 hogares que quedaron libres de ignorados.

Para esta regresión lineal, se utilizó también el procedimiento de "*pasos*", para determinar la importancia de esos errores en los hogares donde hubo diferente número de ellos. Esta variable *(Error_total)* mostró una clara asimetría positiva, por lo que, antes de ejecutar la regresión, se realizaron diferentes cambios de variables con el fin de normalizarla, sin éxito alguno, por lo que se decidió mantenerla tal cual.

2.4 APROXIMACIÓN DE LOS ERRORES NO MUESTRALES

Los errores no muestrales son inevitables y de difícil medición; no obstante, se pueden minimizar a través de procedimientos de control específicos. En este apartado se exponen los resultados obtenidos del empate entre dos Encuestas de Hogares, para medir las diferencias observadas.

El resultado del empate entre las EHPM del 2004 y 2005, fue de 11.061 personas coincidentes, cifra que corresponde a 3.178 hogares diferentes y también coincidentes, medidos respecto al año 2005. Con el término coincidente, se quiere decir, la asociación electrónica realizada entre las mismas viviendas, hogares y personas, entre ambos años.

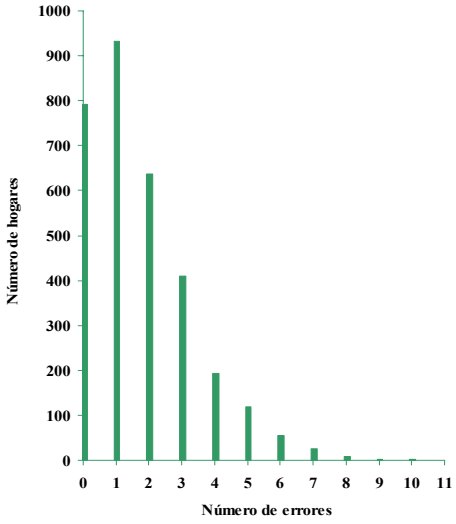
2.4.1 Caracterización de los errores

Los 3.178 hogares coincidentes presentaron un total de 5.436 errores, de los cuales, 4.949 se refieren a errores en 5 variables personales y 487 a errores en 7 variables de la vivienda. La distribución de esos hogares coincidentes, según la cantidad de errores detectados en ellos, presentó una asimetría notable, como se aprecia en el cuadro 2.1 y el gráfico siguiente (2.1). Con un promedio de 1,92 errores y una desviación estándar de 1,73, su distribución indica que una cuarta parte de los hogares no presentaron errores, casi la mitad de ellos tuvieron uno o dos; y la otra cuarta parte tuvo 3 o más. El máximo, de 11 errores se ubica en un solo hogar.

Cuadro 2.1
COSTA RICA: Número de Hogares y de Errores Acumulados
SEGÚN: Cantidad de Errores por Hogar (Absoluto y Relativo)
Empate EHPM 2004-2005

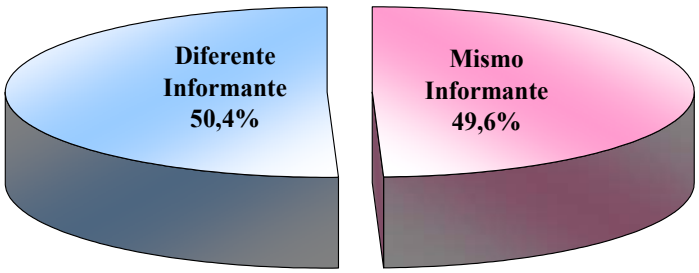
Cantidad de Errores por Hogar	Hogares		Número de Errores Acumulados
	Absoluto	Relativo	
A	B	C	A * B
Total	**3.178**	**100,0**	**5.436**
0	791	24,9	0
1	933	29,4	933
2	638	20,1	1.276
3	409	12,9	1.227
4	194	6,1	776
5	118	3,7	590
6	55	1,7	330
7	26	0,8	182
8	8	0,3	64
9	3	0,1	27
10	2	0,1	20
11	1	0,0	11

Gráfico 2.1
Distribución de hogares, según ocurrencia de errores. 2004-2005



El total de errores estuvo distribuido muy homogéneamente por tipo de informante, tal y como se aprecia en el gráfico siguiente (2.2) [24/].

Gráfico 2.2
Distribución del total de errores por tipo de informante.
Empate EHPM 2004-2005.



FUENTE: Cuadro A 2.1

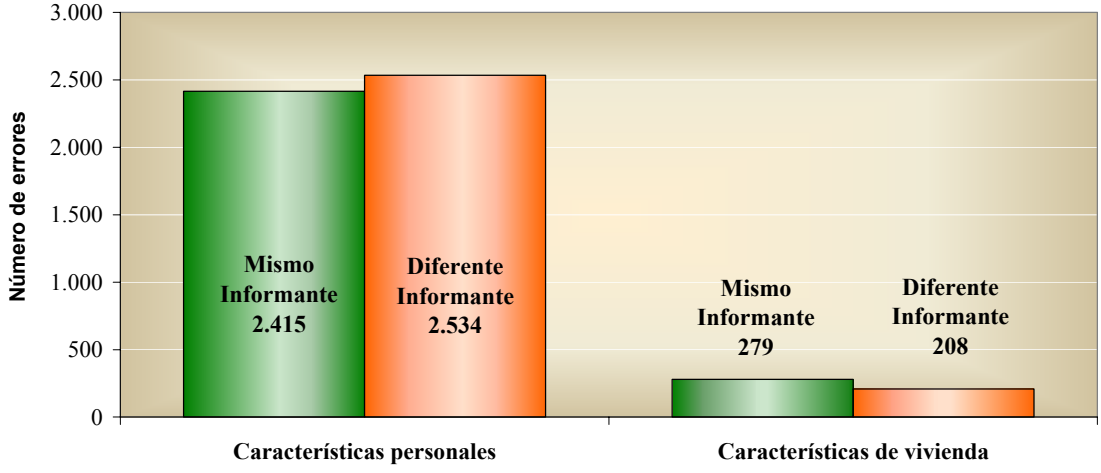
[24/] La letra "A" que antecede al número de cuadro en la fuente de los gráficos, indica que éstos están ubicados en el Anexo de cuadros de este capítulo.

No obstante, al clasificar los hogares de acuerdo con la tenencia de errores, esta distribución varía:

- En los hogares en los que no se encontraron errores, el 70% de las respuestas fueron obtenidas de un mismo informante.
- Por el contrario, en los hogares en los cuales se localizó al menos un error, las respuestas dadas por un mismo informante fue de 55%, lo cual es ligeramente superior a la de los diferentes informantes.

El siguiente gráfico (2.3) señala en términos absolutos, que los errores asociados a las cinco variables de índole personal seleccionadas corresponden al 91% del total, por lo que sería recomendable que en todo el proceso de recolección de datos, se preste más atención a este tipo de variable. El otro 9% tuvo relación con las características de la vivienda, que dada su propia naturaleza poco cambiante, se es más difícil de fallar. En general, los errores fueron producidos proporcionalmente igual, por el mismo informante que por uno diferente.

Gráfico 2.3
Total de errores por tipo de caterística y tipo de informante
Empate EHPM 2004-2005



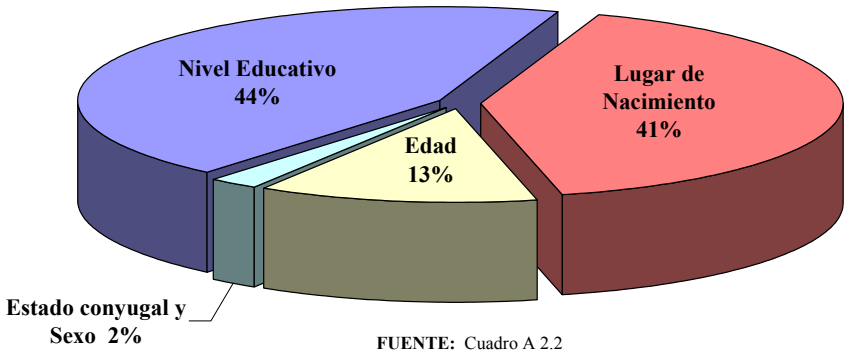
FUENTE: Cuadro A 2.1

2.4.1.1 Errores en las variables de tipo personal

Al empatar la EHPM del 2004 y 2005, se obtuvo una población coincidente de 11.061 personas. De ellas, 4.019 registraron errores en alguna de las cinco características de tipo personal que se seleccionaron para compararlas. No obstante, aparecen totalizados 4.949 errores en las respuestas dadas. Esta cifra no se debe interpretar en términos de personas, por cuanto una misma pudo haber cometido hasta cinco errores correspondientes a las cinco variables escogidas para este estudio. Por ejemplo, un individuo pudo haber declarado mal su estado conyugal y su lugar de nacimiento, de tal manera que su registro presentaría dos errores.

El mayor recuento de errores se concentró en dos variables: *Nivel educativo* y *Lugar de nacimiento*, ambas recogen el 85% del total de estos errores (gráfico 2.4). En el otro extremo, las variables *Sexo* y *Estado conyugal* fueron las de menor incidencia. Hay que recordar que algunos de estos errores, en especial los referentes al *Sexo*, pudieron haberse originado exclusivamente en el procedimiento de empate electrónico entre ambas encuestas.

Gráfico 2.4
Distribución de los errores personales por tipo de variable.
Empate EHPM 2004-2005

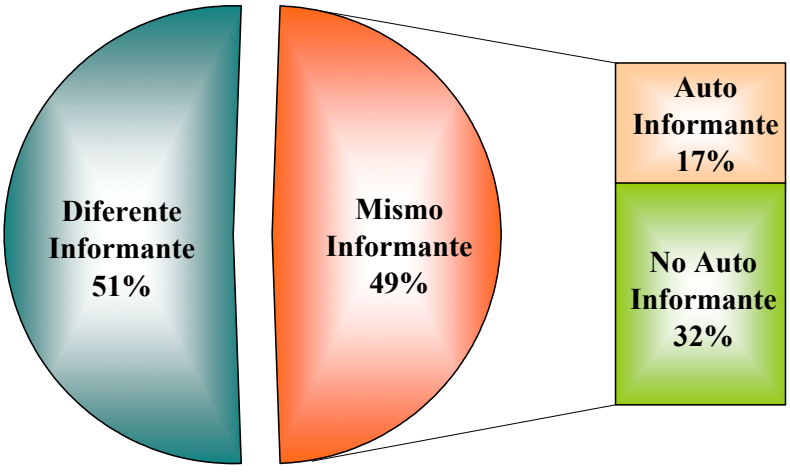


FUENTE: Cuadro A 2.2

Sin importar de cuál tipo de informante provengan los errores de carácter personal, la estructura mostrada por variable estudiada es similar. Por ejemplo, en todos los casos, el *Nivel educativo* y el *Lugar de nacimiento* suman alrededor del 85% de todos los errores, no importa quién sea el informante.

En el nivel de informante, solo en la variable *Edad* suceden los errores mayoritariamente con cambio de informante, mientras que con la variable *Estado Conyugal* sucede al contrario, con el mismo informante (Autoinformante [25/] o no) las personas tienden más a cambiar la información. No obstante, considerando el total de errores, cuando la información se refiere al propio Autoinformante, el porcentaje de errores disminuye alcanzando un 17%. En las demás variables las proporciones son semejantes a la distribución del total mostrada en el siguiente gráfico (2.5).

Gráfico 2.5
Distribución de los errores en las características personales, por tipo de informante. Empate EHPM 2004-2005.

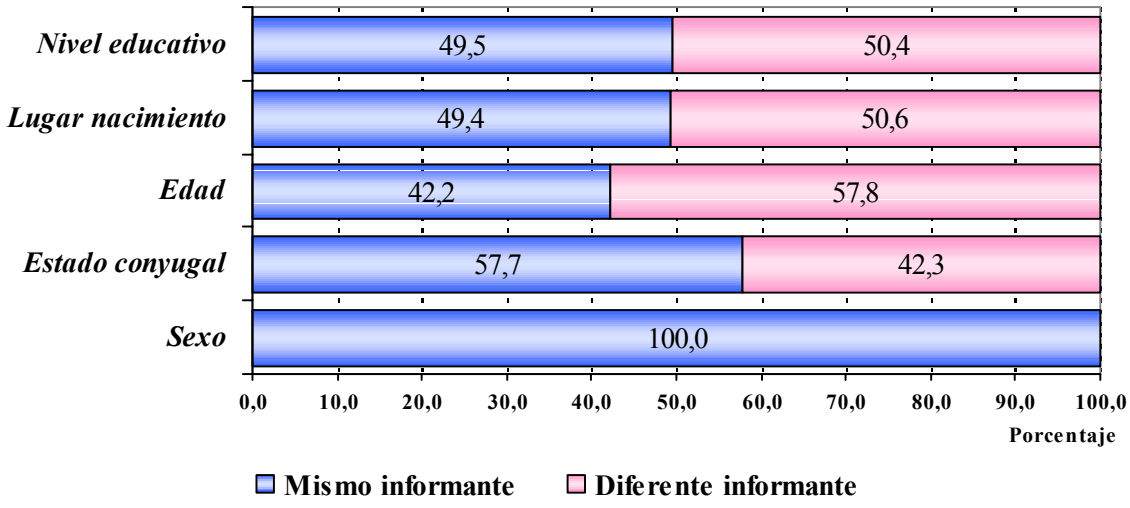


FUENTE: Cuadro A 2.2.

[25/] Persona que proporciona los datos sobre sí misma.

En términos porcentuales, se observa que las variables con mayor cantidad de errores, esto es, *Nivel educativo* y el *Lugar de nacimiento*, se distribuyen casi homogéneamente entre los dos tipos de informante, como se aprecia en el gráfico 2.6. Las siguientes dos variables en orden de importancia respecto a la incidencia de errores (*Edad* y *Estado conyugal*), fluctúan en ± 8 puntos porcentuales alrededor del 50%. La variable *Sexo*, que tiene solamente 3 errores, fueron todos producto del mismo informante, o de errores en el procedimiento del empate electrónico. Un dato interesante respecto al *Estado conyugal* es que cuando la persona declaró acerca de sí misma, tendió más a cambiar la información en esta variable.

Gráfico 2.6
Estructura de los Errores Personales,
por Tipo de variable y Tipo de informante. 2004-2005



FUENTE: Cuadro A 2.2.

2.4.1.2 Errores en las características de vivienda

Los errores detectados en el nivel de vivienda, se deben analizar respecto de cada hogar entrevistado, puesto que son variables que contienen información sobre la estructura de ella. Al igual que con las características personales de cada persona, los errores de vivienda pueden ser acumulativos en cada hogar; esto es, un hogar en particular podría acumular hasta siete errores correspondientes a las siete características de vivienda estudiadas.

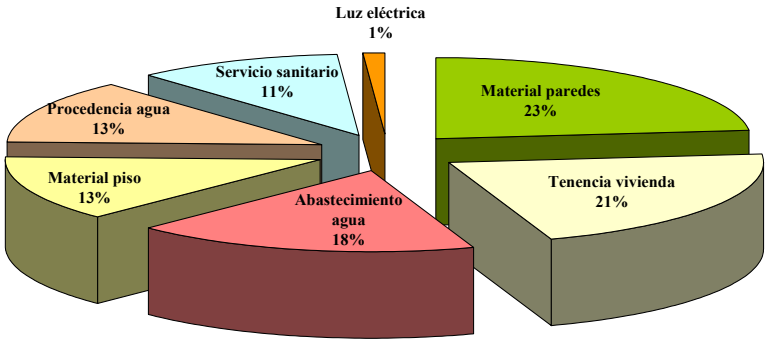
Hay que recordar, tal y como se explicó en el apartado de Metodología, que los criterios utilizados para definir un error en cada una de las variables seleccionadas en esta sección, no son del todo incuestionables. La duda radica en que se podría encontrar alguna situación extrema en la cual el cambio ocurrido en la vivienda se deba a hechos fortuitos y no a un error. No obstante, son analizadas en este apartado por dos razones: la probabilidad de que esas situaciones extremas ocurran es muy baja [26/], y la incidencia de este tipo de error en el total de errores es baja (9%).

Efectivamente, estos errores de vivienda, comparados con los analizados anteriormente, son significativamente menores: 487 de un total de 5.436. Además, la estabilidad que tienen estas variables en el tiempo es lo que permite que, generalmente, todos los miembros del hogar puedan describirlos de forma menos errática.

El *Material de las paredes exteriores* es la variable más afectada en errores, como lo muestra el gráfico 2.7. Esto podría deberse a que el entrevistador pudiera, asignar la respuesta a esta pregunta, mediante una revisión visual que él haga, lo que podría conducir a error; lo mismo podría decirse del *Material de pisos*. Por otra parte, se puede justificar que la *Tenencia de vivienda* sea la segunda variable de mayor incidencia en el número de errores, debido a que algunos informantes podrían confundirse acerca de la pertenencia, sobre todo cuando se está pagando a plazos.

[26/] Así lo señaló el equipo técnico del INEC, que revisó estos criterios.

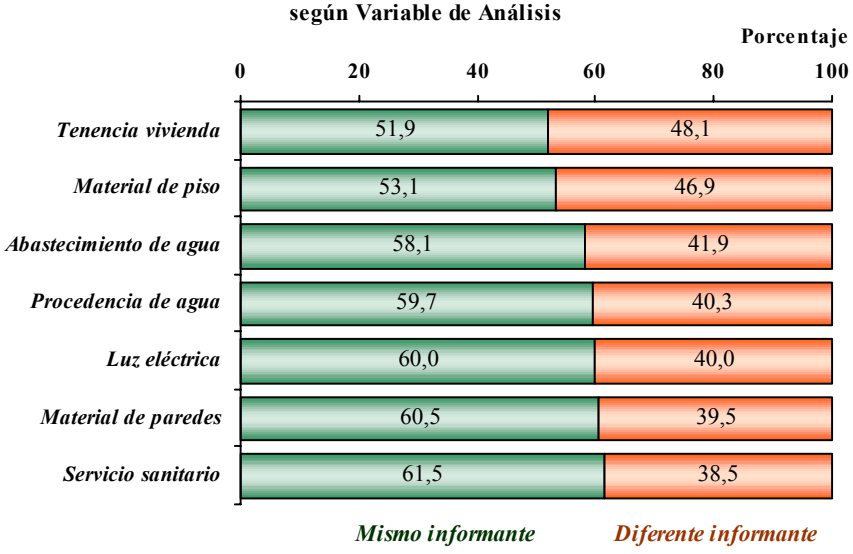
Gráfico 2.7
Distribución de los errores en vivienda por tipo de variable.
Empate EHPM 2004-2005



FUENTE: Cuadro A 2.3.

En la estructura porcentual por tipo de informante, sobresalen los errores que cometen los mismos informantes, de tal forma que su magnitud es mayor a los que cometen los informantes diferentes. Efectivamente, el siguiente gráfico (2.8) indica que en todas las variables, se sobrepasa el 50% en la distribución de errores del Mismo informante.

Gráfico 2.8
Porcentaje de Errores Totales en la Sección Vivienda,
según Variable de Análisis



FUENTE: Cuadro A 2.3

Toda esta información acerca de inconsistencias en las dos encuestas analizadas, es la base para incorporar al análisis de un modelo explicativo a través de regresiones.

2.4.2 Regresión logística

El modelo de regresión logística es utilizado para relacionar una variable dependiente de tipo dicotómica (1 = Sí, 0 = No) con otra u otras variables independientes. En el caso de esta investigación, la variable dependiente se denomina *Error_dicotómico*, y contiene información acerca de la presencia o ausencia de error en un hogar, en forma binaria (1 = sí hay error ó, 0 = no hay error). Las variables independientes que luego de un proceso de selección, ingresaron a la regresión, fueron: *Tipo de informante*, *Edad*, *Relación de parentesco*, *Nivel de educación* y *Estado conyugal*, todas estas variables se refieren al informante; además se incluyeron *Zona*, *Tamaño del hogar*, *Nivel socioeconómico* y *Ln Ingreso per cápita del hogar*. Estas se escogieron como las posibles variables que inciden en la ocurrencia de esos errores. Los resultados obtenidos de este modelo de regresión logística, estiman la probabilidad de hallar al menos un error, en un hogar tomado al azar.

Los hogares seleccionados para la ejecución de esta regresión fueron 3.178. Para conocer cómo se comporta la variable *Error_dicotómico*, se muestra su distribución de frecuencia. Un dato interesante, que ya ha sido mencionado, es que solo el 25% de los hogares no presentaron errores en sus respuestas.

Error Total dicotómico 04-05

Presencia de error	Frecuencia	Porcentaje
Total	**3.178**	**100,0**
1 = Error	2.387	75,1
0 = No error	791	24,9

La regresión se realizó bajo el método de selección *"por pasos"* hacia delante, con el estadístico *razón de verosimilitud* (*RV*) [27/]. Este método permite utilizar criterios estadísticos para que de forma automática, se incluyan en el modelo las variables que son significativas, y se dejen fuera las que no lo son [Pardo y Ruiz, 2002:666]. En el anexo B se presentan los resultados comentados de esta regresión. Seguidamente se exponen los principales:

- El modelo fue de tres pasos; de los cuales, el último dio el mejor ajuste.
- La tabla denominada "de clasificación" indica que el modelo logró clasificar correctamente el 60,9% de resultados de errores y no errores, con un valor de corte de 0,71 [28/], lo que sugiere que el modelo logístico ajusta relativamente bien los datos.

Tabla de clasificación (a)

Observado			Pronosticado		
			Error_dicotómico		Porcentaje correcto
			No error	Error	
Paso 3	Error_dicotómico	No error	240	187	56,2
		Error	408	685	62,7
	Porcentaje global				**60,9**

a El valor de corte es ,710

- Los coeficientes del modelo y de la constante fueron evaluados por la prueba de Wald [29/], y considerados estadísticamente significativos en los tres pasos realizados, excepto la constante. Para efectos gráficos, se presenta únicamente el último paso.

[27/] No obstante, como ejercicio académico, se realizaron todos los posibles métodos por pasos que incorpora el SPSS (adelante condicional, adelante RV, adelante Wald, atrás condicional, atrás RV y atrás Wald), obteniéndose en todos los mismos resultados.

[28/] Aunque con un valor de corte de 0,50, el porcentaje de clasificación asciende a 71,9%, el corte seleccionado (0,71) logra que los porcentajes de clasificación correcta de las dos categorías se equiparen lo mejor posible (56,2% y 62,7%).

[29/] La prueba de Wald contrasta la hipótesis de que un coeficiente vale cero en la población [Pardo y Ruiz, 2002:686].

Variables en la ecuación

		B	E.T.	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso 3	Tipo_informante	-0,507	0,130	15,086	1	0,000 *	0,603	0,467	0,778
	Nivel_educacion	0,032	0,013	6,281	1	0,012 *	1,032	1,007	1,059
	Tamano_hogar	0,216	0,037	34,190	1	0,000 *	1,241	1,154	1,334
	Constante	0,349	0,192	3,302	1	0,069	1,418		

- Conforme con lo esperable, el coeficiente correspondiente a la variable *Tipo de informante* es negativo, lo que indica que cuando se está en presencia del mismo informante, se tiene menor probabilidad de cambiar las respuestas de una encuesta a otra, en las preguntas correspondientes a las variables del modelo.
- Por el contrario, en la variable *Nivel de Educación* se tiene un coeficiente positivo, el cual señala que a mayor nivel de escolaridad, mayor probabilidad de cometer al menos un error en los datos brindados.
- Con la variable *tamaño del hogar* sucede también lo esperable: a mayor cantidad de datos recolectados en un hogar, debido al tamaño de éste, mayor sería la probabilidad de cometer al menos un error.
- Los intervalos de confianza al 95%, para cada valor *Exp(B)* no incluyen el valor 1 [30/], con lo cual se puede concluir que la variable independiente posee un efecto significativo. Si además se considera que *"los intervalos de confianza informan sobre la importancia relativa de las variables independientes"* [31/], se puede afirmar que las tres variables son importantes en términos relativos.

[30/] Contrasta la hipótesis nula de que la razón de ventajas vale 1 en la población, lo cual es equivalente a contrastar con el estadístico de Wald [Pardo y Ruiz, 2002:686].

[31/] Las variables a las que corresponden intervalos que se solapan son variables con un efecto similar; las variable con intervalos que no se solapan son variables con un efecto significativamente distinto (aunque no debe olvidarse que la magnitud de la razón de las ventajas depende de la métrica de las variables) [Pardo y Ruiz, op cit].

- La prueba de bondad de ajuste de Hosmer y Lemeshow, indica que en los tres modelos (pasos) no existen diferencias entre las frecuencias observadas y las esperadas [32/], por lo que se puede asumir que el modelo de interés (el tercero) se ajusta a los datos [Pardo y Ruiz, 2002:684].

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	8,310	4	0,081
2	8,814	6	0,184
3	12,424	8	0,133

- Los resultados obtenidos en los *pseudos* R^2 [33/] denominados estadísticos de ajuste global, permite verificar que el tercer modelo (paso) obtuvo los más altos. No obstante, sus magnitudes son bajas: 4,8% en el R^2 de Cox y Snel y 7,0% en el R^2 de Nagelkerke [34/], lo que indica que la presencia de las variables incluidas determinan aproximadamente el 5 y 7%, respectivamente, de la probabilidad de ocurrencia de al menos un error, con lo que se obtiene una capacidad explicativa del modelo muy baja. Sin embargo, se asegura que estos estadísticos de bondad de ajuste global, para este tipo de modelo, son meramente orientativos, y suelen adoptar valores moderados e incluso bajos, aun cuando el modelo estimado pueda ser apropiado y útil [Pardo y Ruiz, 2002:654]; con lo cual, se podrían obviar estos resultados y continuar con el análisis.

Resumen de los modelos

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	1752,800 (a)	0,034	0,049
2	1736,307 (a)	0,044	0,064
3	1729,928 (a)	**0,048**	**0,070**

a La estimación ha finalizado en el número de iteración 4 porque las estimaciones de los parámetros han cambiado en menos de ,001.

[32] / Se espera que la hipótesis nula no deba ser rechazada, puesto que este estadístico permite contrastar la hipótesis nula de igualdad de distribuciones, es decir, la hipótesis de que la variable dependiente se distribuye de la misma manera en los 10 deciles de riesgo (la muestra se divide en 10 grupos a partir de los deciles de las probabilidades pronosticadas) [Pardo y Ruiz, 2002:684].

[33]/ La medida convencional de la bondad de ajuste, R^2, no es particularmente significativa para los modelos con regresada binaria, por lo que existen una variedad de medidas similares a R^2, llamadas *pseudos* R^2 [Gujarati, 2003:584].

Con los datos obtenidos, el modelo propuesto en el tercer paso obtuvo el mejor ajuste, lo que se expresa como la probabilidad de ocurrencia de error, de la siguiente manera:

$$Probabilidad\ (\ Error\ ) \ = \ \frac{1}{1 \ + \ e^{\ -(\ -0.507\ TI\ \ +0{,}216\ TH\ \ +0{,}032\ NE\ )}}$$

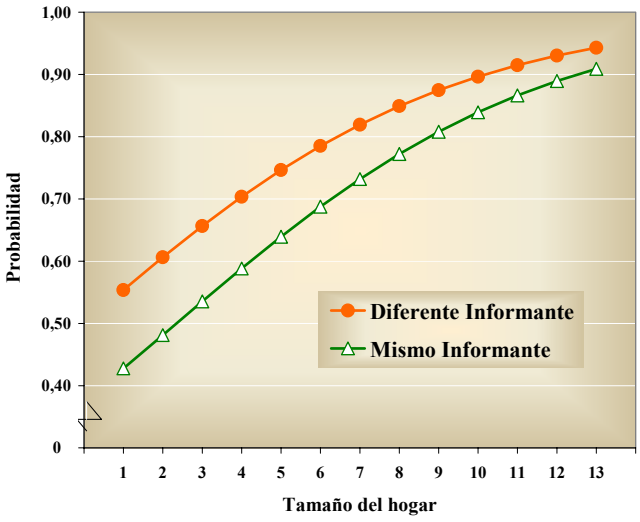
Donde: TI = Tipo de informante
TH = Tamaño del hogar
NE = Nivel de educación

El cálculo de las probabilidades de que se presente algún error, para diferentes tamaños de hogar, con distinto o igual informante, controlando por nivel de educación, puede ilustrarse en el siguiente cuadro (2.2) y gráfico (2.9):

Cuadro 2.2
COSTA RICA: Probabilidades de error
SEGÚN: Tamaño del hogar
POR: Tipo de informante
Empate EHPM 2004-2005

Tamaño del hogar	Probabilidad de error	
	Mismo Informante	Diferente Informante
1	0,428	0,554
2	0,481	0,606
3	0,535	0,657
4	0,588	0,703
5	0,639	0,746
6	0,688	0,785
7	0,732	0,819
8	0,772	0,849
9	0,808	0,875
10	0,839	0,897
11	0,866	0,915
12	0,889	0,930
13	0,909	0,943

Gráfico 2.9
Probabilidades de error,
según Tamaño del hogar, por Tipo de informante.
2004 - 2005



[34]/ La R^2 de Nagelkerke es una modificación de la R^2 de Cox-Snell, sus valores pueden ir de 0 a 1, y tienen el mismo significado que la R^2 en la regresión lineal [García, Palma, Rodríguez y Sarría, 2003:3].

La tabla anterior señala que, si se controla por nivel de educación, en el mejor de los casos, con el mismo informante y con solo un miembro de hogar, la probabilidad de encontrarse con al menos un error, en este empate de encuestas, es del 43%. Si se considera que en promedio los miembros por hogar son 3,8, según la tabla, se estarían manejando, en promedio, cifras de riesgos entre 53% y 70%.

Los coeficientes resultantes del modelo se pueden interpretar mediante cocientes de probabilidades, para obtener los llamados "*odds*" o "*ventajas*" de la ocurrencia de un evento. En el modelo desarrollado en este estudio, el cociente de la probabilidad de obtener un error [35/] a la probabilidad de no obtenerlo, estaría determinado por la siguiente expresión:

$$odds = \frac{Prob\,(evento)}{Prob\,(no\ evento)} = e^{(\,-0{,}507\ TI\ +\ 0{,}216\ TH\ +\ 0{,}032\ NE\,)}$$

No obstante, la *razón de odds* (*odds ratio*) proporciona la magnitud de asociación entre dos variables [Jovell, 2006:29]. Los resultados, mencionados en la salida del *SPSS* como *Exp(B)*, muestran los siguientes datos:

Tipo de informante	0,603
Tamaño del hogar	1,241
Nivel de educación	1,032

Estos datos lo que señalan es que el riesgo de obtener algún error en alguna de las variables atinentes, es 40% menor si el informante es el mismo a que si fuera uno distinto. De igual manera, ese riesgo es 24% mayor cada vez que se incluye un miembro de hogar y 3% mayor con cada año de escolaridad del informante.

[35/] Se define así, por cuanto el código 1 de la variable *Error dicotómico* corresponde a la presencia de algún error en las variables estudiadas dentro del hogar.

Con el objetivo de profundizar en la importancia que estas variables tienen en cuanto a la presencia o no de errores, se procedió a realizar una regresión lineal, que inicia con las trece variables independientes de la regresión logística.

2.4.3 Regresión lineal

El número de errores totales encontrados en los hogares (y no solamente su presencia) puede analizarse a través de una regresión lineal múltiple. Este procedimiento estima la media poblacional del *Error_total* en términos de los valores conocidos de las trece variables explicativas originales utilizadas en la regresión logística anterior. La aceptación de estos tipos de modelos requiere el cumplimiento de ciertos supuestos que son analizados en las salidas de los datos.

La regresión lineal múltiple se aplicó a los mismos 3.178 hogares, a través del paquete SPSS, con el método *por pasos hacia delante*. Se obtuvieron cinco modelos diferentes (cinco pasos), de lo que resulta el último, con el mejor ajuste y los siguientes resultados:

- El quinto modelo presentó el mejor ajuste con cinco variables independientes, además de la constante. Éstas fueron: *Tamaño del hogar, Tipo de informante*, *Ln Ingreso per cápita*, *Edad* y *Nivel de pobreza.*
- No obstante la significancia global alta (0,000), que permite deducir una relación lineal significativa entre el *Error Total* y el conjunto de las variables independientes, esas cinco variables tomadas en conjunto, logran explicar solamente un 12,5% de la varianza del *Error Total*, pues R^2 corregida, vale 0,125. Sin embargo, se observaron otros indicios asociados a los diferentes supuestos del modelo de regresión lineal.

Resumen del modelo (f)

Modelo	R	R cuadrado	R cuadrado corregida	Error típico de la estimación	Estadísticos de cambio				
					Cambio en R cuadrado	Cambio en F	gl1	gl2	Sig. del cambio en F
1	,285 (a)	0,081	0,081	1,545	0,081	280,846	1	3176	0,000
2	,330 (b)	0,109	0,108	1,522	0,027	97,401	1	3175	0,000
3	,345 (c)	0,119	0,118	1,514	0,010	36,433	1	3174	0,000
4	,351 (d)	0,123	0,122	1,510	0,004	16,085	1	3173	0,000
5	,356 (e)	0,127	0,125	1,507	0,003	12,273	1	3172	0,000

a Variables predictoras: (Constante), Tamaño del hogar
b Variables predictoras: (Constante), Tamaño del hogar, Tipo informante
c Variables predictoras: (Constante), Tamaño del hogar, Tipo informante, Ln Ingreso percápita hogar
d Variables predictoras: (Constante), Tamaño del hogar, Tipo informante, Ln Ingreso percápita hogar, Edad
e Variables predictoras: (Constante), Tamaño del hogar, Tipo informante, Ln Ingreso percápita hogar, Edad, Nivel de pobreza
f Variable dependiente: Error total

- Se obtuvieron diferentes indicadores para probar los supuestos del modelo:
  - *Normalidad de los residuos:* el histograma revela una asimetría positiva, y el gráfico de probabilidad normal P-P denota cierta lejanía a la diagonal, de los residuos.
  - *Independencia entre los residuos:* el estadístico Durbin-Watson fue de 1,928, valor [36/] que según Pardo y Ruiz [2002:373-374], se encuentra dentro del rango aceptable (1,5 y 2,5) para suponer que los residuos son independientes.
  - *Colinealidad:* el estadístico *Tolerancia* ($1-R^2$), con valores muy pequeños, indica que esa variable puede explicarse por una combinación lineal del resto, fueron superiores a 0,434 en el quinto paso. Asimismo, el estadístico *FIV (Factores de Inflación de la Variancia)*, valores inversos a la *Tolerancia*, que debe ser menor al valor $1/(1-R^2)$ para no presentar colinealidad, casi se mantienen iguales.

    Otros indicadores de este tipo son presentados en el diagnóstico de colinealidad. Uno de los que señala la no colinealidad, son las dimensiones que explican gran cantidad de variancia de un solo

coeficiente (excepto la constante) [Pardo y Ruiz, 2002:380]; no obstante, en el quinto paso, de interés, esto se cumple.

- *Homocedasticidad:* el gráfico de dispersión, para la variable *Error Total*, entre residuos y pronósticos, no debería presentar ningún patrón de comportamiento para afirmar que la variancia de los residuos es constante. Sin embargo, el gráfico obtenido de los datos presenta patrones lineales, como puede verse en el anexo C.

A pesar de que el quinto paso es el mejor modelo de los obtenidos, éste carece de solidez, además, viola varios supuestos:

- Aunque los coeficientes fueron significativamente distintos de cero, lo que contribuye al ajuste del modelo, y, de tener una alta significancia global (el estadístico F permite decidir que sí existe relación lineal significativa entre la variable dependiente y el conjunto de variables independientes tomadas juntas), el modelo logra explicar solamente un 12,5% de la variabilidad.
- La distribución de los residuos es asimétrica positiva, lo que contradice uno de los supuestos más importantes del modelo de regresión múltiple: la normalidad.
- Los gráficos de regresiones parciales, para cada una de las variables dependientes, no presentan asociación lineal evidente con el Error total estandarizado, lo que contradice también otro de los supuestos del modelo.

Por todo esto, se puede afirmar que no se obtienen resultados satisfactorios de este modelo lineal, y que la explicación de los errores debe circunscribirse al modelo de regresión logística.

---

[36]/ Las tablas D.5A y D.5B en Gujarati [2004:942-945], presenta valores entre 1,718 y 1,820 para n=200 y k=5, con un nivel de significancia del 0,05; y de 1,623 y 1,725 para un nivel de significancia del 0,01.

2.5 REFLEXIONES FINALES

No se encontraron errores en el 25% de los hogares coincidentes, y de estos, el 70% de las respuestas fueron obtenidas de un mismo informante en ambas encuestas. Por el contrario, en los hogares en los cuales se localizó al menos un error, las respuestas fueron dadas por un mismo informante, en una proporción ligeramente superior a la de los informantes distintos.

Se obtuvo un total de 5.436 errores en 3.178 hogares coincidentes. La distribución de ellos indicó que casi una cuarta parte de los hogares no presentaron errores; el 50% tuvieron uno o dos, y la otra cuarta parte tuvo 3 o más, con un máximo de 11 ubicado, en un solo hogar. El 90% de estos errores surgieron de las cinco variables de índole personal en estudio; el otro 10% de las siete características seleccionadas sobre la vivienda. Considerando que en esta base de datos coincidentes se tiene un promedio de 3,3 miembros por hogar, existe una desproporción evidente entre esos porcentajes de errores encontrados, por lo que esto marca un primer indicio acerca del esfuerzo por prestar especial atención a las variables de tipo personal, ubicadas en el Módulo Socioeconómico de la EHPM.

El 85% de los errores sobre características personales recayeron en dos de las cinco variables estudiadas: *Nivel educativo* y *Lugar de nacimiento*. Esto sin importar quién es el informante. Es interesante además que la estructura de errores por variable, presenta pocas diferencias entre los diferentes tipos de informantes.

Los datos indican que, cuando el "mismo informante" está declarando acerca de sí mismo, tiende más a cambiar la información sobre *Estado conyugal* que cuando declara sobre esta misma variable para otro miembro del hogar (35 de 60 errores).

Casi la mitad de los errores de vivienda se refieren a la *Tenencia* de ella y al *Material de las paredes externas*, y para todas las variables de este tipo, proporcionalmente

hay más errores en las obtenidas con el mismo informante, lo cual es llamativo, porque lo esperable era que los errores tendieran a disminuir.

El modelo de regresión logístico logró clasificar correctamente el 60,9% de resultados de errores y no errores, con un buen ajuste de los datos. Las variables que resultaron significativas en el modelo fueron *Tipo de informante*, *Tamaño del hogar* y *Nivel de educación*. El que la primer variable resultara con coeficiente negativo y las otras dos con coeficientes positivos, significa que con un informante diferente, y conforme aumenta el *Tamaño del hogar* y el *Nivel de educación*, la probabilidad de encontrar al menos un error, aumenta.

En el mejor de los casos y controlando por *Nivel de educación*, la probabilidad de error es de un 42,8%, con un mismo informante y un miembro de hogar. Dado que en promedio los hogares tienen 3,8 miembros, se puede decir que se estarían manejando probabilidades entre el 53% y 70% de cometer al menos un error en cualquiera de las variables consideradas, en un mismo hogar.

Las *razones de odds* (*odds ratio*) señalan que el riesgo de obtener un error en alguna de las variables estudiadas, es 40% menor si el informante es el mismo a que si fuera uno diferente. De igual manera, ese riesgo se incrementa en 1,241 veces el *Tamaño del hogar* y, 1,032 veces el *Nivel de educación* del informante.

Los anteriores resultados confirman lo que se ha creído hasta el momento: que el tener a un mismo informante garantiza cierta calidad en las respuestas dadas; y que a mayor cantidad de miembros de hogar, mayor probabilidad de cometer al menos un error, a partir del llenado del cuestionario. No así con el *Nivel de educación*: el riesgo de obtener al menos un error se incrementa conforme aumenta el nivel de escolaridad formal del informante.

La regresión multilineal no produjo resultados satisfactorios en cuanto a la bondad de ajuste del modelo, la cual fue muy baja (12,5%), ni tampoco en lo que respecta al

cumplimiento de algunos de los supuestos de este modelo, por lo que la explicación de los errores debe centrarse en el modelo de regresión logístico anterior.

Dado que la muestra es muy grande, lo apropiado hubiera sido seleccionar dos muestras del archivo original, con el fin de ajustar el modelo a una de ellas y evaluar la capacidad predictiva en la otra muestra. El hecho de que la muestra sea muy grande incide en que prácticamente todos los coeficientes sean significativos. Esto refuerza la necesidad de trabajar con una muestra menor.

Aunque efectivamente, para la predicción de los errores totales se utilizó la regresión lineal múltiple, eventualmente se puede hacer uso de otros tipos de modelos que se basan en la ocurrencia de eventos en espacios de tiempo, tal como la regresión de Poisson.

Con respecto a las tareas operativas que viene realizando el INEC, hay que resaltar que se han implementado una serie de tareas en cada etapa de la encuesta, con el fin de disminuir el error no muestral; sin embargo, este esfuerzo no queda documentado en la publicación oficial de la encuesta. Sería interesante que el texto evidenciara estos procesos, que persiguen el objetivo de minimizar este tipo de error, y que sistemáticamente se llevan a cabo.

Finalmente, hay que considerar posible que algunas de estas conclusiones podrían variar, si el procedimiento de empate entre encuestas cambiara.

BIBLIOGRAFÍA

Alvarado Salas Yadira Mª, 2004. **Construcción de Indicadores de Calidad para Evaluar la Encuesta de Hogares de Costa Rica.** Curso Práctica Profesional 2, Universidad de Costa Rica, Documento inédito. San José, Costa Rica.

García Lucía, Palma Matilde, Rodríguez Abel, Sarría Antonio, 2003. **Análisis de la mortalidad intrahospitalaria de la cirugía de revascularización coronaria.** Instituto de Salud Carlos III. Madrid, España. www.revespcardiol.org/cgi-bin/wdbcgi.exe/cardio/mrevista_cardio.fulltext?*pident* =13049651. Consultado en enero de 2008.

Gujarati Damodar N., 2003. **Econometría**. Mc Graw Hill Interamericana, México.

Hernández Rodríguez Óscar, 1998. **Temas de Análisis Estadístico Multivariado**. Editorial de la Universidad de Costa Rica, San José, Costa Rica.

Hosmer David W. y Lemeshow Stanley, 2000. **Applied Logistic Regression**. Segunda edición, Wiley-Interscience Publication, John Wiley & Sons, Inc. United States of America..

INEC, 2005a (Instituto Nacional de Estadística y Censos). **Encuesta de Hogares de Propósitos Múltiples julio 2005. Principales Resultados.** Instituto Nacional de Estadística y Censos, San José, Costa Rica.

INEC, 2005b (Instituto Nacional de Estadística y Censos). **Imputación de Datos de Mercado de Trabajo.** Instituto Nacional de Estadística y Censos, Ponencia ante el 14º Taller de MECOVI, San José, Costa Rica. www.eclac.cl/deype/mecovi/taller14.htm

Jovell Albert J., 2006. **Cuadernos Metodológicos. 15 Análisis de regresión logística**. Centro de Investigaciones Sociológicas. Madrid, España.

Kalton G. y Moser C.A., 1975. **Survey Methods in Social Investigation.** H.Z.B. Paperback. London.

ONU, 1983 (Naciones Unidas). **Errores no muestrales en las encuestas de hogares: fuentes, evaluación y control. Programa para desarrollar la capacidad nacional de efectuar encuestas de hogares. Versión preliminar.** Departamento de Cooperación Técnica para el Desarrollo. Nueva York, Estados Unidos de América.

Pardo Merino Antonio, Ruiz Díaz Miguel Ángel, 2002. **SPSS 11 Guía para el análisis de datos**. Editorial Mc Graw Hill. España.

ANEXOS

A. Resultados Absolutos

Cuadro A 2.1
COSTA RICA: Hogares coincidentes y Errores encontrados
SEGÚN: Hogares con o sin errores y Tipo de error
POR: Coincidencias en el informante
2004 - 2005 (Absolutos y relativos)

Variables	Absoluto			Relativo		
	Total	Tipo de Informante		Total	Tipo de Informante	
		Mismo Informante	Diferente Informante		Mismo Informante	Diferente Informante
Total de Hogares Coincidentes	**3.178**	**1.879**	**1.299**	**100,0**	59,1	40,9
Hogares sin errores	791	555	236	**100,0**	70,2	29,8
Hogares con errores	2.387	1.324	1.063	**100,0**	55,5	44,5
Hogares con errores en vivienda */	434	247	187	**100,0**	56,9	43,1
Hogares con errores en personas */	2.274	1.246	1.028	**100,0**	54,8	45,2
Total de Errores	**5.436**	**2.694**	**2.742**	**100,0**	49,6	50,4
Errores en características personales	4.949	2.415	2.534	**100,0**	48,8	51,2
Errores en características de vivienda	487	279	208	**100,0**	57,3	42,7

*/ La suma de estos dos rubros no coincide con el Total Hogares con errores porque un mismo hogar podría tener tanto errores en características personales como en características de la vivienda.

Cuadro A 2.2
COSTA RICA: Errores encontrados en la Población Total Coincidente de la EHPM
SEGÚN: Población coincidente con o sin error y
Variables personales objeto de análisis
POR: Coincidencias en el informante
2004 - 2005 (Valores absolutos y relativos)

Variables	Total	Tipo de Informante			
		Mismo Informante			Diferente Informante
		Total	Auto Informante	No Auto Informante	
Población Total Coincidente	**11.061**	**6.165**	**1.879**	**4.286**	**4.896**
Con errores	4.019	2.028	730	1.298	1.991
Sin errores	7.042	4.137	1.149	2.988	2.905
Total Errores Personales	**4.949**	**2.415**	**856**	**1.559**	**2.534**
Sexo	3	3	3	0	0
Edad	618	261	67	194	357
Lugar de Nacimiento	2.041	1.009	355	654	1.032
Estado Conyugal	104	60	35	25	44
Nivel Educativo	2.183	1.082	396	686	1.101
Relativo respecto al Tipo de informante					
Total Errores Personales	**100,0**	48,8	17,3	31,5	51,2
Sexo	**100,0**	100,0	100,0	0,0	0,0
Edad	**100,0**	42,2	10,8	31,4	57,8
Lugar de Nacimiento	**100,0**	49,4	17,4	32,0	50,6
Estado Conyugal	**100,0**	57,7	33,7	24,0	42,3
Nivel Educativo	**100,0**	49,6	18,1	31,4	50,4
Relativo respecto a las Variables utilizadas					
Total Errores Personales	**100,0**	**100,0**	**100,0**	**100,0**	**100,0**
Sexo	0,1	0,1	0,4	0,0	0,0
Edad	12,5	10,8	7,8	12,4	14,1
Lugar de Nacimiento	41,2	41,8	41,5	41,9	40,7
Estado Conyugal	2,1	2,5	4,1	1,6	1,7
Nivel Educativo	44,1	44,8	46,3	44,0	43,4

Nota: El registro para una persona podría contener de uno a cinco errores, es decir, tantos errores como variables analizadas, por lo que el total de errores no es un total de personas; aunque para cada variable en particular, sí lo es.

Cuadro A 2.3
COSTA RICA: Errores encontrados en los Hogares Coincidentes
SEGÚN: Hogares coincidentes con o sin error y
Características de la vivienda objeto de análisis
POR: Coincidencias en el informante
2004 - 2005 (Valores absolutos y relativos)
(Valores absolutos y relativos)

Variables	Total	Coincidencias en el Informante	
		Mismo Informante	Diferente Informante
Total de Hogares Coincidentes	**3.178**	**1.879**	**1.299**
Con errores en vivienda	434	247	187
Sin errores en vivienda	2.744	1.632	1.112
Total Errores de Vivienda	**487**	**279**	**208**
Tenencia de vivienda	104	54	50
Material de paredes	114	69	45
Material de piso	64	34	30
Abastecimiento de agua	86	50	36
Procedencia de agua	62	37	25
Servicio sanitario	52	32	20
Luz eléctrica	5	3	2
Relativo respecto al Tipo de informante			
Total Errores de Vivienda	**100,0**	57,3	42,7
Tenencia de vivienda	**100,0**	51,9	48,1
Material de paredes	**100,0**	60,5	39,5
Material de piso	**100,0**	53,1	46,9
Abastecimiento de agua	**100,0**	58,1	41,9
Procedencia de agua	**100,0**	59,7	40,3
Servicio sanitario	**100,0**	61,5	38,5
Luz eléctrica	**100,0**	60,0	40,0
Relativo respecto a las Variables utilizadas			
Total Errores de Vivienda	**100,0**	**100,0**	**100,0**
Tenencia de vivienda	21,4	19,4	24,0
Material de paredes	23,4	24,7	21,6
Material de piso	13,1	12,2	14,4
Abastecimiento de agua	17,7	17,9	17,3
Procedencia de agua	12,7	13,3	12,0
Servicio sanitario	10,7	11,5	9,6
Luz eléctrica	1,0	1,1	1,0

Nota: El registro para un hogar podría contener de uno a siete errores, es decir, tantos errores como variables analizadas sean, por lo que el total de errores no es un total de hogares; aunque para cada variable en particular, sí lo es.

B. Regresión Logística

Los resultados de la prueba ómnibus, que es una prueba de ajuste global, informan paso a paso, de las variaciones producidas en el ajuste como consecuencia de la incorporación (o eliminación) de cada variable. El estadístico utilizado es el Chi-cuadrado, el cual permite contrastar la hipótesis de que la mejora obtenida en el ajuste es nula [Pardo y Ruiz, 2000:668].

En el primer paso se incluye la variable *Tamaño_del_hogar* y su inclusión supone una mejora significativa del ajuste (Sig.=0,000). En el segundo paso se incorpora a la anterior, la variable *Tipo de informante*, su inclusión también supone una mejora significativa del ajuste respecto al paso anterior. El modelo finalmente consta de 3 pasos y es en éste donde se consigue el mejor ajuste, introduciendo la variable *Nivel de educación*.

Prueba Omnibus de los coeficientes del modelo (contrastes de ajuste global)

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	52,417	1	0,000
	Bloque	52,417	1	0,000
	Modelo	52,417	1	0,000
Paso 2	Paso	16,493	1	0,000
	Bloque	68,909	2	0,000
	Modelo	68,909	2	0,000
Paso 3	Paso	6,379	1	0,012
	Bloque	75,289	3	0,000
	Modelo	75,289	3	0,000

En los estadísticos de ajuste global adjuntos, también se puede apreciar que el ajuste global del modelo (R) mejora en cada paso, al comprobar que el valor de la razón de verosimilitudes (-2 log de la verosimilitud) disminuye paulatinamente.

Resumen de los modelos
(estadísticos de ajuste global)

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	1752,800 (a)	0,034	0,049
2	1736,307 (a)	0,044	0,064
3	1729,928 (a)	0,048	0,070

a La estimacón terminó en la iteración número 4 porque los parámetros estimados cambiaron a menos de 0,001.

En la tabla de clasificación de los casos se cruza el resultado observado en la variable dependiente con el resultado pronosticado por el modelo respectivo. De esta manera, la tabla muestra el porcentaje de casos correctamente clasificados en cada uno de los grupos. El punto de corte utilizado (0,71) permite acercar los porcentajes de clasificación correcta de las dos categorías. 56,2% en la categoría No error y 62,7% en la categoría de Error.

Tabla de clasificación (a)

Observado			Pronosticado		
			Error_dicotómico		Porcentaje correcto
			No error	Error	
Paso 1	Error_dicotómico	No error	286	141	67,0
		Error	574	519	47,5
	Porcentaje global				**53,0**
Paso 2	Error_dicotómico	No error	240	187	56,2
		Error	409	684	62,6
	Porcentaje global				**60,8**
Paso 3	Error_dicotómico	No error	240	187	56,2
		Error	408	685	62,7
	Porcentaje global				**60,9**

a El valor de corte es ,710

La tabla siguiente informa sobre las variables incorporadas al modelo en cada uno de los pasos. También reporta las estimaciones de los coeficientes y su significación. El paso que interesa es el último, pues contiene el modelo final. De las 9 variables independientes seleccionadas para el análisis, desde el método por pasos, se han seleccionado tres y la constante queda excluida por carecer de significancia.

Variables en la ecuación

		B	Error estándar	Wald	gl	Sig.	Exp(B)	I.C. 95,0% para EXP(B)	
								Inferior	Superior
Paso 1 (a)	Tamano_hogar	0,252	0,037	47,205	1	0,000	1,287	1,197	1,383
	Constante	0,139	0,125	1,236	1	0,266	1,149		
Paso 2 (b)	Tipo_informante	-0,520	0,130	15,984	1	0,000	0,594	0,461	0,767
	Tamano_hogar	0,220	0,037	35,008	1	0,000	1,246	1,158	1,340
	Constante	0,588	0,168	12,241	1	0,000	1,801		
Paso 3 (c)	Tipo_informante	-0,507	0,130	15,086	1	0,000	0,603	0,467	0,778
	Nivel_educacion	0,032	0,013	6,281	1	0,012	1,032	1,007	1,059
	Tamano_hogar	0,216	0,037	34,190	1	0,000	1,241	1,154	1,334
	Constante	0,349	0,192	3,302	1	0,069	1,418		

a Variable(s) introducida(s) en el paso 1: Tamano_hogar.
b Variable(s) introducida(s) en el paso 2: Tipo_informante.
c Variable(s) introducida(s) en el paso 3: Nivel_educacion.

Para calcular el estadístico chi-cuadrado de Hosmer y Lemeshow se divide la muestra en 10 grupos a partir de los deciles de las probabilidades pronosticadas, denominándoseles deciles de riesgo. En cada decil de riesgo se calcula el número de casos observados y el número de casos pronosticados. Ambos se comparan, en cada una de las 20 casillas definidas (2 categorías de la variable dependiente y 10 deciles de riesgo), mediante el estadístico chi-cuadrado de Pearson. Este estadístico permite contrastar la hipótesis de que la variable dependiente se distribuye de la misma manera en los 10 deciles de riesgo (o, lo que es

Prueba de Hosmer y Lemeshow

Paso	Chi-cuadrado	gl	Sig.
1	8,310	4	0,081
2	8,814	6	0,184
3	**12,424**	8	**0,133**

equivalente, que no existen diferencias entre las frecuencias observadas y las esperadas) [Pardo y Ruiz, 2002:684]. En este ejercicio, el estadístico chi-cuadrado toma el valor 12,424 en el tercer modelo, asociado a un nivel crítico (Sig.) de 0,133 por lo que no se puede rechazar la hipótesis nula de igualdad de distribuciones y, en consecuencia, se puede asumir que el modelo se ajusta a los datos.

Las correlaciones entre las estimaciones de las variables independientes [37/] deben ser pequeñas, pues una correlación elevada entre dos coeficientes debe interpretarse como un indicio de colinealidad. Cuando existe colinealidad, la estimación del coeficiente relativo a una variable puede estar demasiado afectada (sesgada) por la presencia de las otras variables [Pardo y Ruiz, 2002:686]. En el ejercicio presente, las correlaciones entre las diferentes variables son pequeñas. Esto indica que la relación de asociación entre cada par de variables es de tal magnitud que no cabe pensar en la existencia de colinealidades.

Matriz de correlaciones

		Constante	Tamano_ hogar	Tipo_ informante (1)	Nivel_ educacion
Paso 1	Constante	1			
	Tamano_hogar	-0,885	1		
Paso 2	Constante	1			
	Tamano_hogar	-0,771	1		
	Tipo_informante (1)	-0,672	0,193	1	
Paso 3	Constante	1			
	Tamano_hogar	-0,652	1		
	Tipo_informante (1)	-0,604	0,189	1	
	Nivel_educacion	-0,488	-0,039	0,035	1

[37/] El término constante no es más que un factor de escala que refleja la métrica del conjunto de variables independientes [Pardo y Ruiz, 2002:686].

La tabla adjunta muestra información sobre las variables no incluidas en los modelos en cada paso. El estadístico de puntuación de Rao [38/] permite apreciar cuál variable será incluida en el siguiente paso: aquella a la que corresponde el mayor estadístico de puntuación (siempre que éste sea significativo).

En el ejercicio realizado, se puede ver que, de las variables no incluidas en el primer paso (ya está incluida *Tamano_hogar*), la variable *Tipo_informante* es la que tiene el estadístico de puntuación más alto (16,168); como además es significativo (Sig.=0,000), ésa será la variable incorporada al modelo en el siguiente modelo [Pardo y Ruiz, 2002:670].

En el segundo paso es la

Variables que no están en la ecuación (a)

			Puntuación	gl	Sig.
Paso 1	Variables	Tipo_informante (1)	16,168	1	0,000
		Edad	0,001	1	0,976
		Nivel_educacion	7,238	1	0,007
		Estado_conyugal	9,201	5	0,101
		Estado_conyugal (1)	1,318	1	0,251
		Estado_conyugal (2)	0,107	1	0,743
		Estado_conyugal (3)	1,145	1	0,285
		Estado_conyugal (4)	2,049	1	0,152
		Estado_conyugal (5)	2,697	1	0,101
		Zona (1)	3,842	1	0,050
		Nivel_socioeconomico	7,160	5	0,209
		Nivel_socioeconomico (1)	0,002	1	0,966
		Nivel_socioeconomico (2)	0,810	1	0,368
		Nivel_socioeconomico (3)	2,168	1	0,141
		Nivel_socioeconomico (4)	0,691	1	0,406
		Nivel_socioeconomico (5)	3,842	1	0,050
		Ln_Ingreso_percapita_hogar	6,900	1	0,009
Paso 2	Variables	Edad	0,155	1	0,694
		Nivel_educacion	6,309	1	0,012
		Estado_conyugal	8,718	5	0,121
		Estado_conyugal (1)	0,544	1	0,461
		Estado_conyugal (2)	0,972	1	0,324
		Estado_conyugal (3)	1,891	1	0,169
		Estado_conyugal (4)	1,111	1	0,292
		Estado_conyugal (5)	1,493	1	0,222
		Zona (1)	3,913	1	0,048
		Nivel_socioeconomico	6,764	5	0,239
		Nivel_socioeconomico (1)	0,003	1	0,960
		Nivel_socioeconomico (2)	0,873	1	0,350
		Nivel_socioeconomico (3)	2,030	1	0,154
		Nivel_socioeconomico (4)	0,477	1	0,490
		Nivel_socioeconomico (5)	3,913	1	0,048
		Ln_Ingreso_percapita_hogar	6,068	1	0,014
Paso 3	Variables	Edad	2,333	1	0,127
		Estado_conyugal	6,884	5	0,229
		Estado_conyugal (1)	0,957	1	0,328
		Estado_conyugal (2)	1,548	1	0,213
		Estado_conyugal (3)	0,958	1	0,328
		Estado_conyugal (4)	0,871	1	0,351
		Estado_conyugal (5)	0,517	1	0,472
		Zona (1)	1,490	1	0,222
		Nivel_socioeconomico	4,056	5	0,541
		Nivel_socioeconomico (1)	0,000	1	0,999
		Nivel_socioeconomico (2)	0,037	1	0,847
		Nivel_socioeconomico (3)	1,158	1	0,282
		Nivel_socioeconomico (4)	0,101	1	0,751
		Nivel_socioeconomico (5)	1,490	1	0,222
		Ln_Ingreso_percapita_hogar	1,517	1	0,218

a No se calculan los chi-cuadrado residuales a causa de las redundancias.

[38] / Este procedimiento va incorporando aquellas variables cuya puntuación, siendo significativa, posee la probabilidad asociada más pequeña [Pardo y Ruiz, 2002:667].

variable *Nivel_educacion* la que queda seleccionada para ingresar al siguiente paso. En el tercer paso, ya no queda ninguna variable significativa, por lo que ahí concluye las iteraciones de los modelos.

C. Regresión lineal

Estadísticos descriptivos

	Media	Desviación típica	N
Error total	1,71	1,612	3178
Tipo de informante	0,59	0,492	3178
Sexo	0,33	0,472	3178
Edad	44,54	16,045	3178
Relación de parentesco	1,64	0,693	3178
Condición de actividad	2,01	0,987	3178
Nivel educación	8,20	4,249	3178
Estado conyugal	2,84	1,650	3178
Región	2,40	1,663	3178
Zona	0,43	0,495	3178
Tamaño del hogar	3,90	1,714	3178
Nivel socioeconómico	4,52	1,802	3178
Nivel de pobreza	2,68	0,590	3178
Ln Ingreso percápita hogar	10,8797	0,9320	3178

Resumen del modelo (f)

Modelo	R	R cuadrado	R cuadrado corregida	Error típico de la estimación	Estadísticos de cambio					Durbin-Watson
					Cambio en R cuadrado	Cambio en F	gl1	gl2	Sig. del cambio en F	
1	,285 (a)	0,081	0,081	1,545	0,081	280,846	1	3176	0,000	
2	,330 (b)	0,109	0,108	1,522	0,027	97,401	1	3175	0,000	
3	,345 (c)	0,119	0,118	1,514	0,010	36,433	1	3174	0,000	
4	,351 (d)	0,123	0,122	1,510	0,004	16,085	1	3173	0,000	
5	,356 (e)	0,127	0,125	1,507	0,003	12,273	1	3172	0,000	1,928

a Variables predictoras: (Constante), Tamaño del hogar
b Variables predictoras: (Constante), Tamaño del hogar, Tipo informante
c Variables predictoras: (Constante), Tamaño del hogar, Tipo informante, Ln Ingreso percápita hogar
d Variables predictoras: (Constante), Tamaño del hogar, Tipo informante, Ln Ingreso percápita hogar, Edad
e Variables predictoras: (Constante), Tamaño del hogar, Tipo informante, Ln Ingreso percápita hogar, Edad, Nivel de pobreza
f Variable dependiente: Error total

Coeficientes (a)

Modelo		Coeficientes no estandarizados		Coeficientes estanda-rizados	t	Sig.	Intervalo de confianza para B al 95%		Estadísticos de colinealidad		$\frac{1}{(1-R^2)}$
		B	Error típico	Beta			Límite inferior	Límite superior	Tolerancia $(1-R^2)$	VIF Factores de inflación de la varianza	
1	(Constante)	0,666	0,068		9,791	0	0,533	0,800			
	Tamaño del hogar	0,268	0,016	0,285	16,758	0	0,237	0,299	1	1	1
2	(Constante)	1,083	0,079		13,669	0	0,928	1,239			
	Tamaño del hogar	0,244	0,016	0,260	15,322	0	0,213	0,275	0,977	1,023	1,024
	Tipo de informante	-0,548	0,056	-0,167	-9,869	0	-0,657	-0,439	0,977	1,023	1,024
3	(Constante)	-0,947	0,345		-2,741	0,006	-1,624	-0,270			
	Tamaño del hogar	0,265	0,016	0,282	16,338	0	0,233	0,297	0,933	1,072	1,072
	Tipo de informante	-0,527	0,055	-0,161	-9,528	0	-0,636	-0,419	0,973	1,027	1,028
	Ln Ingreso percápita hogar	0,178	0,029	0,103	6,036	0	0,120	0,236	0,954	1,048	1,048
4	(Constante)	-1,451	0,367		-3,954	0	-2,170	-0,731			
	Tamaño del hogar	0,286	0,017	0,305	16,807	0	0,253	0,320	0,841	1,189	1,189
	Tipo de informante	-0,554	0,056	-0,169	-9,955	0	-0,663	-0,445	0,960	1,042	1,042
	Ln Ingreso percápita hogar	0,189	0,030	0,109	6,389	0	0,131	0,247	0,946	1,057	1,057
	Edad	0,007	0,002	0,071	4,011	0	0,004	0,011	0,877	1,140	1,140
5	(Constante)	-2,041	0,403		-5,064	0	-2,832	-1,251			
	Tamaño del hogar	0,291	0,017	0,310	17,066	0	0,258	0,325	0,836	1,197	1,196
	Tipo de informante	-0,547	0,056	-0,167	-9,848	0	-0,656	-0,438	0,959	1,043	1,043
	Ln Ingreso percápita hogar	0,301	0,044	0,174	6,912	0	0,216	0,386	0,434	2,305	2,304
	Edad	0,007	0,002	0,067	3,787	0	0,003	0,010	0,873	1,145	1,145
	Nivel de pobreza	-0,237	0,068	-0,087	-3,503	0	-0,370	-0,104	0,450	2,224	2,222

a Variable dependiente: Error total

ANOVA (f)

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	670,394	1	670,394	280,846	,000 (a)
	Residual	7581,275	3176	2,387		
	Total	8251,669	3177			
2	Regresión	896,046	2	448,023	193,386	,000 (b)
	Residual	7355,623	3175	2,317		
	Total	8251,669	3177			
3	Regresión	979,520	3	326,507	142,507	,000 (c)
	Residual	7272,149	3174	2,291		
	Total	8251,669	3177			
4	Regresión	1016,199	4	254,050	111,409	,000 (d)
	Residual	7235,470	3173	2,280		
	Total	8251,669	3177			
5	Regresión	1044,086	5	208,817	91,899	,000 (e)
	Residual	7207,583	3172	2,272		
	Total	8251,669	3177			

a Variables predictoras: (Constante), Tamaño del hogar

b Variables predictoras: (Constante), Tamaño del hogar, Tipo informante

c Variables predictoras: (Constante), Tamaño del hogar, Tipo informante, Ln Ingreso percápita hogar

d Variables predictoras: (Constante), Tamaño del hogar, Tipo informante, Ln Ingreso percápita hogar, Edad

e Variables predictoras: (Constante), Tamaño del hogar, Tipo informante, Ln Ingreso percápita hogar, Edad, Nivel de pobreza

f Variable dependiente: Error total

Correlaciones

	Error total	Tipo de informante	Sexo	Edad	Relación de parentesco	Condición de actividad	Nivel educación	Estado conyugal	Región	Zona	Tamaño del hogar	Nivel socio-económico	Nivel de pobreza	Ln Ingreso percápita hogar
Correlación de Pearson														
Error total	1													
Tipo de informante	-0,207	1												
Sexo	0,019	-0,145	1											
Edad	-0,057	0,163	0,124	1										
Relación de parentesco	0,107	-0,168	-0,371	-0,502	1									
Condición de actividad	-0,038	0,102	-0,384	0,130	0,265	1								
Nivel educación	0,048	-0,066	0,000	-0,323	0,136	-0,211	1							
Estado conyugal	-0,055	-0,027	-0,013	-0,038	0,172	0,033	0,053	1						
Región	-0,015	0,007	0,054	-0,035	-0,041	-0,005	-0,183	-0,069	1					
Zona	0,057	-0,027	-0,049	0,052	0,027	-0,045	0,320	0,083	-0,240	1				
Tamaño del hogar	0,285	-0,151	-0,091	-0,320	0,289	0,023	0,027	-0,225	0,015	-0,059	1			
Nivel socioeconómico	-0,050	0,023	0,059	-0,034	-0,027	0,034	-0,244	-0,072	0,191	-0,953	0,046	1		
Nivel de pobreza	0,013	-0,015	0,018	-0,079	0,076	-0,141	0,286	0,016	-0,119	0,124	-0,083	-0,086	1	
Ln Ingreso percápita hogar	0,050	-0,029	0,056	-0,025	0,041	-0,209	0,519	0,081	-0,200	0,355	-0,205	-0,275	0,737	1

Diagnósticos de colinealidad (a)

Modelo	Dimensión	Autovalor (eigenvalue)	Indice de condición	Proporciones de la varianza					
				(Constante)	Tamaño del hogar	Tipo informante	Ln Ingreso percápita hogar	Edad	Nivel de pobreza
1	1	1,915	1	0,04	0,04				
	2	0,085	4,757	0,96	0,96				
2	1	2,571	1	0,02	0,02	0,05			
	2	0,359	2,675	0,02	0,12	0,74			
	3	0,070	6,054	0,97	0,85	0,21			
3	1	3,519	1	0	0,01	0,02	0		
	2	0,372	3,074	0	0,08	0,81	0		
	3	0,105	5,787	0,01	0,81	0,15	0,02		
	4	0,003	32,821	0,99	0,10	0,02	0,98		
4	1	4,402	1	0	0,01	0,01	0	0	
	2	0,373	3,437	0	0,06	0,81	0	0	
	3	0,169	5,101	0	0,38	0,16	0	0,26	
	4	0,053	9,111	0,02	0,41	0,01	0,04	0,68	
	5	0,003	37,810	0,98	0,14	0,01	0,96	0,06	
5	1	5,346	1	0	0	0,01	0	0	0
	2	0,382	3,741	0	0,05	**0,83**	0	0	0
	3	0,172	5,582	0	**0,43**	0,14	0	0,21	0
	4	0,080	8,155	0	0,24	0	0	**0,53**	0,09
	5	0,018	17,338	0,10	0,18	0,01	0,02	0,24	**0,48**
	6	0,002	53,465	0,90	0,10	0,01	**0,97**	0,02	0,43

a Variable dependiente: Error total

Histograma de los Residuos estandarizados

Variable dependiente: Error total

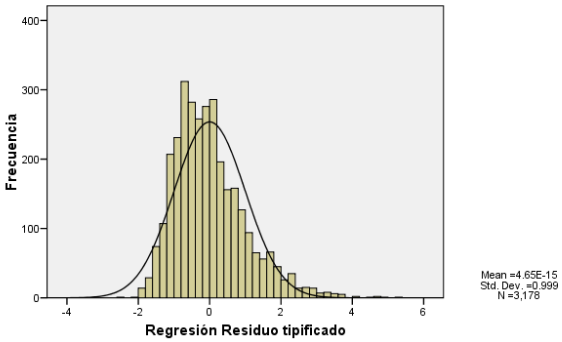


Gráfico P-P normal de regresión Residuo tipificado

Variable dependiente: Error total

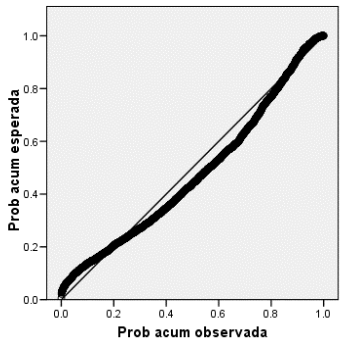
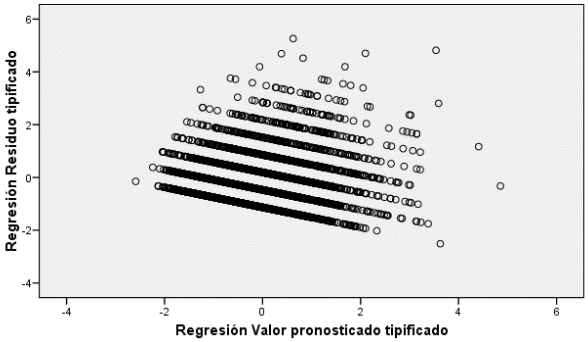


Gráfico de dispersión Pronósticos estandarizados por Residuos estandarizados

Variable dependiente: Error total



D. Sintaxis de los Errores no muestrales

a. Recodificaciones y cálculos

```
* Recodificación del Nivel de Educación *
RECODE
  b08
  (0=0)     (1=1)     (2=1)
  (11=2)    (12=3)    (13=4)    (14=5)   (15=6)
                                 (16=7)   (19=2)
  (21=8)    (22=9)    (23=10)  (24=11)  (25=12)
                                        (29=8)
  (31=8)    (32=9)    (33=10)  (34=11)  (35=12)
                                 (36=13)  (39=8)
  (41=13)  (42=14)  (43=15)  (44=16)  (49=13)
  (51=13)  (52=14)  (53=15)  (54=16)  (55=17)
           (56=18)  (57=19)  (58=20)  (59=13)
  (99=99)
  INTO  Educa_re .
VARIABLE LABELS
     Educa_re 'Nivel educación recodificada'.
EXECUTE .

* Creación de los estratos *
IF (nivel = 1 & region = 1) Estrato = 11 .
IF (nivel = 1 & region = 2) Estrato = 12 .
IF (nivel = 1 & region = 3) Estrato = 13 .
IF (nivel = 1 & region = 4) Estrato = 14 .
IF (nivel = 1 & region = 5) Estrato = 15 .
IF (nivel = 1 & region = 6) Estrato = 16 .
IF (nivel = 2 & region = 1) Estrato = 21 .
IF (nivel = 2 & region = 2) Estrato = 22 .
IF (nivel = 2 & region = 3) Estrato = 23 .
IF (nivel = 2 & region = 4) Estrato = 24 .
IF (nivel = 2 & region = 5) Estrato = 25 .
IF (nivel = 2 & region = 6) Estrato = 26 .
IF (nivel = 3 & region = 1) Estrato = 31 .
IF (nivel = 3 & region = 2) Estrato = 32 .
IF (nivel = 3 & region = 3) Estrato = 33 .
IF (nivel = 3 & region = 4) Estrato = 34 .
IF (nivel = 3 & region = 5) Estrato = 35 .
IF (nivel = 3 & region = 6) Estrato = 36 .
IF (nivel = 4 & region = 1) Estrato = 41 .
IF (nivel = 4 & region = 2) Estrato = 42 .
IF (nivel = 4 & region = 3) Estrato = 43 .
IF (nivel = 4 & region = 4) Estrato = 44 .
IF (nivel = 4 & region = 5) Estrato = 45 .
IF (nivel = 4 & region = 6) Estrato = 46 .
IF (nivel = 5 & region = 1) Estrato = 51 .
IF (nivel = 5 & region = 2) Estrato = 52 .
IF (nivel = 5 & region = 3) Estrato = 53 .
IF (nivel = 5 & region = 4) Estrato = 54 .
IF (nivel = 5 & region = 5) Estrato = 55 .
IF (nivel = 5 & region = 6) Estrato = 56 .
IF (nivel = 6 & region = 1) Estrato = 61 .
IF (nivel = 6 & region = 2) Estrato = 62 .
IF (nivel = 6 & region = 3) Estrato = 63 .
IF (nivel = 6 & region = 4) Estrato = 64 .
IF (nivel = 6 & region = 5) Estrato = 65 .
IF (nivel = 6 & region = 6) Estrato = 66 .
EXECUTE .

* Determinación de los informantes sección B en
     el 2004 *
   * Informante_04 = informante de la sección b *
   * lininfo_b_04 = línea del miembro del hogar
     que proporcionó la información *
   * linea_b_2004 = línea de registro actual *
COMPUTE Informante_04 = 0 .
EXECUTE .
IF (lininfo_b_04 = linea_b_2004)
     Informante_04 = 1 .
EXECUTE .

* Determinación de los informantes sección B en
     el 2005 *
   * Informante_05 = informante de la sección b *
   * lininfo_b_05 = línea del miembro del hogar
     que proporcionó la información *
   * linea_b_2005 = línea de registro actual *
COMPUTE Informante_05 = 0 .
EXECUTE .
IF (lininfo_b_05 = linea_b_2005)
    Informante_05 = 1 .
EXECUTE .

* Cálculo de Autoinformantes sección B en el
     2004 *
    * Autoinformante_b_04 = Autoinformante de
      la sección b en el 2004 *
COMPUTE Autoinformante_b_04 = 0 .
EXECUTE .
IF ( lininfo_b_04 = linea_b_2004 )
   Autoinformante_b_04 = 1 .
EXECUTE .
```

```
* Cálculo de Autoinformantes sección B en el
      2005 *
    * Autoinformante_b_05 = Autoinformante de
      la sección b en el 2005 *
COMPUTE Autoinformante_b_05 = 0 .
EXECUTE .
IF ( lininfo_b_05 = linea_b_2005 )
   Autoinformante_b_05 = 1 .
EXECUTE .

* Cálculo de los mismos Autoinformantes en el
      2004 y 2005 *
    * MismoAutoinf_b = Misma persona que
      proporcionó los datos en ambos años *
COMPUTE MismoAutoinf_b = 0 .
EXECUTE .
IF ( Autoinformante_b_04  = 1  &
     Autoinformante_b_05 = 1 )
   MismoAutoinf_b = 1 .
EXECUTE .

*  Relación de parentesco  informante 04 *
COMPUTE RelParen_infor_04 = 0 .
EXECUTE .
IF (Informante_04 = 1)
     RelParen_infor_04 = b03_04 .
EXECUTE .

* Relación de parentesco informante 05 *
COMPUTE RelParen_infor_05 = 0 .
EXECUTE .
IF (Informante_05 = 1)
    RelParen_infor_05 = b03_05 .
EXECUTE .

* Edad del informante 04 *
COMPUTE Edad_infor_04 = 0 .
EXECUTE .
IF (Informante_04 = 1)
    Edad_infor_04 = b05_04 .
EXECUTE .

* Edad del informante 05 *
COMPUTE Edad_infor_05 = 0 .
EXECUTE .
IF  (Informante_05 = 1)
    Edad_infor_05 = b05_05 .
EXECUTE .

* Sexo del informante 04 *
COMPUTE Sexo_infor_04 = 0 .
EXECUTE .
IF (Informante_04 = 1)
    Sexo_infor_04 = b04_04 .
EXECUTE .

* Sexo del informante 05 *
COMPUTE Sexo_infor_05 = 0 .
EXECUTE .
IF (Informante_05 = 1)
    Sexo_infor_05 = b04_05 .
EXECUTE .

* Recodificación Condición de Actividad 2004 *
RECODE
  condact_04
    (0=0)  (1=1)  (2 thru 7=2)  (8 thru 14=3)  (9=9)
INTO cond_act_04 .
VARIABLE LABELS
  cond_act_04 'Condición de actividad 04'.
EXECUTE .

* Recodificación de la Condición de Actividad
      2005 *
RECODE
  condact_05
     (0=0)  (1=1)  (2 thru 7=2)
     (8 thru 14=3) (9=9)
  INTO  cond_act_05 .
VARIABLE LABELS cond_act_05
     'Condición de actividad 05'.
EXECUTE .

* Condición de Actividad del informante 04 *
COMPUTE CondAct_infor_04 = 0 .
EXECUTE .
IF (Informante_04 = 1)
    CondAct_infor_04 = cond_act_04 .
EXECUTE .

* Condición de Actividad del informante 05 *
COMPUTE CondAct_infor_05 = 0 .
EXECUTE .
IF(Informante_05 = 1)
    CondAct_infor_05 = cond_act_05 .
EXECUTE .

* Recodificación del Nivel de Educación 04 *
RECODE
  b08_04
    (0=0)     (1=1)     (2=1)
    (11=2)    (12=3)    (13=4)    (14=5)    (15=6)
```

```
                                (16=7)  (19=2)
   (21=8)  (22=9)  (23=10) (24=11) (25=12)
                                   (29=8)
   (31=8)  (32=9)  (33=10) (34=11) (35=12)
                                (36=13) (39=8)
   (41=13) (42=14) (43=15) (44=16) (49=13)
   (51=13) (52=14) (53=15) (54=16) (55=17)
           (56=18) (57=19) (58=20) (59=13)
   (99=99)
  INTO  Educa_re_04 .
VARIABLE LABELS
   Educa_re_04 'Nivel educación recodificada'.
EXECUTE .

* Recodificación del Nivel de Educación 05 *
RECODE
  b08_05
   (0=0)   (1=1)   (2=1)
   (11=2)  (12=3)  (13=4)  (14=5)  (15=6)
                            (16=7)  (19=2)
   (21=8)  (22=9)  (23=10) (24=11) (25=12)
                                   (29=8)
   (31=8)  (32=9)  (33=10) (34=11) (35=12)
                            (36=13) (39=8)
   (41=13) (42=14) (43=15) (44=16) (49=13)
   (51=13) (52=14) (53=15) (54=16) (55=17)
           (56=18) (57=19) (58=20) (59=13)
   (99=99)
  INTO  Educa_re_05 .
VARIABLE LABELS
   Educa_re_04 'Nivel educación recodificada'.
EXECUTE .

* Nivel de Educación del informante 04 *
COMPUTE NivelEduc_infor_04 = 0 .
EXECUTE .
IF (Informante_04 = 1)
   NivelEduc_infor_04 = b08_04 .
EXECUTE .
* Educación del informante 04 recodificada *
COMPUTE Educ_re_infor_04 = 0 .
EXECUTE .
IF (Informante_04 = 1)
   Educ_re_infor_04 = Educa_re_04 .
EXECUTE .

* Nivel de Educación del informante 05 *
COMPUTE NivelEduc_infor_05 = 0 .
EXECUTE .
IF(Informante_05 = 1)
   NivelEduc_infor_05 = b08_05 .
EXECUTE .
* Educación del informante 05 recodificada *
COMPUTE Educ_re_infor_05 = 0 .
EXECUTE .
IF (Informante_05 = 1)
   Educ_re_infor_05 = Educa_re_05 .
EXECUTE .

* Estado Conyugal del informante 04 *
COMPUTE EstConyug_infor_04 = 0 .
EXECUTE .
IF (Informante_04 = 1)
   EstConyug_infor_04 = b13_04 .
EXECUTE .

* Estado Conyugal del informante 05 *
COMPUTE EstConyug_infor_05 = 0 .
EXECUTE .
IF (Informante_05 = 1)
   EstConyug_infor_05 = b15_05 .
EXECUTE .

* Llave hogar para el año 2005 *
COMPUTE Llave_diferente_05 =
     segmento_05 * 10000
     + vivienda_2005 * 10
     + hogar_2005 .
EXECUTE .

* Cálculo de variables del informante para todo
el hogar *
AGGREGATE
 /OUTFILE=*
 MODE=ADDVARIABLES
 /BREAK=llave_hogar_2005
 /RelParen_infor_04_max =
                    MAX(RelPare_infor_04)
 /RelParen_infor_05_max =
                    MAX(RelPare_infor_05)
 /Edad_infor_04_max = MAX(Edad_infor_04)
 /Edad_infor_05_max = MAX(Edad_infor_05)
 /Sexo_infor_04_max = MAX(Sexo_infor_04)
 /Sexo_infor_05_max = MAX(Sexo_infor_05)
 /CondAct_infor_04_max =
                    MAX(CondAct_infor_04)
 /CondAct_infor_05_max =
                    MAX(CondAct_infor_05)
 /NivelEduc_infor_04_max =
                   MAX(NivelEduc_infor_04)
 /NivelEduc_infor_05_max =
                   MAX(NivelEduc_infor_05)
```

```
 /NivelEduc_re_infor_04_max =
                    MAX(NivelEduc_re_infor_04)
 /NivelEduc_re_infor_05_max =
                    MAX(NivelEduc_re_infor_05)
 /EstConyug_infor_04_max =
                      MAX(EstConyug_infor_04)
 /EstConyug_infor_05_max =
                      MAX(EstConyug_infor_05)

* Errores en Sexo = 1 *
COMPUTE Error_Sexo = 1 .
EXECUTE .
IF ( b04_04 = b04_05 )
   Error_Sexo = 0 .
EXECUTE .

* Errores en Edad (hasta 2 años de más) =1 *
COMPUTE Error_Edad = 1 .
EXECUTE .
IF ( b05_05 = b05_04
     | b05_05 = ( b05_04 + 1 )
     | b05_05 = ( b05_04 + 2 ) )
   Error_Edad = 0 .
EXECUTE .
* Tratamiento de los ignorados *
IF ( b05_04 = 98
     | b05_04 = 99
     | b05_05 = 98
     | b05_05 = 99 )
   Error_Edad = 0 .
EXECUTE .

* Errores en Lugar de Nacimiento = 1 *
COMPUTE Error_LugarNacim = 1 .
EXECUTE .
IF ( b06_05 = b06_04 )
   Error_LugarNacim = 0 .
EXECUTE .
* Tratamiento de los ignorados *
IF ( b06_04 = 999 | b06_05 = 999 )
   Error_LugarNacim = 0 .
EXECUTE .

* Errores en Estado Conyugal = 1 *
COMPUTE Error_EstadoConyug = 0 .
EXECUTE .
IF ( b15_05 = 6
     & ( b13_04 = 1
         | b13_04 = 2
         | b13_04 = 3
         | b13_04 = 4
         | b13_04 = 5 ) )
   Error_EstadoConyug = 1 .
EXECUTE .
* Tratamiento de los ignorados *
IF ( b13_04 = 9 | b15_05 = 9 )
   Error_EstadoConyug = 0 .
EXECUTE .

* Errores en Nivel de Educación (con un año de
    más) = 1 *
COMPUTE Error_NivelEduc = 1 .
EXECUTE .
** Mismo grado académico en ambos años (6.897
   casos) **
   IF ( b08_04 = b08_05 )
      Error_NivelEduc = 0 .
   EXECUTE .
** Ningún año, preparatoria y enseñanza especial
     en 2004 (7.410 casos) **
   * Avanzaron al siguiente nivel *
   IF ( ( b08_04 = 0  &  b08_05 = 1 )
         | ( b08_04 = 1  &  b08_05 = 11 )
         | ( b08_04 = 2  &  b08_05 = 11 )  )
      Error_NivelEduc = 0 .
   EXECUTE .
** Primaria conocida en el 2004 (coincidencias =
    9.694 casos) **
   * Avanzaron al siguiente nivel *
   IF ( ( b08_04 = 11  &  b08_05 = 12 )
         | ( b08_04 = 12  &  b08_05= 13 )
         | ( b08_04 = 13  &  b08_05 = 14 )
         | ( b08_04 = 14  &  b08_05 = 15 )
         | ( b08_04 = 15  &  b08_05 = 16 )
         | ( b08_04 = 16  &  b08_05 = 21 )
         | ( b08_04 = 16  &  b08_05 = 31 )  )
      Error_NivelEduc = 0 .
   EXECUTE .
** Primaria desconocida en el 2004
     (coincidencias = 9.696 casos) **
   IF ( b08_04 = 19
        &  ( b08_05 = 11
             | b08_05 = 12
             | b08_05 = 13
             | b08_05 = 14
             | b08_05 = 15
             | b08_05 = 16
             | b08_05 = 21
             | b08_05 = 31  )  )
      Error_NivelEduc = 0 .
   EXECUTE .
```

```
** Secundaria académica conocida en el 2004
    (coincidencias = 10.658 casos) **
   * Avanzaron al siguiente nivel *
   IF ( ( b08_04 = 21  &  b08_05 = 22 )
           | ( b08_04 = 22  &  b08_05 = 23 )
           | ( b08_04 = 23  &  b08_05 = 24 )
           | ( b08_04 = 24  &  b08_05 = 25 )
           | ( b08_04 = 25  &  b08_05 = 41 )
           | ( b08_04 = 25  &  b08_05 = 51 )  )
      Error_NivelEduc = 0 .
   EXECUTE .
** Secundaria técnica conocida en el 2004
    (coincidencias = 10.740 casos) **
   * Avanzaron al siguiente nivel *
   IF ( ( b08_04 = 31  &  b08_05 = 32 )
           | ( b08_04 = 32  &  b08_05 = 33 )
           | ( b08_04 = 33  &  b08_05 = 34 )
           | ( b08_04 = 34  &  b08_05 = 35 )
           | ( b08_04 = 35  &  b08_05 = 36 )
           | ( b08_04 = 36  &  b08_05 = 41 )
           | ( b08_04 = 36  &  b08_05 = 51 )  )
      Error_NivelEduc = 0 .
   EXECUTE .
** Secundaria académica y técnica desconocida
    en el 2004 (coincidencias = 10.746 casos) **
   IF ( ( b08_04 = 29  |  b08_04 = 39 )
          &  ( b08_05 = 21
                 | b08_05 = 22
                 | b08_05 = 23
                 | b08_05 = 24
                 | b08_05 = 25
                 | b08_05 = 31
                 | b08_05 = 32
                 | b08_05 = 33
                 | b08_05 = 34
                 | b08_05 = 35
                 | b08_05 = 36
                 | b08_05 = 41
                 | b08_05 = 51  )  )
      Error_NivelEduc = 0 .
   EXECUTE .
** Traslados entre Secundaria académica y
    técnica (coincidencias = 10.973 casos) **
  IF ( ( b08_04 = 21
              & (b08_05 = 31 | b08_05 = 32 ) )
        | ( b08_04 = 22
              &  ( b08_05 = 32  |  b08_05 = 33 ) )
        | ( b08_04 = 23
              &  ( b08_05 = 33  |  b08_05 = 34 ) )
        | ( b08_04 = 24
              &  ( b08_05 = 34  |  b08_05 = 35 ) )
        | ( b08_04 = 25
              &  ( b08_05 = 35  |  b08_05 = 36 ) )
        | ( b08_04 = 31
              &  ( b08_05 = 21  |  b08_05 = 22 ) )
        | ( b08_04 = 32
              &  ( b08_05 = 22  |  b08_05 = 23 ) )
        | ( b08_04 = 33
              &  ( b08_05 = 23  |  b08_05 = 24 ) )
        | ( b08_04 = 34
              &  ( b08_05 = 24  |  b08_05 = 25 ) )
        | ( b08_04 = 35  &   b08_05 = 25  ) )
      Error_NivelEduc = 0 .
   EXECUTE .
** Parauniversitaria conocida en el 2004
    (coincidencias = 10.978 casos) **
   IF ( ( b08_04 = 41
              &  ( b08_05 = 42  |  b08_05 = 51 ) )
        | ( b08_04 = 42
              &  ( b08_05 = 43  |  b08_05 = 51 ) )
        | ( b08_04 = 43
              &  ( b08_05 = 44  |  b08_05 = 51 ) )
        | ( b08_04 = 44  &   b08_05 = 51  )  )
      Error_NivelEduc = 0 .
   EXECUTE .
** Parauniversitaria desconocida en el 2004
    (coincidencias = 10.978) **
   IF ( b08_04 = 49
          & ( b08_05 = 41
                | b08_05 = 42
                | b08_05 = 43
                | b08_05 = 44
                | b08_05 = 51  )  )
      Error_NivelEduc = 0 .
   EXECUTE .
** Universitaria conocida en el 2004
    (coincidencias = 11.279 casos) **
   IF ( ( b08_04 = 51  &  b08_05 = 52 )
           | ( b08_04 = 52  &  b08_05 = 53 )
           | ( b08_04 = 53  &  b08_05 = 54 )
           | ( b08_04 = 54  &  b08_05 = 55 )
           | ( b08_04 = 55  &  b08_05 = 56 )
           | ( b08_04 = 56  &  b08_05 = 57 )
           | ( b08_04 = 57  &  b08_05 = 58 )  )
      Error_NivelEduc = 0 .
   EXECUTE .
** Universitaria desconocida en el 2004
    (coincidencias = 11.290 casos) **
   IF ( b08_04 = 59
          &  ( b08_05 = 51
                 | b08_05 = 52
                 | b08_05 = 53
```

```
                | b08_05 = 54
                | b08_05 = 55
                | b08_05 = 56
                | b08_05 = 57
                | b08_05 = 58
                | b08_05 = 41  )  )
      Error_NivelEduc = 0 .
   EXECUTE .
** Tratamiento de la Educación desconocida**
IF ( b08_04 = 99  |  b08_05 = 99 )
   Error_NivelEduc = 0 .
EXECUTE .
** Tratamiento de los menores de 5 años en 2004
**
IF ( b05_04 < 5 )
   Error_NivelEduc = 0 .
EXECUTE .

* Suma de errores personales (para cada
      registro) *
COMPUTE Error_Total_Persona = 0 .
EXECUTE .
COMPUTE
Error_Total_Persona =
    ( Error_Sexo
      + Error_Edad
      + Error_LugarNacim
      + Error_EstadoConyug
      + Error_NivelEduc ) .
EXECUTE .

* Errores en Tenencia de Vivienda = 1 *
COMPUTE Error_TenenViv = 0 .
EXECUTE .
IF ( ( v02_04 = 1 | v02_04 = 2 )
        &  ( v02_05 >= 3) )
   Error_TenenViv = 1 .
EXECUTE .
* Tratamiento de los ignorados *
IF ( ( v02_04 = 9  |  v02_05 = 9 ) )
   Error_TenenViv = 0 .
EXECUTE .

* Errores en Material Predominante en Paredes
      Exteriores *
COMPUTE Error_MaterialPared = 0 .
EXECUTE .
IF ( ( v03_04 = 1  |  v03_04 = 2  |  v03_04 = 4 )
       &  ( v03_05 = 3
              | v03_05 = 5
              | v03_05 = 6
              | v03_05 = 7 ) )
  Error_MaterialPared = 1 .
EXECUTE .
* Tratamiento de los ignorados *
IF ( v03_04 = 9  |  v03_05 = 9 )
   Error_MaterialPared = 0 .
EXECUTE .

* Errores en Material Predominante en el Piso *
COMPUTE Error_MaterialPiso = 0 .
EXECUTE .
IF ( ( v06_04 = 1  |  v06_04 = 2 )
   &  ( v06_05 = 3  |  v06_05 = 4  |  v06_05 = 5 ) )
   Error_MaterialPiso = 1 .
EXECUTE .
* Tratamiento de los ignorados *
IF ( v06_04 = 9  |  v06_05 = 9 )
   Error_MaterialPiso = 0 .
EXECUTE .

* Errores en Abastecimiento de Agua *
COMPUTE Error_AbastecimAgua = 0 .
EXECUTE .
IF ( v10_04 = 1  &  ( v10_05 ~= 1 ) )
   Error_AbastecimAgua = 1 .
EXECUTE .
* Tratamiento de los ignorados *
IF ( v10_04 = 9  |  v10_05 = 9 )
   Error_AbastecimAgua = 0 .
EXECUTE .

* Errores en Procedencia del Agua *
COMPUTE Error_ProcedenAgua = 0 .
EXECUTE .
IF ( ( v11_04 >= 1  &  v11_04 <= 4 )
       &  ( v11_05 >= 5  &  v11_05 <= 7 ) )
   Error_ProcedenAgua = 1 .
EXECUTE .
* Tratamiento de los ignorados *
IF ( v11_04 = 9  |  v11_05 = 9 )
   Error_ProcedenAgua = 0 .
EXECUTE .

* Errores en Tenencia de Servicio Sanitario *
COMPUTE Error_ServicioSanitario = 0 .
EXECUTE .
IF ( ( v12_04  = 1  |  v12_04  = 2 )
       &  ( v12_05 >= 3  &  v12_05 <= 5 ) )
   Error_ServicioSanitario = 1 .
EXECUTE .
* Tratamiento de los ignorados *
```

```
IF ( v12_04 = 9  |  v12_05 = 9 )
   Error_ServicioSanitario = 0 .
EXECUTE .

* Errores en Tenencia de Luz Eléctrica *
COMPUTE Error_LuzEléctrica = 0 .
EXECUTE .
IF ( ( v14_04 >= 1  &  v14_04 <= 9 )
        &  ( v14_05   = 7  |   v14_05   = 8 ) )
   Error_LuzEléctrica = 1 .
EXECUTE .
* Tratamiento de los ignorados *
IF ( v14_04 = 99  |  v14_05 = 9 )
   Error_LuzEléctrica = 0 .
EXECUTE .

* Suma de errores de vivienda por hogar *
COMPUTE Error_Total_Vivienda = 0 .
EXECUTE .
COMPUTE
Error_Total_Vivienda =
     ( Error_TenenViv
       + Error_MaterialPared
       + Error_MaterialPiso
       + Error_AbastecimAgua
       + Error_ProcedenAgua
       + Error_ServicioSanitario
       + Error_LuzEléctrica ) .
EXECUTE .

* Ordenamiento del archivo por
       Llave_hogar_2005 *
SORT CASES BY
  Llave_diferente_05 (A) .

* Suma por hogar de los errores personales *
AGGREGATE
  /OUTFILE=*
  MODE=ADDVARIABLES
  /BREAK=Llave_diferente_05
  /Error_Total_Persona_sum =
SUM(Error_Total_Persona).

* Agregación de todos los errores, los personales
      y los de la vivienda *
COMPUTE Error_Total= 0 .
EXECUTE .
COMPUTE Error_Total =
      Error_Total_Persona_sum
     + Error_Total_Vivienda .
EXECUTE .

* Selección de hogares y frecuencia de errores
      por hogar *
USE ALL.
COMPUTE filter_$=(Hogar_diferente = 1).
VARIABLE LABEL
    filter_$ 'Hogar_diferente = 1 (FILTER)'.
VALUE LABELS
    filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .
FREQUENCIES
  VARIABLES=Error_Total
  /ORDER=  ANALYSIS .

* Selección de hogares para regresiones (3178),
      eliminando ignorados *
USE ALL.
COMPUTE
 filter_$ =
 ( Hogar_diferente = 1
 & Sexo_infor_05_max > 0
 & Edad_infor_05_max < 98
 & NivelEduc_re_infor_05_max < 99
 & EstConyug_infor_05_max < 9
 & tamahog_05 < 98
 & (nivpob_05 > 0 & nivpob_05 < 9)
 & (ingperca_05 > 0 & ingperca_05 < 99999999)).
VARIABLE LABEL
    filter_$
      'Hogar_diferente = 1
       & Sexo_infor_05_max > 0
       & Edad_infor_05_max < 98
       & NivelEduc_re_infor_05_max < 99
       & Niv... (FILTER)'.
VALUE LABELS
    filter_$  0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .

* Creación del Error Total como variable
       dicotómica *
COMPUTE Error_dicotomico = 0 .
EXECUTE .
IF (Error_Total > 0)
   Error_dicotomico = 1 .
EXECUTE
```

b. Regresiones

```
* Selección de hogares *
USE ALL.
COMPUTE
 filter_$ = (Hogar_diferente_05 = 1
    & RelParen_infor_05_max > 0
    & (Edad_infor_05_max > 12
        &  Edad_infor_05_max < 98)
    & Sexo_infor_05_max > 0
    & CondAct_infor_05_max > 0
    & NivelEduc_re_infor_05_max < 99
    & (EstConyug_infor_05_max > 0
        & EstConyug_infor_05_max < 9)
    & tamahog_05 < 98
    & (nivpob_05 > 0 & nivpob_05 < 9)
    & (ingperca_05 > 0
        & ingperca_05 < 99999999)).
VARIABLE LABEL
 filter_$ 'Hogar_diferente_05 = 1
    & RelParen_infor_05_max > 0
    & (Edad_infor_05_max > 12
    &  Edad_infor_05_max < 98)
    & Se... (FILTER)'.
VALUE LABELS filter_$  0 'Not Selected' 1
'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE .

* Cálculo de error dicotómico *
COMPUTE Error_dicotomico = 0 .
EXECUTE .
IF (Error_Total_05 > 0) Error_dicotomico = 1 .
EXECUTE .

* Regresión logística *
LOGISTIC REGRESSION VARIABLES
  Regr_Error_Total_dicotomico
  /METHOD = FSTEP(LR)
    MismoInf_recod
    Edad_infor_05_max
    RelParen_recod
    NivelEduc_re_infor_05_max
    EstConyug_recod
    Zona_recod
    tamahog_05
    Nivel_recod
    Ln_Ingperca_05
  /CONTRAST (MismoInf_recod)=Indicator(1)
  /CONTRAST (RelParen_recod)=Indicator(1)
  /CONTRAST (EstConyug_recod)=Indicator(1)
  /CONTRAST (Zona_recod)=Indicator(1)
  /CONTRAST (Nivel_recod)=Indicator(1)
  /CLASSPLOT /CASEWISE OUTLIER(2)
  /PRINT = GOODFIT CORR ITER(1) CI(95)
  /CRITERIA = PIN(.05) POUT(.10)
   ITERATE(20) CUT(.71) .

* Regresión lineal con el Error total *
REGRESSION
  /DESCRIPTIVES MEAN STDDEV
   CORR SIG N
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS CI BCOV R
   ANOVA COLLIN TOL CHANGE ZPP
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT Error_total
  /METHOD=STEPWISE
   MismoInf_b_max
   Sexo_infor_05_max
   Edad_infor_05_max
   RelParen_infor_05_max
   ConAct_infor_05_max
   NivelEduc_re_infor_05_max
   EstConyug_infor_05_max
   Region
    Zona
    Nivel
    Tamahog_05
   Nivpob_05
   Ln_Ingrperca_05
  /PARTIALPLOT ALL
  /SCATTERPLOT=(*ZRESID ,*ZPRED )
  /RESIDUALS DURBIN HIST(ZRESID)
   NORM(ZRESID)
  /CASEWISE PLOT(ZRESID) OUTLIERS(3) .
```

CAPÍTULO 3
LA RELACIÓN ENTRE LA EDUCACIÓN FORMAL DE PAREJA

3.1 INTRODUCCIÓN

La conformación de las parejas es uno de los acontecimientos vitales más programados por hombres y mujeres de todas las culturas. El conocimiento popular al respecto sostiene que las uniones de pareja vinculadas centralmente por amor, no tiene que ver con edad, razas, poder económico. No obstante, la realidad da cuenta de que las relaciones de pareja están permeadas y se desarrollan dentro de ambientes comunes para ambos actores, configurándose relaciones explicadas y caracterizadas por compartir espacios físicos y psicológicos similares, y por tanto, económicos y sociales; es decir, dentro de un mismo estrato social o educacional.

El concepto de aleatoriedad, o el amor ciego en su formulación romántica, aunque relegado a la ficción, sigue siendo analíticamente útil para nuestros efectos al contribuir a establecer la división entre aquello que deberíamos esperar en condiciones de independencia y lo observado en los datos, a la luz de la argumentación teórica que explica la diferencia.

Esta investigación realiza un análisis de la asociación en la formación de las uniones de las parejas casadas o unidas consensualmente o de hecho, a partir del nivel educativo alcanzado por los miembros de esta y su influencia en la reproducción de la estructura social del país. A la luz de los argumentos teóricos hallados en la literatura y con la ayuda de los modelos log-lineales, examinamos la formación de pareja con el mismo o diferente nivel de escolaridad, para determinar el grado de simetría en torno a la incorporación de la educación en la selección conyugal; el trabajo se enmarca para el período de 2004.

La investigación consta de cinco partes. En la primera, se definen los objetivos. En la segunda parte, tratamos las cuestiones teóricas relacionadas con el estudio de los patrones en la formación de las uniones a partir del nivel educativo, importancia de las implicaciones individuales y colectivas; además, ilustramos la situación con datos para Costa Rica. En la tercera parte, tratamos las cuestiones de orden metodológico: obtención de los datos, procesamiento y presentación de los "modelos". En la cuarta parte, analizamos los resultados, y en la quinta, trazamos las conclusiones finales y apuntamos elementos para la discusión.

3.1.1 Formulación y justificación del problema

Las pautas de formación de la familia, de lo cual Costa Rica no es la excepción, han estado cambiando significativamente durante décadas. La postergación del primer matrimonio, particularmente entre las mujeres, así como el aumento de la unión consensual, son algunos de los cambios destacados. Existen estudios que contribuyen a explicar esas transformaciones, enfatizando la posición de la mujer, y en particular en cómo estos cambios se vinculan con incrementos en credenciales educativos, y en las tasas de participación de actividad femenina en el mercado laboral, *(...) El cambio, que se produjo sin intervención del Estado, a favor o en contra, estuvo provocado fundamentalmente por la rápida expansión de la educación superior femenina y por su participación como fuerza de trabajo. Ello fue expresión, ante todo, de una postergación del matrimonio y de la maternidad, que gracias a la llegada de la píldora anticonceptiva no las obligaba a postergar también las relaciones sexuales* (Therborn: 2004, 30).

Los datos para Costa Rica en la dinámica demográfica se vinculan con cambios importantes en la formación de los hogares. La tasa global de fecundidad es de 3,3 en 1987 y se redujo a 1,9 en 2006; la tasa de nupcialidad registrada en 1985 fue de 7,5, con un comportamiento descendente, que para 2005 es de 5,7. El promedio era de 4 hijos(as) por mujer en su vida fértil en 1980, y bajó a 2 en 2006. También se redujo el tamaño medio de

los hogares, que pasó de 4 miembros en 2001 a 3,8 en 2005. También la ruptura voluntaria de las uniones ha mostrado cierto aumento: la proporción de personas divorciadas por cada cien matrimonio inscritos en 1985 fue de 13,4% y de 40% para 2005, según datos del Instituto Nacional de Estadística y Censos.

Las cifras disponibles en la encuesta de fecundidad de 1999 en Costa Rica muestran un aumento en la proporción de uniones libres, comparada con la encuesta de 1992, lo cual indica un cambio en el comportamiento conyugal de las mujeres costarricenses. En 1992 se reportó un 20% de mujeres entre 20 a 49 años en unión libre, aumentando para 1999 a 28% en mujeres entre los 18 y 44 años de edad.

Las transformaciones encuentran vínculos internacionales visibles e intensos como destaca el autor Therborn, tal es el caso *de la tercera oleada de cambios, gran parte de la conexión fue aportada por el entrecruzamiento institucional global y por movimientos de dimensión mundial. El Año de la Mujer, instituido por la Organización de las Naciones Unidas en l975, y el Decenio de las Naciones Unidas para la Mujer (1975-l985) pusieron en movimiento una multitud de estudios nacionales, conferencias, organizaciones e iniciativas orientadas a las relaciones de género. El Banco Mundial, algunas oficinas de ayuda nacionales y fundaciones privadas dotadas de recursos como el Consejo de Población y, gradualmente, la Organización de las Naciones Unidas misma, permitieron que el control de la natalidad, invento posterior a la segunda guerra mundial, dirigido por el Estado, alcanzara un ímpetu intercontinental, con apoyo de especialistas y recursos económicos. En el mundo desarrollado, el feminismo, en su calidad de corriente cultural amplia que recoge las aspiraciones de las mujeres a la educación superior, el trabajo, las carreras profesionales y la autonomía personal, aportó una dirección transnacional* (Therborn: 2004, 31)

Los cambios en la composición de los vínculos de pareja son la manifestación de una serie de condicionantes psico-sociales, económicos y políticos, que interactúan dialécticamente y subyacen en cada uno de los individuos integrantes de la pareja, así como

la suma de las mismas se traslucen y traducen en el comportamiento social que muestran las sociedades. De manera que emparejarse en un período histórico dado nos refiere a múltiples factores macrosociales y microsociales, que determinan diferencialmente a las familias, dependiendo de su pertenencia a una determinada cultura, estrato social o etapa del desarrollo social, donde se repite, regenera, crea y transforman valores, conductas interrelaciones y procesos de intergeneraciones.

El concepto de emparejarse lo entendemos en esta investigación como la unión de los miembros de la sociedad en parejas heterosexuales, cuya vinculación primordial para este estudio es el nivel educativo de los miembros que la integran. Se tomó las jefaturas de hogar compuestas solo por hombres, en el tanto que la proporción de mujeres jefas de hogar resultó sumamente baja, con un valor del 4% en la categoría de casadas o en unión libre.

Partiendo de que las personas se emparejan al azar, con independencia de su origen, religión, posición económica, estatus social, cultura u otra característica adquirida o adscrita, entonces la probabilidad de formar una unión de pareja estaría determinada por la disponibilidad combinada de las variables de referencia anteriores; es decir, por los condicionantes estructurales del mercado matrimonial. Sin embargo, entre los científicos sociales existe un amplio consenso en torno a que más allá del azar, individuos con características similares tienden a unirse entre ellos formando pareja homógamas. Piani (2003) y Esteve (2005) indican que no solo los determinantes estructurales, sino, también, las preferencias personales, la mediación de terceras partes (históricamente familia), Estado e Iglesia, influyen en el proceso de escogencia conyugal.

En ese sentido, la macro tarea de aprehender la fenomenología de la sociedad, en el tema de las variables que intervienen en la selección de pareja, la circunscribimos al ámbito no de categorías o *variables individuales sino hacia un enfoque colectivo*, como lo señala Cabella, citado por Piani (2003:3), *un enfoque centrado en el núcleo familiar. Esta tendencia se basa en la observación de que muchos de los comportamientos y características macro de la sociedad se gestan en el entorno micro de la familia.* Agrega la

misma, *la familia es una de las principales instituciones de la sociedad y como tal tiene un carácter netamente dinámico; está en permanente proceso de cambio en las sociedades occidentales el núcleo familiar ha transitado por grandes transformaciones que se vinculan fuertemente con la evolución de la estructura productiva*.

Los alcances de los patrones de unión en la estructura productiva son múltiples y variados; como afirma Piani (2003: 2), e*studios previos indican que para la sociedad uruguaya el grado de similitud de ciertas características de las parejas, como por ejemplo el nivel de escolaridad, acarrea consecuencias importantes sobre los patrones de unión en la estructura de la sociedad y su descendencia*, es decir, las características que se pueden tomar para definir las uniones como la edad, la educación, la inteligencia, la pertenencia política, la religión, etcétera.

En Costa Rica, las tipologías de las uniones conyugales y las características a nivel educativo de los miembros fundadores de la pareja se trasluce en la estructura social de su progenie, y por tanto, en la estructura productiva de la sociedad. Como lo menciona Castro y Gutiérrez (2007: 10), *la educación juega un papel doble, pues por un lado es una de las vías de la movilidad social ascendente, pero por otro es un elemento diferenciador creciente que consolida los procesos adscriptivos en la estratificación social. Por adscriptivo se entendería la permanencia en un grupo socioocupacional similar en los hijos respecto de los padres.* En este estudio, la tipología de uniones conyugales se cataloga en homogéneos, con similar nivel educativo entre miembros de la pareja, y otra por los heterogéneos, con diferente nivel educativo.

Esta investigación plantea el análisis de las similitudes o diferencias en la formación de las uniones, según el nivel educativo alcanzado por los miembros de la pareja y su reproducción en la estructura social relación del nivel educativo y su clasificación con la situación de pobreza reflejada por el hogar. En ese sentido, se detecta ausencia de estudios en esta línea para Costa Rica. Esto permite aportar información y elementos de reflexión sobre las tendencias que distinguen la formación de uniones conyugales en un marco de

desigualdad social y de género, de la creciente participación de la mujer en el mercado laboral, del aumento en los indicadores de educación de las féminas y, por tanto, de las modificaciones en los nexos de pareja y de los vínculos familiares, característico de los países de la región. Así, Therborn (2004:30) sintetiza *los actuales modelos y tendencias de la familia pueden resumirse en tres palabras, complejidad, contingencia y contradicción. Complejidad en el sentido de la coexistencia y entrelazamiento de las formas familiares; contingencia de relaciones, por las opciones y accidentes que siguen al debilitamiento de la regulación institucional; y contradicción entre preferencias, entre situaciones y recursos. La familia es una de las instituciones y uno de los acontecimientos más importantes que modelan el curso vital de los individuos, y pesa considerablemente tanto sobre los parámetros del poder mundial como sobre la política interna en la mayoría de los países. La persistente importancia de la familia no debe sorprender. Se trata después de todo del vínculo entre dos instintos básicos del género humano, sexo y poder.*

El estudio del tema de la pareja con base en la variable del nivel educativo de sus integrantes, posibilita plantear, posteriormente, otro tipo de interacciones entre grupos sociales de la formación de las uniones conyugales entre miembros de distintos colectivos étnico-religiosos en otras poblaciones del país, analizar la postergación matrimonial y su relación con el logro educativo, las relaciones en el momento de la formación de pareja y el sistema educativo de acuerdo con variables demográficas (edad, ocupación, número de la unión (primera, segundas nupcias o unión de hecho; vale decir, relación conyugal)).

La problemática analizada en esta investigación puede sintetizarse en: ¿cuál es la asociación de la educación en las pautas de formaciones de uniones conyugales?; esto es, el cometido final del presente estudio es examinar la asociación - relación entre la educación y las pautas de la formación de las uniones conyugales para responder si los individuos con mayor educación tienen oportunidades mayores o menores de casarse o establecer uniones con personas menos educadas y su repercusión o incidencia en la reproducción de la

estructura social; es decir, en la situación de pobreza, en la sociedad costarricense durante el año 2004.

Específicamente, pretende aproximarse al cuestionamiento de ¿quién se casa o se une con quién?, en la cual se establecerán los patrones de asociación entre el nivel educativo de la jefatura de hogar y el de su cónyuge en torno a la formación de la pareja, caracterizando las uniones en función de la educación y, por lo tanto, una categoría estaría formada por los homogéneos (similar nivel educativo) y otra por los heterogéneos (diferente nivel educativo), tratando de considerar todas las combinaciones posibles.

Un producto esperado es estimar las probabilidades entre el nivel educativo del varón y su cónyuge en la formación de pareja y determinar, a partir de los datos, el modelo conceptual de emparejamiento vigente para la sociedad costarricense en un momento dado (2004), haciendo un análisis de corte transversal. Además, al analizar parejas de diferentes cohortes, habrá otras variables intervinientes y diferentes influencias temporales.

En Costa Rica se ha hecho poca reflexión sistemática o científico-académica que permita establecer y entender cuáles han sido los efectos en la estructura de la sociedad que ha generado la formación de uniones conyugales. Desde el punto de vista de la realidad social, Costa Rica muestra rasgos de una sociedad tradicional, donde son evidentes algunos patrones reproductivos de carácter económico y cultural que no han sido analizados a la luz de la asociación del eje educativo y los vínculos de pareja que se desarrollan en torno a lo que podemos denominar como la formación de estratos sociales, o élites socioeconómicas.

3.1.2 Objetivos

Objetivo General

1. Examinar la relación del nivel educativo en la pareja para el año 2004, empleando un modelo log-lineal de corte transversal, a partir de la Encuesta de Hogares de Propósitos Múltiples, del Instituto Nacional de Estadística y Censos de Costa Rica.

Objetivos Específicos

1. Determinar, a partir de la especificación de los modelos indicados por Park (1991), el modelo estadístico que mejor explique la realidad costarricense de acuerdo con el nivel educativo de los integrantes de la pareja.

2. Relacionar los niveles de escolaridad de la pareja y la situación de pobreza que refleja la reproducción de la estratificación social en la estructura familiar costarricense.

3.2 REFERENTE TEÓRICO

3.21 Reproducción de clase o estratificación

Con el propósito de entender la fenomenología de la sociedad, los científicos sociales de las diferentes escuelas del pensamiento han tratado de formular posibles interpretaciones del fenómeno de la organización social y sus divisiones internas clasificándola, según razones epistemológicas y fundamentos filosóficos importantes, en estratos o clases sociales. Estas teorías buscan explicar la desigualdad en la sociedad, y por tanto, su reproducción social.

Tales modelos o paradigmas de análisis con algunos matices importantes hacia dentro, han sido extensamente discutido, en lo fundamental por dos grandes vertientes: materialismo histórico y estructural funcionalismo o positivismo (autores como Marx y Weber son representantes claves de cada alternativa) que se oponen radicalmente en los fundamentos históricos de sus propuestas, pero que en sus contenidos intermedios ofrecen aspectos que pueden ser tratados con alguna libertad en lo metodológico, los cuales enriquecen el análisis de algunos temas específicos. Para Marx, una clase social es un grupo que tiene relación con los medios de producción; de ahí su distinción en las sociedades capitalistas de dos clases sociales, en referencia a la posesión de los medios de producción; situación que da origen a la denominada categoría de luchas de clases dado el conflicto social a partir de la posesión de los medios de producción. Según Castro (2007, 31) *las clases sociales constituyen para la perspectiva de Marx grupos reales de agentes, definidos principalmente, pero no en forma exclusiva, por su lugar en las relaciones de producción y, concretamente, por la propiedad o no de los medios de producción.*

Para Weber, *las divisiones de clase se derivan no solo del control sobre los medios de producción, sino de diferencias económicas que nada tienen que ver con la propiedad,*

como es el caso de los conocimientos técnicos y las credenciales o cualificaciones, como diplomas y títulos académicos, que les permiten a ciertos individuos tener condiciones de trabajo más favorables y mejores remuneraciones (Rodríguez, 1997: 43).

La propuesta de clases sociales ha sido más fecunda en los temas políticos y en los del desarrollo-subdesarrollo. Entre tanto, es el paradigma de la estratificación el más desarrollado en los métodos cuantitativos. Se entiende como estratificación social la conformación en estratos (grupos verticales) bien diferenciados, de acuerdo con una multiplicidad de criterios, la mayoría de ellos perfectamente cuantificables, establecidos y reconocidos, que ayudan a estudiar la composición de un entorno social complejo, el cual debe ser agrupado según diversos criterios para lograr su estudio, descripción y comprensión. En otras palabras, la estratificación es la distribución de los individuos en grupos o estratos jerárquicamente definidos, mediante la incidencia de una serie de variables interrelacionadas de las dimensiones económica y social. Entre ellas, podemos citar la edad, estrato o procedencia social, nivel educativo o instrucción, ingreso, ocupación, actividad económica y género; por ejemplo, *(...) el estudio de la estratificación social y la educación nos obliga a analizar: el hecho de las diferencias sociales, las modalidades de estratificación, sus rasgos y estructura. A partir de aquí podremos entender las pautas educativas de cada clase social, sus posibilidades, la caracterización psicológica de los individuos de cada clase social, sus actitudes y la consiguiente repercusión en la educación de los hijos/as (...)* (García: 2001, 2).

Existen múltiples factores que interactúan en la reproducción social, pero en este documento es de especial interés entender la reproducción social como el mecanismo mediante el cual el nivel de escolaridad determina y especifica la ubicación en un grupo social, según las características que tienen los miembros de la pareja que son transmitidas a sus hijos. El papel de la educación juega un papel decisivo en la dinámica propia de la reproducción de las relaciones de clases de la producción de bienes materiales y simbólicos; y la clase dominante puede definir e imponer su modelo de individuo y

sociedad, así como seleccionar y controlar los medios por los cuales la educación los realiza.

En esta línea, Varela (2005) indica, en relación con la teoría de reproducción esbozada por Bourdieu y Passeron, que del mismo modo que las instituciones económicas y su lógica de funcionamiento favorecen a la población estudiantil que ya poseen capital económico, las instituciones educativas están estructuradas para favorecer a la población estudiantil que ya poseen capital cultural. Asimismo, plantea la *Escuela reproduce las desigualdades sociales al reforzar el habitus que corresponde a las familias de la clase media. De este modo la Escuela no es el lugar en el que se producen las desigualdades sociales sino donde se legitiman esas desigualdades. El capital cultural, definido "arbitrariamente" como "la cultura legítima" por los grupos dominantes y como el que debe ser transmitido a todos los escolares -que supuestamente acceden a él en régimen de igualdad- es el capital cultural de las clases medias. La Escuela puede así naturalizar y ocultar las desigualdades sociales al transformar las diferencias de clase en desigualdades individuales, en desigualdades de talento y de capacidades individuales en el acceso y apropiación de "la cultura"*.

Bourdieu, quien analiza fundamentalmente el análisis de la reproducción cultural, coincide con el planteamiento acerca de que la educación determina las tendencias de las estructuras a reproducirse a sí mismas, mediante la producción de actores sociales o grupos sociales que engendran prácticas mediadoras entre la reproducción social y cultural. Como lo expone Bourdieu, (1978: 257) *el sistema educativo reproduce de modo perfectísimo la estructura de la distribución del capital cultural entre las clases, ya que la cultura que transmite está más próxima a la cultura dominante y el sistema de inculcación a que recurre está menos alejado del sistema de inculcación practicado por la familia.*

En este contexto, independientemente de cuáles sean los factores que explican las pautas de composición de las parejas en las sociedades occidentales contemporáneas, los investigadores se han centrado en implicaciones individuales y sociales. Así, a nivel

individual, la composición de las parejas se analiza en el marco del estudio de la familia, asumiendo que las características de los cónyuges y su combinación pueden determinar las decisiones familiares que estos tomarán posteriormente. Por su parte, los sociólogos han hecho especial énfasis en el papel del matrimonio en la reproducción de la estructura social (Kalmijn: 1991), desde la formación de la pareja como un recurso utilizado por los individuos para consolidar o mejorar su condición social individual o familiar, tomando en cuenta el nivel educativo de las partes (miembros de la pareja).

De manera que la variedad de posibilidades educativas al cual los individuos tengan acceso, (entre otros factores), refleja los patrones educativos de cada clase o estrato social, y por consiguiente la desigualdad de clases que origina la desigualdad de oportunidades educativas, además de reproducir los patrones o pautas culturales de su grupo social.

En el desarrollo de todo país, avanzar a una sociedad que provea la igualdad de oportunidades y la participación de los distintos sectores sociales, constituyen elementos importantes que posibilitan el desarrollo y acceso a los recursos económicos, mejorando la calidad de vida. En ese marco, la esfera de la educación es una de las principales áreas que coadyuva en la consecución de tal fin, *(...) la educación constituye el principal instrumento para el progreso individual y, en segundo término, cristaliza a la vez las posibilidades de una colectividad para dotarse de mejores condiciones en su calidad de vida. El papel de la educación, entonces, constituye, de alguna manera, un termómetro de las posibilidades que existen en una sociedad en cuanto al desarrollo humano (...)* (Ruiz: 2004, 11).

En este contexto, en la transmisión intergeneracional se ubican las pautas y comportamientos de la reproducción de las relaciones sociales que ha transmitido la pareja a su progenie, respecto a aspectos simbólicos, objetivos y subjetivos. Tal como se documenta en el estudio de la Comisión Económica para América Latina (CEPAL) sobre el desarrollo social de América Latina, donde se expone*: pese a los esfuerzos realizados en la región por masificar el acceso al sistema educativo formal, no se ha podido evitar que la adquisición de capital educativo continúe estando condicionada por dinámicas*

intergeneracionales. El hecho de que las oportunidades de educación y, por consiguiente, las de acceso a empleos más estables y mejor remunerados, sean en alto grado heredadas, constituye un elemento clave para que las desigualdades socioeconómicas actuales se reproduzcan indefinidamente en las sucesivas generaciones más jóvenes. En efecto, la probabilidad de recibir un mínimo adecuado de educación está determinada en gran medida por el grado de educación de los padres y por la capacidad económica del hogar de origen: hacia fines de los años noventa, alrededor de 75% de los jóvenes urbanos provenían de hogares en que los padres disponían de una educación insuficiente —menos de 10 años de estudio— y, en promedio, más de 45% no habían terminado el ciclo secundario, que equivale en la mayoría de los países a 12 años de estudio, y que hoy se considera el umbral educativo mínimo, en las zonas urbanas, para acceder al bienestar (...) el hecho de que las oportunidades de educación y, por consiguiente, las de acceso a empleos más estables y mejor remunerados, sean en alto grado heredadas, constituye un elemento clave para que las desigualdades socioeconómicas actuales se reproduzcan indefinidamente en las sucesivas generaciones más jóvenes. En efecto, la probabilidad de recibir un mínimo adecuado de educación está determinada en gran medida por el grado de educación de los padres y por la capacidad económica del hogar de origen (....) (CEPAL: 2004-1, 27).

3.2.2 Educación y su incidencia en la ubicación social

Según Esteve (2005), dada la trascendencia de la formación de pareja, el estudio de la homogamia en particular, ha captado el interés de varias disciplinas, desde la demografía, la antropología y, por supuesto, la sociología. Fundamentalmente, la semejanza de los cónyuges se analiza, en primer lugar, respecto a características adscritas, tales como la posición social definida por la adscripción familiar, la religión o la etnia y, en segundo lugar, respecto a características adquiridas, entre las que destaca el nivel educativo. Las barreras en la composición de pareja existentes entre individuos de distintos grupos sociales, religiosos o étnicos, pierden relevancia en las sociedades contemporáneas,

mientras que los elementos de identificación sociales adquiridos, como el nivel de instrucción, son cada vez más determinantes (Kalmijn, 1998).

Para entender este proceso, es importante entender la educación como una variable que informa no solo de la calificación de los individuos sino, también, indirectamente, de su posición socioeconómica y de su capacidad de promoción social y profesional. Por este motivo, el análisis de la concordancia entre los niveles educativos de los miembros de la pareja es un eje prioritario en el estudio de las uniones conyugales.

En Costa Rica, el papel del Estado en la vida educativa del país en las diferentes administraciones ha contribuido a consolidar una importante carrera educativa: apertura de escuelas y colegios, y generación de leyes (Ley de Educación Común de la reforma educativa) desde finales del siglo XIX, construyendo así una plataforma educativa, *(...) Históricamente, Costa Rica ha apostado por la educación como elemento clave para promover el desarrollo humano. El decidido impulso que se dio a la educación primaria a finales del siglo XIX fue uno de los factores que marcó la diferencia en materia de alfabetización que exhibió el país frente al resto de Centroamérica, pues ya para inicios del siglo XX había superado su rezago (...)* (Estado de la Educación Costarricense: 2005, 68).

En esta perspectiva histórica, señala González (2003: 292), la Asamblea Constituyente de 1949 puede considerarse el primer paso revolucionario que permite modificar la concepción educativa del país, pues, además de dar autonomía a la Universidad, crea el Ministerio de Educación Pública y propone un Consejo Superior de Educación (Dirección General de Enseñanza Oficial) (art. 81); declara la gratuidad de la preescolar y secundaria (art. 78); organiza la educación como un proceso integral correlacionado (desde la maternal o educación preescolar hasta la universitaria) (art. 77) y estimula la participación de las comunidades en la organización educativa.

Es así como bajo el amparo del Estado benefactor se crean varias instituciones autónomas y semiautónomas, con carácter social, en el cual surge el Instituto Nacional de Aprendizaje (1965), preocupado en la necesidad de formar individuos, con una especialidad técnica y vocacional con perfiles que respondan a las demandas del sistema que no se agotan en la educación formal, ni en los niveles de educación secundaria o superior; esto, con el fin de aumentar la productividad en diversos sectores de la economía.

Las diferentes concepciones de Estado, desde un Estado desarrollista e interventor hasta un Estado empresario, que han caracterizado las diferentes administraciones de los gobiernos costarricenses, como lo señala Rojas y Retana (2003: 486), han hecho un esfuerzo por la democratización de la educación, entendida ésta como la ampliación de oportunidades educativas en todos los niveles de la educación nacional para los pobladores de este país, aunque, durante los últimos años, se han visto determinadas por las mismas limitaciones del país en el plano económico y social; vale decir, la educación y la política nacional plantearon nexos entre educación, Estado, sociedad y economía y sus necesarias implicaciones para el desarrollo del país.

En el contexto de oportunidades educativas, podemos ubicar la vida universitaria, la cual data desde 1914 con la formación de docentes en la Escuela Normal de Costa Rica (que posteriormente es retomada por las universidades); además a partir de la mitad del siglo XX se diversifica la oferta de estudios universitarios y parauniversitarios. Los estudios superiores estatales y privado se inician con la creación de la Universidad de Costa Rica en 1940, el Instituto Tecnológico de Costa Rica en 1971, la Universidad Nacional en 1973 y por último la Universidad Estatal a Distancia en 1977, instituciones que componen el sistema de educación superior pública universitaria de Costa Rica, aparte de la creciente cantidad de universidades privadas, lo cual se traduce en décadas efectivas, durante las cuales en Costa Rica se ha producido un importante contingente social, con grados académicos, *(...) aprovechando los resultados del censo 2000, las tasas brutas de escolaridad en este nivel educativo, obtenidas tras comparar la matrícula con la población*

nacional de 18 a 24 años, experimentaron un notable crecimiento entre 1984 y 2000: prácticamente duplicaron su valor, al pasar de 15,% a 29,6% en ese período. Este importante avance se dio en realidad a partir de 1970, y está asociado a la creación de diversas instituciones de estudios superiores en esa época (...) (Estado de la Educación Costarricense: 2005, 98).

Asimismo, a partir de los noventas el Estado neoconservador-neoliberal ha asumido prácticas contrapuestas al accionar de los gobiernos precedentes. No obstante, no resta importancia que el sistema educativo en todos los ámbitos individuales y colectivos continúa siendo un factor relevante que posibilita e interviene en la estructura productiva y sociopolítica del país. Actualmente, como señala el Informe del Estado de la Educación (2005: 53), el reto de las políticas educativas está dirigido a tres áreas principales: acceso, permanencia y éxito, que se enfocan a aumentar la cobertura de la enseñanza preescolar, la secundaria y la instrucción no tradicional, así como a permitir el acceso a la educación a grupos con necesidades especiales y en condición de pobreza. En lo que respecta al contenido del currículo, se dan directrices para la formación en valores y la incorporación de una perspectiva basada en los principios del desarrollo sostenible, además de competencias para el trabajo, con la incorporación de idiomas, informática, habilidades y conocimientos técnicos.

Entonces, el nivel escolar en la formación de los integrantes de la pareja en el estado actual de para efectos de nuestro tema de estudio, es una dimensión relevante. Como se indica en el Informe sobre la Educación (2005: 65), la educación impacta la sociedad en diversos planos: tiene efectos sociales (promueve la movilidad social, contribuye a disminuir la pobreza y crea un clima propicio para estimular la formación educativa, pues se ha comprobado que a mayor nivel educativo de los padres y madres, crecen las probabilidades de que los hijos alcancen ese nivel, situación que no sucede en hogares donde los progenitores tienen menores niveles de formación); tiene efectos económicos (una mano de obra más calificada, más productiva y capaz de generar mayores ingresos) y

tiene, asimismo, efectos políticos (en secundaria es donde ocurre la socialización política básica). A su vez, el informe destaca que ningún país ha logrado un avance significativo, y menos aún sostener sus logros, sin una fuerte y continuada inversión pública en educación.

3.2.3 Mercado matrimonial

Existen distintas teorías que consideran el papel que la educación puede jugar como determinante de las opciones matrimoniales. Según la teoría clásica de la Nueva Economía de la Familia, los individuos proceden a seleccionar una pareja de acuerdo con sus afinidades y preferencias individuales. Depende de un supuesto de racionalidad individual en la toma de decisiones afectivo-familiares, encabezada por el economista estadounidense Gary Becker (1987), se considera que estas preferencias procuran maximizar el intercambio de recursos que se produce en la formación de la pareja. Sin embargo, la satisfacción de estas preferencias mediante una opción óptima no siempre es posible. Ante un contexto de posibilidades matrimoniales limitadas y sometidas a la eventual presión de terceras partes, los candidatos se conforman con una buena opción.

En otras palabras, el modelo complementario de pareja, aquel en el que el hombre se especializa en tareas productivas y la mujer en las tareas reproductivas, sería el óptimo para Becker, teoría que se podría denominar "tradicional". *De acuerdo con este esquema, las mujeres valoran en el hombre su capacidad de éxito en la esfera productiva, altamente correlacionada con su nivel de formación, vale decir, el nivel educativo del marido sería más valorado que el de la esposa y por consiguiente se esperaría una tendencia al patrón clásico de hipergamia femenina donde la mujer tiene un nivel de instrucción menor al del hombre, este modelo asume que las preferencias conyugales son asimétricas entre hombres y mujeres. Sin embargo, como apunta Oppenheimer (1994), la incorporación de la mujer al mercado de trabajo, y por ende la independencia económica obliga a replantear este principio de complementariedad puesto que el trabajo extradoméstico adquiere valor específico*, además de que la variación de los patrones de comportamiento de la mujer

aumentarán la exigencia en la selección conyugal, paralelamente retrasarán la edad al matrimonio. *En este nuevo contexto, en el que las parejas optimizan sus recursos sin necesidad de especializarse, el nivel de instrucción de la mujer pasa a ser tan valorado como el del hombre de tal manera que sería de esperar un aumento de los niveles de homogamia igual nivel de instrucción para ambos miembros de la pareja (Es*teve: 2005,4).

Asimismo, Binstock (2004: 55) cita que entre *las perspectivas conceptuales más difundida para interpretar los cambios en las pautas de formación de la familia, es la que pone el foco en el crecimiento de la independencia femenina, formulada por Becker (1981). Esta perspectiva sostiene que uno de los mayores beneficios del matrimonio deviene de la mutua dependencia que surge de la división de roles de los cónyuges. El matrimonio es más atractivo o ventajoso cuando los miembros de la pareja tienen diferentes atributos que intercambiar (la actividad económica el varón y la doméstica la mujer). En la medida en que la mujer adquiere mayor educación –y concomitantemente mejores oportunidades laborales - disminuye su especialización en la esfera doméstica y aumenta su independencia económica, lo cual reduce los beneficios y el atractivo del matrimonio. Esta perspectiva, entonces, predice que el aumento de los logros educativos de las mujeres se asociará con mayores oportunidades de postergar el matrimonio, o incluso renunciar al mismo. La postergación del matrimonio de la mujer, por ende, se traducirá también en similares predicciones para el caso de los varones.*

Agrega la autora, *otra perspectiva para interpretar la relación entre la formación familiar y la educación es la elaborada por Oppenheimer, quien argumenta que la transición al matrimonio está directamente ligada a las incertidumbres en torno al futuro económico del potencial cónyuge, a edades jóvenes. Sintéticamente, Oppenheimer sostiene que, en sociedades con alta diferenciación de roles de género en las cuales el varón tiene el rol exclusivo de proveedor de ingresos, la edad al matrimonio -tanto para la mujer como para el varón- va a estar fuertemente asociada a la transición del varón hacia una situación de empleo e ingresos relativamente estable que, a su vez, va a estar altamente*

determinada por la edad a la que se completa la educación. Cuando el rol de la mujer cambia, y los patrones de su participación económica comienzan a asemejarse a los de los varones, las características consideradas importantes en un potencial cónyuge se vuelven más similares para ambos sexos. Las mujeres, argumenta Oppenheimer, serán evaluadas con más frecuencia por sus logros y su potencial económico, más que por características tradicionales como la atracción física o la familia de origen (Binstock 2004: 56).

En este sentido, el potencial económico de la mujer deja de estar disponible a edades jóvenes, particularmente porque dicho potencial, a su vez, también aumenta al prolongarse la educación, en virtud de la cual es predecible que el matrimonio se postergue. Con mayor educación y, concomitantemente, con mayores ingresos potenciales, las mujeres se vuelven más atractivas en el mercado matrimonial si ambos miembros de la pareja pueden beneficiarse a partir de reunir y compartir los recursos (Oppenheimer, 1988). En este sentido, a medida que aumenta la contribución potencial de la mujer a la economía familiar, podría esperarse que la posición del varón cobre menor relevancia para la formación de la familia.

De este modo, la presente investigación pretende confirmar y darle énfasis a la categoría educación, por cuanto esta categoría da cuenta sobre los patrones y tendencias que subyacen en la formación de la pareja asociada a la educación, que lleva a reproducir en su progenie la ubicación de clase social o de estratificación. De ahí que la incidencia o impacto de la formación educativa de las parejas y la influencia ejercida y heredada de padres a hijos, se toma como un aspecto determinante en el resultado social de la ubicación socioeconómica, (*...) existen evidencias de que la escolaridad de los padres incide fuertemente en el nivel educativo de los hijos, lo cual plantea una preocupación, ... que muestra cómo en Costa Rica, a menor escolaridad de los padres, mayor es la probabilidad de que los hijos no asistan al sistema, especialmente a partir de la secundaria (...)* (Estado de la Educación Costarricense: 2005, 68).

La variable del nivel educativo de la pareja representa, entre otras, variables características de los distintos grupos que definen diferentes niveles de acceso de las personas (pareja y progenie) a bienes materiales y culturales, de manera que van determinando ciertos patrones de consumo que los identifica con estilos de vida que comparten preferencias, gustos y actitudes, entre otras, condición que les especifica niveles de prestigio que les permite disfrutar cierto estatus social, al cual se refiere Castro (2007, 161), para quien *la educación, junto con el nivel de ingreso, es uno de los factores que manifiesta la diferencia y distancia social entre estratos sociales (...). La educación, si bien puede propiciar procesos de movilidad social, también es un profundo diferenciador social en particular en una sociedad que valora cada vez más el conocimiento.*

3.2.4 Estado de la situación

Con el propósito de obtener un perfil del escenario de Costa Rica en lo relacionado con la composición de parejas, se presentan cifras provenientes de la Encuesta de Hogares de Propósitos Múltiples, realizada por el Instituto Nacional de Estadística y Censos en el año 2004.

Analíticamente, nuestro objetivo principal es observar las tendencias generales de interacción entre los distintos grupos educativos; es decir, examinar el efecto del nivel educativo en la pareja por la interacción entre los distintos grupos educativos. En ese sentido, tratar las diferencias en función del tipo de unión u otras variables, supondría una desviación de dicho objetivo, aunque es materia susceptible de ser tratada en futuras investigaciones, a pesar de que las uniones de hecho son un estado en la conformación de las parejas que viene alcanzando cifras crecientes en las últimas décadas. En ese sentido, la investigación toma como referencia la población masculina con jefatura de hogar, que obtiene una mayor proporción en la encuesta y como cónyuges a las mujeres (patrón dominante en la conformación de las parejas); quedan por fuera de la investigación las

jefaturas de mujeres, cifra que viene acompañada de un lento pero continuo incremento, que alcanzó para el 2004, un 7% (2.856) en relación con la población total encuestada.

Para el año de estudio (2004), los datos presentan un comportamiento similar en la variable estado civil de la siguiente manera: en las categorías uniones de hecho el 14% y casados(as) una tercera parte del total de la población encuestada, tanto para hombres como mujeres.

El comportamiento de la relación de parentesco reporta del total de población muestral un 32% (11.366) como jefes de hogar y cónyuges un 22% (7.903), de los cuales la población de interés para el objeto de estudio se calcula del total de jefes y cónyuges, una proporción de jefes hombres del 88,63% (7.543) y una proporción cónyuges mujeres del 99,52% (7.487) según se expone en el cuadro 3.1.

Cuadro 3.1
COSTA RICA: Distribución de la población muestral
SEGÚN: Sexo y Estado conyugal
POR: Relación de parentesco
2004 (Cifras relativas en porcentajes)

Sexo y Estado conyugal	Total	Jefatura	Cónyuge	Otro
Total	**100,00**	**100,00**	**100,00**	**100,00**
Nº casos	35.698	11.366	7.903	16.429
Unión libre	13,63	19,51	27,40	2,93
Casado (a)	33,43	51,07	71,95	2,71
Otro	52,94	29,42	0,65	94,36
Hombres	**100,00**	**100,00**	**100,00**	**100,00**
Nº casos	17.715	8.510	380	8.825
Unión libre	13,68	22,98	58,16	2,79
Casado (a)	33,59	65,65	37,89	2,49
Otro	52,73	11,37	3,95	94,72
Mujeres	**100,00**	**100,00**	**100,00**	**100,00**
Nº casos	17.983	2.856	7.523	7.604
Unión libre	13,58	9,17	25,85	3,10
Casado (a)	33,28	7,60	73,67	2,96
Otro	53,14	83,23	0,48	93,94

Nota: La categoría "Otro" incorpora a hijo(a), yerno o nuera, nieto(a), padres o suegros, otros familiares, pensionista, otros no familiares, hermano(a).

Con el propósito de obtener un panorama sobre la distribución de la población muestral según ubicación geográfica (cuadro 3.2), los datos reportan para los totales de la población muestral un valor del 41% se encuentra situado en la zona urbana y el 59% pertenece a la zona rural, así como en la categoría de jefes representa un 27% y los cónyuges representa un 18% en la zona urbana. Respectivamente, en la zona rural los jefes representan una proporción del 18% y el 47% para las cónyuges.

De la composición de la población muestral, las proporciones de los hombres jefes y cónyuges representan alrededor de un 40% en la zona urbana y en la zona rural, similar comportamiento muestran las proporciones de las mujeres con un 48% y 47% para cada una de las zonas respectivamente, lo cual obedece al diseño muestral implementado.

Asimismo, la composición de la población objeto de interés (destacada en tonos grises en el cuadro 3.2) muestra que las proporciones de los hombres jefes y cónyuges representan un 40% y el valor para las mujeres en esas categorías es de un 47%.

Cuadro 3.2
COSTA RICA: Distribución de la población muestral
SEGÚN: Sexo y Relación de parentesco
POR: Zona
2004 (Cifras relativas en porcentajes)

Sexo y Relación de parentesco	Total	Urbano	Rural
Total			
Total por Zona	**100,00**	**41,11**	**58,89**
Nº casos	43.779	17.996	25.783
Total por Relación	**100,00**	**100,00**	**100,00**
Jefe	25,96	26,96	18,36
Cónyuge	18,05	17,60	47,48
Otro	55,99	55,44	34,16
Hombres			
Total por Zona	**100,00**	**42,33**	**57,67**
Nº casos	21.831	8.706	13.125
Total por Relación	**100,00**	**100,00**	**100,00**
Jefe	38,98	38,71	39,16
Cónyuge	1,74	2,25	1,40
Otro	59,28	59,04	59,44
Mujeres			
Total por Zona	**100,00**	**42,33**	**57,67**
Nº casos	21.948	9.290	12.658
Total por Relación	**100,00**	**100,00**	**100,00**
Jefe	13,01	15,95	10,85
Cónyuge	34,28	31,99	35,95
Otro	52,71	52,06	53,20

El comportamiento para el total de la población encuestada según nivel de pobreza (cuadro 3.3) se tiene como jefes de hogar es un 26% y cónyuges un 18%, de los cuales la población de interés (sobresaltada en el cuadro en tonos grises) para el objeto de estudio, se calcula del total de jefes y cónyuges, una proporción de jefes hombres del 39% y una proporción cónyuges mujeres del 35%. En este caso, 1.660 hombres jefes y 1.510 mujeres cónyuges se encuentran en estado de pobreza.

Cuadro 3.3
COSTA RICA: Distribución de la población muestral
SEGÚN: Sexo y Nivel de pobreza
POR: Relación de parentesco
2004 (Cifras relativas en porcentajes)

Sexo y Nivel de pobreza	Total	Jefe	Cónyuge
Total	**100,00**	**26,06**	**18,24**
Nº casos	38.533	10.041	7.027
No pobre	73,57	19,79	14,14
Pobre	26,43	6,27	4,10
Hombres	**100,00**	**39,36**	**1,65**
Nº casos	19.198	7.556	317
No pobre	74,30	30,71	1,29
Pobre	25,70	8,65	0,36
Mujeres	**100,00**	**12,85**	**34,70**
Nº casos	19.335	2.845	6.710
No pobre	72,84	8,94	26,89
Pobre	27,16	3,92	7,81

En términos generales, la variable nivel educativo contiene cinco categorías: primaria incompleta, primaria completa, secundaria incompleta, secundaria completa y universidad, las cuales representan el 28%, 34%, 15%, 10% y 13%, respectivamente del total de la población encuestada. A su vez, los datos muestran el 60% y el 40% del total de la población entre jefes y cónyuges.

En términos de la estructura educativa, se destaca que las proporciones más altas las ocupan los niveles más bajos de educación (primaria completa e incompleta) acorde con la estructura socioeducativa del país.

Cuadro 3.4
COSTA RICA: Distribución de la población muestral
SEGÚN: Relación de parentesco
POR: Nivel educativo
2004 (Cifras absolutas y relativas en porcentajes)

Relación de parentesco	Total	Primaria incompleta	Primaria completa	Secundaria incompleta	Secundaria completa	Universidad
Absoluto						
Total	**19.269**	**5.409**	**6.571**	**2.947**	**1.916**	**2.426**
Jefe	11.366	3.392	3.755	1.670	1.048	1.501
Cónyuge	7.903	2.017	2.816	1.277	868	925
Relativo						
Total	**100,00**	**28,07**	**34,10**	**15,29**	**9,94**	**12,59**
Jefe	58,99	17,60	19,49	8,67	5,44	7,79
Cónyuge	41,01	10,47	14,61	6,63	4,50	4,80

3.3 METODOLOGÍA

Para la explicación y comprensión del fenómeno que se estudia, es necesario tomar en cuenta dos aspectos: el primero, referido a los cuadros y resultados que se exponen corresponden a ***cifras muestrales***, en el tanto que no se trabajó con la base de datos expandida. No obstante, las características del diseño muestral establecen que la muestra es autoponderada por estratos (INEC-EHPM: 2004,20), lo cual nos permite reflejar lo que puede pasar en la población, para nuestro interés.

Por otro lado, la muestra contempla una diferenciación tácita de partida, respecto de la diferencia entre las jefaturas ocupadas por hombres y mujeres, ya que la proporción de jefaturas de mujeres es bastante baja en la composición de las parejas a nivel de la población; ello, en el tanto que las mujeres al declararse jefas de hogar en la categoría: de solteras, divorciadas y separadas, asumen que no tienen pareja. La proporción que presenta la característica para nuestra población objetivo es sumamente baja con un 4% para el año que nos ocupa, resultado de las 2.856 mujeres que se declaran jefas de hogar.

3.3.1 Fuente de datos

Los datos de los cuales se parte para efectuar el análisis de interés para la investigación, son los obtenidos de la Encuesta de Hogares de Propósitos Múltiples del año 2004; esta encuesta se realiza anualmente en el mes de julio en todo el país, labor que lleva a cabo el Instituto Nacional de Estadística y Censos (INEC), ente oficial encargado de la recolección de los datos y del diseño muestral aplicado en esta. El archivo de datos está formado por un total de 43.779 observaciones y 179 variables.

3.3.2 Población de estudio

Para los efectos de la investigación, la población de estudio se circunscribe a los hogares formados por los jefes hombres y cónyuges mujeres que forman pareja,

independientemente del estado conyugal; es decir, se toman en cuenta las personas casadas y las uniones de hecho reportadas en la encuesta[39].

Se definió el interés de la composición de la pareja donde el jefe es hombre y se excluyó a las mujeres jefas de hogar, con el propósito de que los modelos muestren el patrón de género tradicional en la formación de pareja.

3.3.3 Variables de estudio

Derivadas de los objetivos del presente estudio, las variables de interés son el *nivel de educativo* que lleva al análisis de la asociación del nivel de escolaridad en la formación de la pareja; en lo relacionado con la reproducción de clase, se utiliza la variable denominada "situación de pobreza", referida a las condiciones materiales de existencia asociada con el nivel educativo de la pareja.

Para obtener la variable *nivel educativo* del jefe de hogar y su cónyuge, se procedió a construir una nueva variable, retomando la variable de la Encuesta de Hogares "nivel de instrucción"; que atiende cuál es el último año o grado aprobado por cada miembro del hogar categorizado en cinco niveles (el nivel de instrucción ignorado por lo bajo, 0,4% del total de la muestra, se incluyó en primaria incompleta); seguidamente, se presenta la variable nivel educativo, recodificada, compuesta por cinco categorías o niveles de educación:

[39] El nivel educativo de los miembros de la pareja en relación con el parentesco asocia el nivel educativo respecto al estado actual de la pareja y no de constituida la pareja (2004).

Nivel Educativo	Años de Escolaridad
1 Primaria incompleta	De 0 a 5
2 Primaria completa	6
3 Secundaria incompleta	De 7 a 10
4 Secundaria completa	De 11 a 12
5 Universitaria	De 13 y más

Como se mencionó, la otra variable de interés es la "situación de pobreza", referida a las condiciones materiales de existencia, cuyo punto de partida para su construcción la variable "nivel de pobreza"[40] de la Encuesta de Hogares. La selección de la variable obedece a criterios de disponibilidad, dado que es la variable con la que se tiene un mayor acercamiento a partir de los datos de la Encuesta. A pesar de las limitaciones que posee respecto a las variables incorporadas en su medición, dadas las restricciones en la definición de ingreso y las variables que se toman en la construcción de esta (actualmente, está en discusión la necesidad de incluir otras variables denominadas ayudas estatales, y por tanto, la redefinición conceptual del ingreso a partir de lo que expone coincidentemente Juan Diego Trejos Solórzano y el vicepresidente del INEC, Víctor Hugo Céspedes (La Nación, 22 enero de 2007), con el propósito de subsanar las desventajas del indicador). No obstante, según declaraciones del Vicepresidente del INEC, la nueva medición de pobreza se aplicará hasta el 2008 debido a que requiere de una redefinición de la Encuesta de Hogares.

Así, la metodología empleada en el cálculo de la variable es considerar pobre el hogar cuyo ingreso no le alcanza para cubrir el costo de un grupo de bienes y servicios básicos, y en extrema pobreza a aquel al que ni siquiera le alcanza para la canasta básica

[40]/ El INEC en la medición de la variable nivel de pobreza utiliza el método de línea de pobreza, que consiste en comparar el ingreso de cada hogar con el costo de un grupo de bienes y servicios requeridos para satisfacer necesidades básicas. El costo se determina a partir de la canasta básica. El ingreso incluye los recursos que reciben las personas por sus ocupaciones principales y algunas transferencias

alimentaria. De modo que, la variable "nivel de pobreza" se codificó en dos categorías pobre y no pobre, de la siguiente manera:

Situación de pobreza

Nivel de pobreza	Calificación según EHPM	Nueva categoría
No aplica	0	Eliminados
Extrema pobreza	1	Pobres
No satisfacen necesidades básicas	2	
No pobres	3	No pobres
Ignorado	9	Eliminados

3.3.4 Evaluación de las variables de interés

De las variables último año aprobado, relación de parentesco, y estado conyugal, podemos decir, a partir de lo que muestra el archivo de datos de la Encuesta de Hogares, que el comportamiento de dichas variables en relación a la no respuesta es confiable, por cuanto las personas entrevistadas declaran sus niveles educativos y estado civil sin mayor problema. De igual manera, se ha detectado que sin importar cuál sea el informante, estas personas generalmente conocen la información de los miembros del hogar. Finalmente, es importante señalar que dada la naturaleza de las variables, estas no implican un mayor grado de complejidad y compromiso para responder; además de encontrarse situadas en el bloque del inicio del cuestionario, facilitando que la persona informante no se canse.

Para la variable "nivel de pobreza", se tiene una tasa de no respuesta de la variable recodificada del 10,8%, que incorpora a los que no tienen ingresos reportados (812 casos de

estatales, como pensiones, subsidios y becas. La medición utiliza un grupo de bienes y servicios basados en las costumbres de las familias de referencia de 1988 y 1989.

la muestra) del total de 7.515 parejas potenciales de la población objetivo del estudio; la literatura señala que se encuentra en un rango aceptable, por lo cual podemos concluir que se está trabajando con una base de datos confiable.

Finalmente, la publicación de resultados de la Encuesta de Hogares de Propósitos Múltiples no muestra los niveles de error y confiabilidad para las variables respecto de los errores no de muestreo; no obstante, es de hacer notar que los datos no están exentos de dichos errores.

3.3.5 Procesamiento del archivo de datos

El archivo de datos aportada por el INEC, procedente de la Encuesta de Hogares del 2004; se reconfiguró a la luz del interés del presente estudio. El procesamiento de los datos se realizó en dos etapas. Las operaciones realizadas en esta fase del proceso para obtener los datos requeridos, fueron:

1. Construir una variable de identificación a partir de los datos de las variables número de vivienda y número de hogar, que se utilizará como variable clave para unir los archivos.

2. Construir las variables de interés “nivel de educativo” y “situación de pobreza” (ambas categóricas).

3. Filtrar el archivo con la variable relación de parentesco y sexo para obtener respectivamente dos archivos que contienen la base de datos de jefes de hogar hombres y la base de datos de cónyuges mujeres.

4. Unir los archivos con cada uno de los datos utilizando la variable clave, que da como resultado una base de datos formada por la variable común de identificación que unifica en una línea los datos de cada unión (jefe y cónyuge), con las otras variables de interés.

5. Las observaciones con valores faltantes en la variable nivel de pobreza después de la unión de archivos se eliminan.

La segunda etapa partió del archivo de datos depurada, para realizar la corrida de los modelos con datos categóricos, usando la técnica de los modelos log-lineales para analizar el efecto del nivel educativo en las uniones de pareja y la reproducción de la estructura productiva.

El archivo de datos consta de 50 casos que resume un total de 6.703 hogares, desglosados en las diferentes combinaciones del nivel educativo de los jefes de hogar y de las cónyuges; es decir, es producto de una tabla de frecuencias o de contingencia de doble entrada, que permite examinar las frecuencia observadas, las cuales pertenecen a cada una de las combinaciones especificas de cada nivel educativo de los jefes de hogar y de las cónyuges, producto de 5 niveles educativos para cada entrada, donde cada celda o casilla de la tabla resultante represente una única combinación de las variables cruzadas; de ahí que son 25 celdas para los hogares pobres y 25 celdas para los hogares no pobres.

3.3.6 Técnica de análisis

Con el objetivo de analizar los efectos de la educación se utilizará la técnica para el análisis de datos categóricos conocida como modelos log-lineales, dado que las variables de análisis son nominales o categóricas. Tal como lo plantea Catena (2003), el análisis de frecuencias tiene el objetivo fundamental de comprobar la existencia de relaciones entre variables categóricas (o categorizadas); en dicho caso, el interés es el de hacer predicciones sobre las frecuencias de las casillas.

Se debe distinguir que la frecuencia de cada celda se considera como la variable dependiente, la variable a explicar, mientras que los factores o la combinación en los niveles educativos de la pareja son los posibles agentes explicativos, independientes. De ahí que el análisis de datos categóricos intenta proporcionar información sobre los efectos de

una variable dependiente y de los parámetros necesarios para formular el modelo, que representan la magnitud de los efectos de la fuerza de asociación entre variables.

Los parámetros son los factores explicativos que contribuyen en alguna medida a las frecuencias observadas. La contribución de cada factor es capturada por un conjunto de coeficientes, parámetros, tanto como valores tenga. La combinación de parámetros permitirá explicar las frecuencias observadas. La comparación entre los tamaños de los parámetros posibilita hacerse una idea bastante completa de la importancia de cada factor para la explicación, de modo que mayores coeficientes son indicadores de una mayor importancia de los niveles de un factor.

Como lo retoma Esteve y Cortina (2005), a partir de los autores Knoke y Burke, los modelos log-lineales son comúnmente usados para analizar las pautas de interacción entre dos o más variables. Los modelos log-lineales, según esos autores, son apropiados, dado que no asumen una relación de causalidad entre variable dependiente e independiente, sino que miden la asociación entre dos o más variables, más allá de lo que se relacionarían por la simple intervención del azar, permitiendo así formulaciones teóricas más flexibles.

Los modelos log-lineales tratan, en este caso particular, del mercado matrimonial de forma integral; es decir, considerando todas las interacciones posibles, sin necesidad de fragmentarlas para ser adaptadas a otro tipo de técnicas. Se trata pues de una visión más cercana a los mercados matrimoniales, que no obliga a fracturar el análisis en múltiples combinaciones o transiciones, que, en la mayoría de casos, son interdependientes. A su vez, los modelos log-lineales descomponen, jerárquicamente, cada uno de los efectos. Por ejemplo, ofrecen parámetros específicos del efecto de pertenecer a un grupo A, el efecto de pertenecer a un grupo B y el efecto de pertenecer a A y B simultáneamente. Es precisamente este último efecto el que se utiliza como un indicador específico de asociación entre dos grupos, libre del efecto de la estructura o de la distribución de los marginales. De modo que los modelos log-lineales permiten examinar las pautas de

asociación entre grupos educativos y así controlar el efecto de la estructura educativa de las uniones de las parejas.

La metodología que seguimos en la especificación de los modelos es la indicada por Park (1991), y se presenta en el siguiente cuadro, mediante la estructura topológica de uniones según las hipótesis que definen cada modelo, por lo que cada una de las celdas con el mismo número suponen tener el mismo *odds ratio* (se usa en estudio retrospectivos), o riesgo relativo (se usa en estudios prospectivos o de cohorte).

		Nivel educativo del marido																								
		Modelo 1					***Modelo 2***					***Modelo 3***					***Modelo 4***					***Modelo 5***				
		Modelo simétrico S					**Modelo de la diferencia en la diagonal principal H**omogamia					**Modelo diagonal simétrico DS**					**Modelo asimétrico A**					**Modelo de hipergamia-hipogamia HH**				
		I	**II**	**III**	**IV**	**V**	**I**	**II**	**III**	**IV**	**V**	**I**	**II**	**III**	**IV**	**V**	**I**	**II**	**III**	**IV**	**V**	**I**	**II**	**III**	**IV**	**V**
Nivel educativo de la mujer	**I**	**1**	6	10	13	15	**1**	6	7	8	9	**1**	2	3	4	5	**1**	2	3	4	5	**1**	6	7	8	9
	II	6	**2**	7	11	14	6	**2**	6	7	8	2	**1**	2	3	4	5	**1**	2	3	4	10	**2**	6	7	8
	III	10	7	**3**	8	12	7	6	**3**	6	7	3	2	**1**	2	3	6	5	**1**	2	3	11	10	**3**	6	7
	IV	13	11	8	**4**	9	8	7	6	**4**	6	4	3	2	**1**	2	7	6	5	**1**	2	12	11	10	**4**	6
	V	15	14	12	9	**5**	9	8	7	6	**5**	5	4	3	2	**1**	8	7	6	5	**1**	13	12	11	10	**5**

I. primaria incompleta II. primaria completa III. secundaria incompleta IV. secundaria completa V. con estudios superiores

Los modelos mencionados anteriormente examinan distintas hipótesis, partiendo de la hipótesis de independencia. El modelo de independencia o modelo uniforme, en términos de homogamia educativa, asume que no existe relación entre la educación del hombre y la educación de la mujer y que, por tanto, la educación no es una variable por tener en cuenta en la selección de la pareja (queremos saber si la proporción de casos para cada categoría de una de las variables es independiente del valor que toma la otra variable). Este supuesto equivale a decir que la distribución de las parejas por nivel educativo de los cónyuges es el resultado del azar y solo estaría condicionado por los efectivos de hombres y mujeres por nivel educativo.

Para el caso nuestro, tenemos una tabla de contingencia de doble entrada, donde ambas variables comparten exactamente las mismas categorías; es apropiado explorar la condición de independencia, que asume independencia en todas las celdas excepto en las de la diagonal; esto es, las celdas que recogen las parejas homógamas. El modelo independiente es en el que las proporciones de probabilidad para cada una de las celdas son básicamente iguales pero diferentes en un nivel.

Las frecuencias esperadas derivadas de este modelo (independiente), con base en una tabla de doble entrada, se estiman según esta expresión:

$$\log f_{ij} = \mu_0 + \mu_i + \mu_j$$

donde:

$\log f_{ij}$ es el logaritmo natural de la frecuencia esperada de la fila *i* columna *j* ;
μ_0 es la constante;
μ_i es parámetro para la fila *i*;
μ_j es parámetro para la columna *j*

En el otro extremo, tenemos el modelo saturado, que asume una interacción específica para cada una de las combinaciones, o sea, una asociación específica para cada una de las combinaciones de los grupos educativos, calculando, por tanto, un parámetro para cada una de ellas. La expresión matemática de este modelo es:

$$\log f_{ij} = \mu_0 + \mu_i + \mu_j + \mu_{ij}$$

donde:

$\log f_{ij}$ es el logaritmo natural de la frecuencia esperada de la fila *i* columna *j* ;
μ_0 es la constante;
μ_i es parámetro para la fila *i*;

μ_j es parámetro para la columna *j*

μ_{ij} es el parámetro de interacción entre la fila *i* la columna *j*

El modelo saturado, aunque tiene la particularidad de reproducir exactamente los datos, no tiene interés analítico, puesto que consume tantos parámetros como interacciones quiere explicar. Se trata, por lo tanto, de un modelo con parsimonia [41] nula.

Sin embargo, entre el modelo independiente y el modelo saturado existen numerosas especificaciones que, con más o menos parsimonia, tiene interés explicativo en términos sustantivos, los cuales se han definido en el cuadro de la estructura tipológica de la selección de parejas y que son el sustento esencial de interés en la investigación.

El modelo de **asociación simétrica (S)** igualdad en estudios independiente de la combinación o el nivel, comprueba que para aquellas interacciones entre categorías distintas el orden de las variables influye. Si existiera simetría, por ejemplo, las celdas (4,5) y (5,4) mostrarían el mismo nivel de asociación (interacción). Esto significa que la probabilidad de una unión entre personas con distinta categoría de instrucción es independiente del hecho de que sea el hombre o la mujer el que tenga mayor o menor nivel. Para comprobar si existe esta tendencia uniforme en uno de los sexos para unirse hacia arriba o hacia abajo, se calcula la asimetría (A).

El modelo de **asociación en la diagonal simétrica (DS)** asume que la probabilidad de una unión varía en función de cómo se aleja de la diagonal; es decir, que cuanto más dispar es la educación del hombre respecto a la de la mujer o viceversa, más improbable es que exista una unión con esas características. Asume que el patrón de emparejamiento viene de diferente nivel educativo y la probabilidad con un nivel aparte decrece. Así pues, se espera que el parámetro 1 sea mayor al parámetro 2, el 2 mayor al 3, y así sucesivamente.

---

[41] Parsimonia se refiere al grado de sencillez de un modelo: cuanto menos parámetros son necesarios para explicar un fenómeno, más parsimonioso es el modelo.

El modelo de la **diferencia en la diagonal principal de la asociación simétrica (H)** especifica la educación homogamia [42] entre diferentes niveles. Las celdas situadas en la diagonal representan las proporciones de parejas homógamas, aquellas en que ambos cónyuges tienen la misma categoría o nivel educativo.

El modelo asimétrico **(A) de asociación de la diagonal no simétrica** nos indica que la simetría en las diagonales de afuera reflejan los niveles educativos hacia arriba de las parejas (hipergamia [43]) y niveles educativos de emparejamiento hacia abajo (hipogamia [44]), la cual muestra procesos que deben ser examinados separadamente.

El modelo de **asociación no simétrica en la diagonal diferente (HH)** nos señala que los niveles de homogamia, hipergamia o hipogamia son diferentes; por lo tanto, pueden ser examinadas cada una separadamente.

La estructura y resultados de los principales modelos deben ser evaluados mediante los estadísticos razón de verosimilitud entre los grados de libertad (LR^2 = Likelihood Ratio) y el *Criterio Bayesiano de Información* (Bayesian Indicator Criteria BIC), que informan sobre los ajustes por realizar, para lo cual, en ambos casos, cuanto menor sea el valor de estos indicadores, mejor es el ajuste y, por lo tanto, mejor es la capacidad explicativa del modelo.

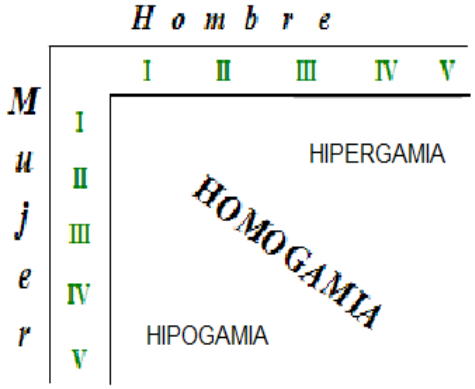
Para efectos del análisis de los datos y los respectivos modelos multivariados sobre la relación en el nivel educativo de las parejas, se cuenta con los siguientes tipos de medición de la homogamia educativa en las parejas (según lo resume Park a partir de los modelos planteados):

---

[42] La homogamia se refiere a la situación donde las mujeres forman pareja con hombres que tiene igual educación.

[43] La hipergamia refiere a la situación en el cual las mujeres forman pareja con hombres que tiene más alta educación.

[44] La hipogamia refiere a la situación el cual las mujeres forman pareja con hombres que tienen más baja educación.



1. ***Homogamia:*** indica por definición las parejas donde sus miembros presentan el mismo nivel educativo.

2. ***Heterogamia*:** señala a los miembros de la pareja que difieren en algún nivel educativo. Dentro de esta categoría existen variantes que se asumieron de la siguiente manera:

α) ***Hipergamia*** refiere a la situación en el cual las mujeres forman pareja con hombres que tiene más alta educación; es decir, las mujeres poseen una educación más baja (inferior) que la de hombre; por lo tanto, las diferentes combinaciones nos muestran que los miembros de la pareja difieren en uno, dos, tres y hasta cuatro niveles educativos, los datos de la condición de hipergamia se encuentran situados en la triangular superior de la matriz de datos.

La diferencia se determina en este caso a partir de la ubicación del nivel educativo del hombre jefe de hogar, donde este posee un nivel educativo superior y la mujer o cónyuge tiene uno, dos, tres y hasta cuatro niveles inferiores respecto de la del jefe de hogar; por lo tanto, si el hombre posee un nivel educativo de universidad: la diferencia de ***un nivel*** significa que la cónyuge se encuentra en un nivel educativo de secundaria completa; si la diferencia es de ***dos niveles,*** la cónyuge está en secundaria incompleta, y así sucesivamente. En síntesis, el punto de base para la definición de los diferentes niveles de diferencia se realiza conforme al nivel educativo de la jefatura de hogar.

β) ***Hipogamia*** refiere a la situación en la cual las mujeres forman pareja con hombres que tienen más baja educación; es decir, las mujeres poseen una educación más alta (superior) que la de hombre; por lo tanto, los miembros de la pareja difieren en uno, dos, tres y hasta cuatro niveles educativos; representados en la matriz de datos en la triangular inferior.

La diferencia se determina en este caso a partir de la ubicación del nivel educativo de la mujer, donde la mujer posee un nivel educativo superior y el hombre o jefe de hogar tiene uno, dos, tres y hasta cuatro niveles inferiores respecto de la mujer; por lo tanto, si la mujer posee un nivel educativo de universidad: la diferencia de ***un nivel*** significa que el hombre se encuentra en un nivel educativo de secundaria completa; si la diferencia es de ***dos niveles,*** el hombre está en secundaria incompleta, y así sucesivamente. En síntesis, el punto de base para la definición de los diferentes niveles de distinción se realiza conforme al nivel educativo de la mujer.

Existen diversas alternativas para presentar y comentar los resultados obtenidos mediante la aplicación de modelos loglineales. En este trabajo, por coherencia argumental, hemos decidido clasificar estos modelos en dos grandes bloques. Con el primer bloque, formado por los modelos 1 y 3, damos respuesta a la siguiente pregunta: ¿en qué medida las personas se unen de forma homógama; es decir, con personas de su mismo nivel de instrucción? El segundo bloque está formado por los modelos 2, 4 y 5. Estos modelos ponen el acento en la obtención de parámetros para evaluar la vigencia de la hipergamia (parejas en la que la mujer tiene un nivel de instrucción inferior al del hombre), o la hipogamia.

Ninguna de las hipótesis presentadas tendría razón de ser verificada si, previamente, no comprobamos que la condición del modelo de independencia no se satisface; es decir, que el nivel de instrucción de los individuos es relevante en la selección conyugal y, por lo

tanto, las frecuencias observadas no son simplemente el resultado de la combinación azarosa entre los grupos educativos. Bajo el supuesto de independencia, controlados los efectos de la estructura, el riesgo de una unión entre una mujer analfabeta y un hombre con estudios universitarios sería exactamente igual a la de una unión entre universitarios. Pero el deficiente ajuste del modelo de independencia obliga a rechazar dicho supuesto e invita a explorar nuevas formulaciones, que expliquen mejor los datos observados. Se justifica así la razón de ser de los modelos indicados por Park, como veremos a continuación.

3.4 DATOS Y RESULTADOS

3.4.1 Nivel educativo del jefe del hogar y su cónyuge

Al efectuar el procesamiento de la Encuesta de Hogares del año 2004 acorde con los propósitos de nuestro estudio resultaron 6.703 casos de interés, que se traducen en 6.703 parejas de la población muestral, distribuidas de la siguiente manera:

Nivel Educativo	Cónyuge %	Jefe %
Total	**100,00**	**100,00**
Nº casos	6.703	6.703
Primaria incompleta	25,50	27,30
Primaria completa	36,00	35,90
Secundaria incompleta	16,20	15,00
Secundaria completa	11,00	9,20
Universidad	11,30	12,60

Como se observa del cuadro anterior, el 61,5 % de las mujeres cónyuges tienen primaria completa o un nivel inferior; para los hombres jefes de hogar resultó en 63,2 %. Las mujeres en el nivel educativo de secundaria completa y de universidad alcanzan un porcentaje similar al 11% para ambas categorías.

La categoría que más predomina es el nivel educativo de primaria completa para ambos sexos, y la categoría que contiene el menor peso es la opción de secundaria completa para ambos, aunque para los hombres es aún más baja, con 9,2 %, respecto de las mujeres, con un 11 %.

Del total de parejas, el resultado de la lectura de las combinaciones que se presentan en el cuadro 3.5 se observa que una baja instrucción para ambos miembros de la pareja es lo más frecuente. Destaca el nivel educativo de primaria completa para ambos miembros, con 18%, seguida de la combinación del nivel educativo de primaria incompleta para ambos miembros con 16%; la combinación del nivel educativo de primaria incompleta para el jefe de hogar y primaria completa de la cónyuge es de 9%. Del total de la muestra el 77,7% cuenta con un nivel educativo bajo, ubicado en categorías menores al nivel educativo de secundaria completa.

En contraposición, las combinaciones con valores más bajos son, a saber: las combinaciones extremas o de las esquinas para la tabla de contingencia, definidas por el nivel educativo universitario del jefe de hogar con nivel educativo primaria incompleta de la cónyuge; el nivel educativo universidad de la cónyuge con nivel educativo primaria incompleta del jefe de hogar, y la interacción entre el nivel educativo de secundaria completa del jefe del hogar y el nivel educativo de primaria incompleta para la cónyuge con porcentajes menores al 1% en las combinaciones descritas. Lo anterior muestra un pequeño cambio en las creencias sobre la construcción de pareja que tienen las mujeres, al tener una proporción más alta que la de los hombres respecto a conformar pareja con un cónyuge que posee más bajo nivel educativo.

Cuadro 3.5
COSTA RICA: Distribución relativa de la población muestral
SEGÚN: Nivel educativo del cónyuge femenino
POR: Nivel educativo del jefe de hogar masculino
2004 (Cifras relativas)

Nivel educativo cónyuge (mujer)	TOTAL	Nivel educativo jefe de hogar (hombre)				
		Primaria incompleta	Primaria completa	Secundaria incompleta	Secundaria completa	Universidad
TOTAL	**100,00**	**27,35**	**35,86**	**15,04**	**9,16**	**12,59**
Primaria incompleta	25,54	15,59	7,67	1,82	0,25	0,21
Primaria completa	35,95	9,16	17,96	5,27	2,21	1,36
Secundaria incompleta	16,23	1,57	6,15	4,45	2,10	1,97
Secundaria completa	10,98	0,58	2,58	2,10	3,01	2,70
Universidad	11,29	0,45	1,51	1,40	1,58	6,36

Para análisis de los modelos, debe previamente probarse que no se satisface el supuesto de independencia; es decir, que la composición de las parejas es dependiente del nivel de instrucción de las personas y, por lo tanto, existe una asociación entre el nivel educativo del jefe y del cónyuge donde las frecuencias observadas no son simplemente el resultado de la combinación azarosa entre los individuos. Los resultados fueron los siguientes:

Indicadores estadísticos del modelo independiente

BIC	Razón de verosimilitud (LR^2)	g.l	LR^2 / g.l	Resultado del modelo
6022,901	3102,02	8	387,75	Modelo no se ajusta

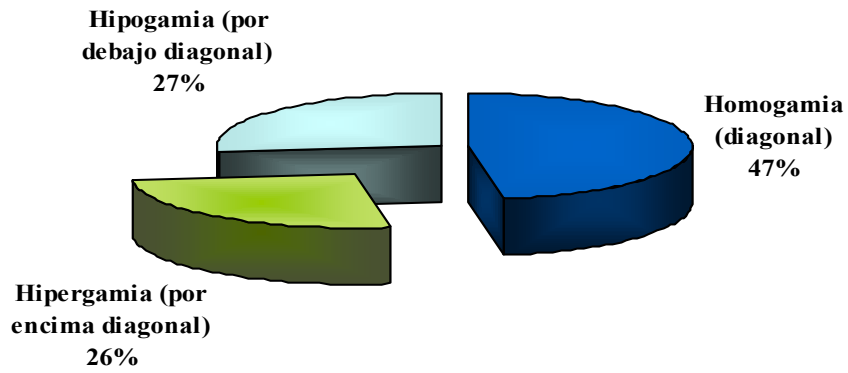
Como se puede observar los estadísticos anteriores indican que el modelo independiente no se ajusta, lo que obliga a rechazar el supuesto de independencia, indicando que existe una relación significativa entre el *nivel educativo*, el cual posibilita la exploración de los otros modelos (Simétrico **S**, Diagonal simétrico **DS**, Diferencia en la diagonal principal **H**, Asimétrico **A** e Hipergamia- hipogamia **HH***),* que puedan explicar mejor los datos observados.

3.4.2 Presencia de homogamia educativa

En cuanto a las diferencias entre el nivel educativo de los jefes y cónyuges, los datos muestran que del total de las 6.703 parejas el 47,4% elige una pareja con el mismo nivel educativo (homógamas) ver gráfico 3.1. El resto está en la categoría de heterogamia, con un 52,6%. De modo que de 3,528 parejas que están en la condición de heterogamia (diferente nivel educativo en la pareja), la categoría de hipergamia acumula un total de 48,6%, en contraste con la hipogamia, que resultó en un 51,4%. El comportamiento bastante parecido entre la hipergamia e hipogamia nos anuncia la presencia de un leve cambio en el comportamiento de la composición de las parejas, según el nivel educativo.

Gráfico 3.1

Distribución de la población según homogamia educativa de la pareja en Costa Rica para el año 2004



Una primera lectura del cuadro 3.6 nos indica que las parejas homogámicas con valores más altos son las parejas en los extremos del nivel educativo; por un lado, las de primaria completa o un nivel inferior educativo; por otro lado, las que poseen el nivel superior universitario.

Así como la distribución de la población en el nivel de homogamia, las categorías que representan los valores más altos en la línea de la diagonal se ubican en primaria

incompleta con un 57% respecto del total de esa misma categoría, seguido por valores cercanos al 50% en primaria completa y estudios superiores en relación con el total de su categoría.

Cuadro 3.6
COSTA RICA: Distribución de la población muestral
SEGÚN: Nivel educativo del cónyuge femenino
POR: Nivel educativo del jefe de hogar masculino
2004 (Cifras aboslutas y relativas)

Nivel educativo cónyuge (mujer)	Nivel educativo jefe de hogar (hombre)					
	TOTAL	Primaria incompleta	Primaria completa	Secundaria incompleta	Secundaria completa	Universidad
Números absolutos						
TOTAL	**6.703**	**1.833**	**2.404**	**1.008**	**614**	**844**
Primaria incompleta	1.712	1.045	514	122	17	14
Primaria completa	2.410	614	1.204	353	148	91
Secundaria incompleta	1.088	105	412	298	141	132
Secundaria completa	736	39	173	141	202	181
Universidad	757	30	101	94	106	426
Números relativos						
TOTAL	**100,00**	**100,00**	**100,00**	**100,00**	**100,00**	**100,00**
Primaria incompleta	25,54	**57,01**	21,38	12,10	2,77	1,66
Primaria completa	35,95	33,50	**50,08**	35,02	24,10	10,78
Secundaria incompleta	16,23	5,73	17,14	**29,56**	22,96	15,64
Secundaria completa	10,98	2,13	7,20	13,99	**32,90**	21,45
Universidad	11,29	1,64	4,20	9,33	17,26	**50,47**

También se aprecia en el cuadro 3.7 que las proporciones con valores más altos se encuentran en las categorías de más bajo nivel educativo (primaria incompleta y completa) para ambas la homogamiay la heterogamia, en relación con el total de cada nivel educativo; es decir, las proporciones disminuyen al incrementarse el nivel educativo por la estructura educacional de la sociedad.

Cuadro 3.7
COSTA RICA: Distribución de parejas
SEGÚN: Nivel educativo
POR: Homogamia y heterogamia
2004

Nivel educativo	Total	Homogamia	Heterogamia	
			Hipergamia	Hipogamia
Total	**100,0**	**47,4**	**25,6**	**27,1**
Nº casos	6.703	3.175	1.713	1.815
Primaria incompleta	25,5	32,9	38,9	0,0
Primaria completa	36,0	37,9	34,6	33,8
Secundaria incompleta	16,2	9,4	15,9	28,5
Secundaria completa	11,0	6,4	10,6	19,4
Universidad	11,3	13,4	0,0	18,2

Al determinar las diferencias del nivel educativo entre los jefes de hogar y sus cónyuges en la formación de pareja, tenemos que para la condición de ***hipergamia***, caracterizada porque el jefe de hogar posee un nivel educativo superior respecto a su cónyuge, se obtuvo 69% con un nivel de diferencia, lo cual significa que los hombres cuentan con máximo un nivel superior de diferencia respecto del cónyuge. Los datos reportan 23% en la diferencia en dos niveles, 6% con tres niveles de diferencia y menos de un 1% para cuatro niveles de diferencia. En este sentido, de los datos se deriva que la diferencia o brecha en el nivel educativo de la mayoría de parejas donde el jefe de hogar tiene un nivel educativo superior es baja, además que la proporción de parejas con el máximo de niveles de diferencia es a su vez bastante baja.

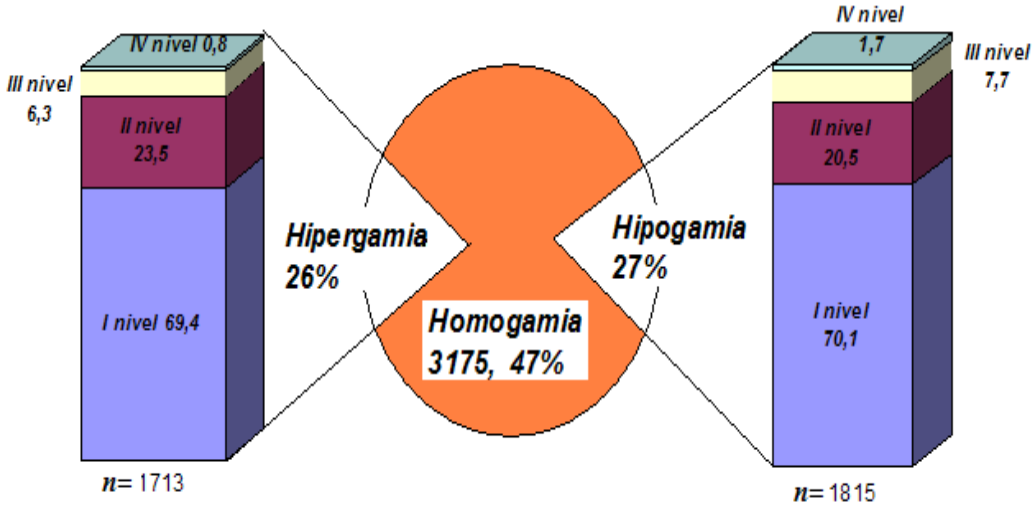
Para la condición de ***hipogamia,*** que es cuando el cónyuge tiene un nivel educativo superior al del jefe de hogar, se totalizan 1,815 parejas, de las cuales 70% tiene un nivel de

diferencia, 20% con dos niveles de diferencia y, 8%, y 2%, con tres y cuatro niveles de diferencia respectivamente (ver gráfico 3.2).

De todo lo anterior, se desprende que el comportamiento de la hipergamia y la hipogamia es similar; la diferencia en los niveles de instrucción se mueve en escala descendente; los valores altos para ambos comportamientos se sitúan en la diferencia de un solo nivel en la composición de las parejas, valores bajos en el extremo de cuatro niveles de diferencia (véase anexo para mayor detalle).

Gráfico 3.2

Distribución de la homogamia e hipergamia de las parejas según niveles de diferencia



Fuente: Cuadro anexo.

3.4.3 Análisis de los modelos multivariados

El análisis log-lineal realizado enfoca primeramente su atención en identificar un modelo apropiado para el emparejamiento educativo, basado en los modelos expuestos en la sección metodológica. Esto exige necesariamente la interpretación o evaluación de la bondad de ajuste de los modelos, mediante los indicadores estadísticos razón de verosimilitud (LR^2 = Likelihood Ratio) y el Criterio Bayesiano de Información (BIC), que informan sobre los ajustes por realizar; cuanto menor es el valor de estos indicadores, mejor es el ajuste y, por lo tanto, mejor es la capacidad explicativa del modelo.

Al no cumplirse el supuesto de independencia, como se dijo anteriormente, se posibilita seguir analizando la composición de las parejas en relación con el nivel de instrucción de las personas, según los modelos definidos. Un resumen de los indicadores de los estadísticos, que permite evaluar los ajustes de los modelos investigados, se presenta seguidamente.

Indicadores estadísticos de los modelos

Modelos	BIC	Razón de verosimilitud (LR^2)	g.l	LR^2 / g.l	Modelo que se ajusta
Simétrico S	2805,6	6325,1	14	451,8	
Diagonal simétrico DS	2866,2	4074,8	4	1018,7	
Diferencia en la diagonal principal,	2808,9	5402,8	8	675,4	
Asimétrico A	2861,9	3630,2	7	518,6	
Hipergamia- hipogamia HH	2803,4	5416,9	12	451,4	*

Como se observa, el modelo de **asociación no simétrica en la diagonal diferente (HH)** (valores para BIC=2803,4 y LR^2/g.l=451,4) resulta ser el modelo que se ajusta, lo que sugiere que el emparejamiento se realiza siguiendo un patrón diferente en el nivel

educativo de los miembros, indistintamente de quién tenga mayor o menor nivel de instrucción; es decir, que los niveles de homogamia, hipergamia o hipogamia son diferentes. Ante este escenario se requiere examinar la diferencia en los niveles de hipergamia y de hipogamia, separadamente.

Para verificarla diferencia entre el modelo de hipergamia y de hipogamia, se efectuó una prueba de hipótesis del modelo de asimetría que hace diferencia entre ellas (hipótesis nula: el modelo se ajusta y la hipótesis alternativa: el modelo no se ajusta); los resultados de la prueba se exponen en el cuadro siguiente, donde los datos muestran evidencia de favorecer el modelo de hipogamia (vp= 0,01; con el coeficiente negativo), que se refiere a la situación en la cual las mujeres forman pareja con hombres que tienen menor nivel educativo formal, o sea, dicho modelo enfoca la atención en un cambio en la unión de las parejas, dado que las mujeres no perfilan su elección de pareja en hombres con mayor educación, que les permita consolidar o mejorar su condición económica y social individual o familiar, porque ellas mismas se las pueden proveer. Ello, por cuanto la variable educación, además de reflejar la calificación de los individuos que conforman la pareja, se asocia con la posición económica y su capacidad de promoción social y profesional

Resultados de estadísticos de prueba para el modelo hipergamia vs hipogamia

Coeficente	Error Estándar	z	P > | z |	Intervalos de confianza	
				Izquierdo	Derecho
-1,01	3,557	-2,83	0,01	-1,713	-0,311

Retomando la divergencia entre los planteamientos de las teorías del mercado matrimonial, postuladas por los autores Becker, Oppenheimer y Binstock, los datos dan razón a los dos últimos autores, donde según los datos de 2004 el modelo de hipogamia que predomina en Costa Rica se explica mediante el crecimiento de la incorporación de la mujer al mercado de trabajo, que al proveerle de una independencia económica obliga a

replantear el principio de complementariedad y del patrón de elección en la formación de pareja; a su vez, el nivel de instrucción de la mujer pasa a ser tan valorado como el del hombre, por lo que se esperaría en este nuevo contexto un aumento de los niveles de hipogamia.

Entonces, en las sociedades contemporáneas se viene reconfigurando la función tradicional que se le ha asignado al varón como proveedor de ingresos, dedicado la mayor parte de su tiempo al trabajo productivo, en contraposición con la incursión de la mujer al trabajo productivo, pasando al terreno que antes era dominio de los varones y dejando las tareas reproductivas; los patrones de participación económica en el trabajo productivo ubican a la mujer en una posición potencial económica. Así, la serie de tiempo de los datos de las Encuestas de Hogares de Propósitos Múltiples muestran para Costa Rica, que la participación de la mujer en el mundo laboral ha crecido. En ese sentido, sobre la evolución de la *población económicamente activa* femenina *(*PEA o fuerza de trabajo que incorpora ocupados, desocupados cesantes y los que buscan empleo por primera vez): en los años cincuenta era del 15%, en los sesenta 16%, en los setenta 19%, en los ochentas del 22%, en los noventa en el orden de 30%, en la década del 2000 alcanzó un 34%, y cinco años más tarde (2005) la PEA femenina se mantiene en un tercio del total de la PEA.

Lo anterior, aunado a la creciente asistencia en el ámbito educativo, tornan a la mujer más atractiva en el mercado matrimonial, donde ambos miembros de la pareja pueden beneficiarse a partir de reunir y compartir los recursos. En este sentido, se explica el modelo de heterogamia educacional, con predominio de la hipogamia, ya que la mujer al contar con mayor educación y por el aumento de su contribución potencial a la economía del hogar, le provee de independencia económica, lo cual implica que la posición del varón asociada al nivel educativo cobra menor relevancia en la formación de pareja. Dicha interpretación no agota la explicación ya que de acuerdo con el modelo de análisis, se puede controlar por otras variables explicativas, como es ciclo de vida de la pareja, haciendo diferenciación entre la población económicamente activa e inactiva, edad, religión

preferencia política, entre otras; ello, porque la interpretación de los fenómenos sociales son complejos y responden a modelos multicausales.

Finalmente, diferente situación se presenta al comparar los datos para la sociedad coreana de los setentas desarrollado en el artículo de Park, donde se examinaron los patrones y tendencias en la asociación entre esposas y esposos del logro educativo durante 1950-79 en Corea, con base en varios censos (que abarcaron varias cohortes). En el cual se concluyó que el matrimonio homogámico fue la forma predominante de apareamiento educativo, comparado con matrimonios hipergámicos o hipogámicos durante 1970. Este resultado es consistente con la perspectiva del conflicto, donde se argumenta la consolidación de los estratos dominantes de la cultura de clase, a través de los matrimonios homogámicos. *Desde 1970-79, los que se casan hacia arriba aumentó en Corea (hipergamia), sosteniendo el patrón tradicional de matrimonio mientras se experimenta la industrialización en Corea En conjunto, el incremento de homogámico educativo entre familias recién forjadas implica un incremento en el efecto del fondo de los padres y en el logro de socio-económico de niños, que ayuda a mantener y reproducirse el orden existente de estatus* (Park: 1991).

3.4.4. Sobre la reproducción de la estratificación a partir del nivel educativo de la pareja

Con el propósito de analizar el comportamiento en la estratificación de clase a partir del efecto del nivel de instrucción de la pareja, utilizamos la variable situación de pobreza (referido a las condiciones materiales con que cuenta un hogar), que proviene de la variable nivel de pobreza de la Encuesta de Hogares, el cual cuenta con limitaciones conceptuales, debido a que la medición del ingreso no contempla otras variables como las ayudas estatales (según declaraciones del Vicepresidente del INEC, la nueva medición de pobreza se aplicará hasta el 2008); no obstante, son los datos con que se disponen.

En ese sentido, se calcula la probabilidad de que las parejas sean pobres en relación con los niveles educativos de los miembros de la pareja, para acercarnos a la explicación de la asociación existente entre el grado o nivel de instrucción de los padres con la reproducción de la estratificación social de la progenie.

En el cuadro 3.8 se destacan las probabilidades de mayor valor, que corresponden a las parejas con menor instrucción. Así, por cada diez parejas donde ambos miembros cuentan con nivel de instrucción primaria incompleta (vp = 0.398), cuatro de esas parejas se encuentran en situación de pobreza. A partir de ello, se puede inferir que la pobreza esta relacionado con el nivel de educativo de la pareja, y sería muy posible que sus hijos(as) tenderán a moverse en esos espacios caracterizados por la escasez material, dada la condición de sus padres, (...) *el hecho de que las oportunidades de educación y, por consiguiente, las de acceso a empleos más estables y mejor remunerados, sean en alto grado heredadas, constituye un elemento clave para que las desigualdades socioeconómicas actuales se reproduzcan indefinidamente en las sucesivas generaciones más jóvenes. En efecto, la probabilidad de recibir un mínimo adecuado de educación está determinada en gran medida por el grado de educación de los padres y por la capacidad económica del hogar de origen* (CEPAL: 2004-1, 27).

Cuadro 3.8
COSTA RICA: Probabilidades asociadas al nivel de pobreza respecto del nivel educativ o
2004

Nivel educativo jefe / Nivel educativo	Total	Primaria incompleta	Primaria completa	Secundaria incompleta	Secundaria completa	Universitaria
Total	0,225	0,353	0,263	0,178	0,084	0,016
Primaria incompleta	0,365	**0,398**	**0,339**	**0,343**	0,154	0,100
Primaria completa	0,235	**0,278**	**0,271**	0,175	0,127	0,020
Secundaria incompleta	0,195	**0,352**	**0,218**	**0,215**	0,078	0,021
Secundaria completa	0,096	0,118	0,128	0,106	0,099	0,028
Universidad	0,020	0,000	0,055	0,053	0,017	0,005

Como se observa en el resumen de la estimación de los modelos, los indicadores estadísticos muestran que al comparar la bondad de ajuste del modelo original sin la variable situación de pobreza con el modelo original con la variable pobreza presenta una mejor bondad de ajuste.

Seguidamente, al comparar el modelo original sin la variable situación de pobreza con el modelo que incorpora la variable situación de pobreza, se define un patrón diferente para los pobres y no pobres; se observa que este último es el que señala un mejor ajuste (BIC=5.95, LR/g.l=317.69). En ese sentido, podemos decir que los patrones de las parejas son diferentes entre pobres y no pobres.

Resumen de las estimaciones de los modelos para el nivel de pobreza

Modelos	Variable	BIC	Razón de verosimilitud (LR^2)	g.l	LR^2 / g.l	Modelo que se ajusta
Hipergamia-hipogamia HH	Modelos originales sin variable situación de pobreza	2803,44	5416,94	12	451,41	
Simétrico S	Modelos originales con variable situación de pobreza	679,56	8472,75	15	564,85	
Simétrico S	Modelos con variable situación de pobreza y patrones particulares	-5,95	9213,03	29	317,69	*

Lo anterior nos posibilita reafirmar nuestra perspectiva de que la educación como referencia del nivel de escolaridad de las parejas contribuye a determinar y especifica la ubicación en un grupo social, según las características que tengan los miembros de la pareja, lo que podría ser transmitido a los hijos. Entonces, el papel de la educación esta asociado a la dinámica propia de la reproducción, de relaciones de clases y de producción de bienes materiales y simbólicos, tal y como lo plantean autores como Bourdieu y Passeron, Varela y otros.

La reproducción social obedece a diversos factores; sin embargo, en este caso dada la relación entre la pobreza y el nivel educativo de la pareja nos lleva a que la educación se asocia no solo con la calificación de los individuos, sino, también, de su posición económica y de su capacidad de promoción social y profesional; así, la inserción ocupacional de los jóvenes refleja la influencia determinante que ejerce la situación socioeconómica y educacional del hogar de origen en las oportunidades de bienestar; diferentes datos de la CEPAL señalan que *habría correlación entre los porcentajes de no superación educativa intergeneracional y la proporción de jóvenes que en los sondeos de opinión manifiestan no tener oportunidades de superar el nivel de vida de sus padres. Cabe destacar también que el porcentaje de jóvenes que no supera la educación de sus padres y tampoco alcanza el capital educativo básico (CEPAL: 2004-2, 36).*

3.5 REFLEXIONES FINALES

Recapitulando sobre el proceso investigativo y el procesamiento de los datos, se desprende lo siguiente:

- En las pautas para la conformación de pareja intervienen una gran cantidad de variables, de modo que es de naturaleza multifactorial; en ese contexto, esta investigación se centra en el estudio de la asociación entre los diferentes niveles educativos de las parejas, como un factor que contribuye a determinar el tipo de uniones.

- Al clasificar a las personas emparejadas en cinco grupos educativos (primaria incompleta, primaria completa, secundaria incompleta, secundaria completa y universidad), jefe hombre-cónyuge mujer, casi la mitad de las parejas se unen con alguien de su mismo grupo educativo, mientras que más de la mitad se unió a parejas con diferente nivel educativo.

- Más de la mitad de las parejas que se identificaron como heterogámicas se distribuyen de la siguiente manera, en función de los niveles de diferencias: en las tres cuartas partes de estos casos la diferencia es de solamente un nivel; menos de una cuarta parte presenta diferencia en dos niveles, y el resto en tres niveles y cuatro niveles de diferencia.

- A más de la mitad de las parejas se les clasifica como heterogámicas; en un poco más de la mitad de estas parejas, las mujeres son quienes tienen un nivel superior educativo.

- Al analizar los diferentes modelos que diferenciaron la homogamia de la heterogamia, el parámetro de asimetría fue significativo, o sea, que la interacción en la pareja es asimétrica, indicando así que la hipergamia pierde vigencia ante la relevancia de la hipogamia.

- Pese a la disminución en parejas homógamas respecto de la parejas hipergámicas, como consecuencia de la creciente incorporación de la mujer al mercado laboral, existe una propensión alta a unirse dentro del mismo grupo educativo.

- El patrón tradicional o el predominio de la hipergamia donde el hombre tenía mayor nivel educativo y las mujeres tenían un nivel de instrucción menor, se hace cada vez menos frecuente, lo cual revela que la hipergamia pierde vigencia para el año en estudio. En ese sentido, parece relevante replicar el estudio para años anteriores y posteriores, de modo que permitan apreciar diferencias de largo alcance.

- Es importante rescatar el supuesto de independencia del modelo, que nos permitió explorar los patrones y tendencias del nivel educativo en la composición de las parejas, dado que la hipotésis de partida planteada fue que la correspondencia azarosa entre los cónyuges estaba condicionada por la distribución por nivel de instrucción de hombres y mujeres, y no, únicamente por el efectivo de hombres y mujeres, donde la educación de los cónyuges es un aspecto relevante en la composición de las parejas.

- Las probabilidades de mayor valor respecto a la situación de pobreza de las parejas se encuentran ubicadas en las parejas con menor instrucción; así, tenemos que para diez parejas en donde ambos miembros cuentan con nivel de instrucción primaria incompleta, la probabilidad (vp = 0.398) indica que cuatro de esas parejas se encuentra en una situación de pobreza.

- Al comparar la bondad de ajuste de los modelos en relación con la situación de pobreza y de categorías que establecen diferencia entre patrones de pobres y no pobres, resultó significativo el modelo que plantea que sí existe una diferencia entre los patrones de parejas pobres y parejas no pobres.

- De la relación entre la variable pobreza y el nivel educativo de la pareja, los resultados proporcionaron evidencia para reafirmar nuestra perspectiva de que la educación formal, como referencia del nivel de escolaridad de las parejas, cumple un

papel muy importante en la ubicación en el grupo social, según las características que tengan los miembros de la pareja y que podría tener incidencia en la situación de las oportunidades que serán transmitidas a sus hijos(as).

- Queda por hacer comparaciones entre generaciones, aplicar modelos para probar la diferencia entre niveles, incluir otras variables relacionadas con el calendario nupcial sobre los patrones de unión, como la postergación en las edades para unirse en cada generación sucesiva (y diferencia de edades entre los cónyuges), etc. Sin embargo, análisis de este tipo contribuyen a acercarse a una descripción sobre la formación de parejas.

A. BIBLIOGRAFÍA

Becker Gary, 1987. **A theory of mariage timing. Journal of Political Economy**.

Binstock Georgina, 2004. **Cambios en las pautas matrimoniales en Buenos Aires: desentrañando el efecto de la educación**. Trabajo presentado en I Congreso de Asociación de Población de Latinoamérica, setiembre 2004. Caxambú – Brasil. Investigadora del Centro de Estudios de Población Buenos Aires. Argentina.

Bourdieu Pierre, y Passeron Jean, 1996. **La reproducción. Elementos para una teoría del sistema de enseñanza**. Editorial Laia S. A. México.

Castro V. Carlos, Gutiérrez E. Ana Lucía, Rodríguez S. Carlos y Barahona M. Manuel, 2007. **Transformaciones en la estructura social en Costa Rica**. Universidad Costa Rica, Costa Rica.

Catena Andrés, Ramos Manuel y Trujillo Humberto, 2003. **Análisis multivariado**. Editorial Biblioteca Nueva Madrid. España.

CEPAL, 2004a. **Una década de desarrollo social en América Latina, 1990-1999**. Libro N.° 77, Santiago de Chile, Chile. Consultado en enero 2008 en http://www.cepal.org/publicaciones.

CEPAL, 2004b: **Transmisión intergeneracional de las oportunidades de bienestar**. Libro N.° 77, Santiago de Chile, Chile. Consultado en enero 2008 en http://www.eclac.org/publicaciones/xml/9/4649/Capitulo_IV_1997.pdf.

Chen Mario, Rosero Luis, Brenes Gilbert, León Miriam, González Ma. Isabel, Pissa Juan Carlos, 2001. **Salud reproductiva y migración nicaragüense en Costa Rica 1999-2000: resultados de una encuesta nacional**. Programa Centroamericano de Población (PCP) e Instituto de Investigaciones en Salud (INISA), Universidad de Costa Rica. San José, Costa Rica.

Estado de la Educación Costarricense, 2005. **Primer Informe sobre el Estado de la Nación en Desarrollo Humano Sostenible**. Costa Rica.

Esteve Albert y Cortina Clara, 2005. **Homogamia educativa en la España contemporánea y tendencias**. Consultado en mayo del 2006 en www.ced.vab.es

García M. Tomás, 2001. **Estratificación social y educación.**

González G. Yamileth, 2003. **Educación diversificada y humanista para una democracia integral (1950-1970).** Capítulo V, "Historia de la educación costarricense". Editorial UNED, Costa Rica, págs. 269-350.

INEC, 2004 (Instituto Nacional de Estadística y Censos). **Encuesta de Hogares de Propósitos Múltiples julio 2004**. Instituto Nacional de Estadística y Censos, San José, Costa Rica.

Kalmijn M., 1998. **Intermarriage and Homogamy: Causes, Patterns, Trends. Annual.** Review of Sociology, 24, pp.395-421.

Kalmijn M., 1991. **Status Homogamy in the United States**. American Journal of Sociology. 97, pp.496-523.

Knoke D. y Burke P., 1980. **Log Linear Models**, Beverly Hills, Sage Publications.

MIDEPLAN, 2007 (Ministerio de Planificación Nacional y Política Económica) **Organización Familiar: Matrimonios, Divorcios, Nacimientos y Pensiones Alimenticias Período: 1975 - 2005**. Costa Rica. Consultado en agosto de 2007 en http://www.mideplan.go.cr/sides/social/08-01.htm.

Oppenheimer Valerie, 1988. **A theory of mariage timing**. The American Journal of Sociology. Vol. 94, N.º 3. pp.563-591.

Park Mee-Hae, 1991. **Patterns and trends of educational mating in Korea.** Korea Journal of population and development. Vol. 20. N.º 2. December.

Piani Georgina, 2003. **¿Quién se casa con quién? Homogamia educativa en las parejas de Montevideo y Zona Metropolitana**. Universidad de la República, Facultad de Ciencias Sociales - Departamento de Economía.

Rodríguez, Carlos, 1997. **Los efectos del ajuste: estratificación y movilidad social en Costa Rica en el período de 1950-1995**. Tesis para optar por el grado de Doctor en Sociología, Programa de Doctorado en Ciencias Sociales, México.

Rojas Yolanda y Retana P., 2003. **Educación para el desarrollo: del intervencionista al neoliberalismo.** Capítulo VIII, "Historia de la educación costarricense". Editorial UNED, Costa Rica, págs. 435-493.

Ruiz Ángel, 2005. **Universalización de la educación secundaria y reforma educativa.** XI Informe sobre el Estado de la Nación en Desarrollo Humano Sostenible.

Therborn Göran, 2004. **Familias en el mundo: historia y futuro en el umbral del siglo XXI**, Serie Seminario y conferencias N.° 42 Santiago, Chile: CEPAL. Consultado en julio 30 2007 en http://www.cepal.org/publicaciones.

Varela Julia, 2005. **Sociología de la Educación. Algunos modelos críticos**. En http://www.ucm.es/info/eurotheo/diccionario/E/educacion_sociologia.htm.

Venegas Ma. Eugenia, 2004. **La evolución del concepto formación en la dimensión educativa costarricense durante la primera mitad del siglo XX.** Revista Actualidades Investigativas en Educación, Instituto de Investigación en Educación, Universidad de Costa Rica, Volumen 4, N.° 2. Costa Rica.

ANEXO

Cuadro A 3.1
COSTA RICA: Distribución de parejas
SEGÚN: Niveles de hipergamia e hipogamia
2004

Categorías		%		%	Nº casos	%
TOTAL					**6703**	**100,0**
Homogamia					3175	**47,4**
Heterogamia					3528	**52,6**
	Hipergamia		**Hipogamia**			
Cónyuge - jefe hogar	1713	**48,6**	1815	**51,4**		
1 nivel de diferencia	**1189**	**69,4**	**1273**	**70,1**		
Prim inc - prim comp	43,2		48,2			
Prim comp - sec inc	29,7		32,4			
Sec inc - sec comp	11,9		11,1			
Sec comp - univ	15,2		8,3			
2 niveles de diferencia	**402**	**23,5**	**372**	**20,5**		
Prim inc - sec inc	30,3		28,2			
Prim comp - sec comp	36,8		46,5			
Sec inc - univ	32,8		25,3			
3 niveles de diferencia	**108**	**6,3**	**140**	**7,7**		
Prim inc - sec comp	15,7		27,9			
Prim comp - univ	84,3		72,1			
4 niveles de diferencia	**14**	**0,8**	**30**	**1,7**		
Prim inc - univ	100,0		100,0			

Fuente: Elaboración propia con base en la Encuesta de Hogares de Propósitos Múltiples julio 2004, del Instituto de Estadística y Censos, Costa Rica.

CAPÍTULO 4

LA INCORPORACIÓN DE LA MUJER UNIVERSITARIA A LA POBLACIÓN ECONÓMICAMENTE ACTIVA

4.1 INTRODUCCIÓN

En las últimas décadas hemos asistido a cambios importantes en el mercado laboral y se destacan, entre otros, la masiva incorporación de las mujeres al mercado de trabajo y la consiguiente feminización del colectivo asalariado, por lo que se encuentra una progresión de la actividad laboral femenina. Pero la feminización del mercado, aunque real, es inacabada e incompleta porque ha tenido lugar bajo un fondo de desigualdad y de precariedad, en donde las mujeres con un nivel educativo universitario han sido protagonistas de dichas situaciones.

Los enormes avances en la formación de las mujeres y en su interés por participar de manera continuada en el mercado podían hacer pensar en una pronta desaparición de las tradicionales desigualdades entre mujeres y hombres en el mercado laboral; pero lo cierto es que no ha ocurrido así. Es decir, la espectacular transformación de la oferta laboral femenina no se ha traducido en una mejora equivalente del lugar que ocupan dentro del mercado. Es cierto que si se analizan uno a uno los diferentes indicadores de la desigualdad, se observa que la situación ha evolucionado, pero en los datos se siguen encontrando diferencias persistentes. Y estas desigualdades parecen hoy más injustas que ayer, precisamente por el enorme esfuerzo que han realizado las mujeres, con sus múltiples funciones dentro de nuestra sociedad: madre, esposa, universitaria y miembro activo de la economía del país.

Además podemos encontrar los problemas derivados de la relación educación – empleo que ocupa una atención creciente en las sociedades, constituyendo desafíos para la investigación, la planificación y la evaluación educativa. Desde este marco, el presente

trabajo aborda esa relación, pero circunscribiéndose al rol que la educación superior juega dentro de la población femenina. Somos conscientes de que centrándonos en el mercado de trabajo hacemos un análisis parcial, puesto que las relaciones entre trabajo familiar y trabajo de mercado siguen teniendo una gran importancia, sobre todo para las mujeres.

4.1.1 Justificación del tema

El fin esencial de realizar el presente estudio es conocer la importancia de algunos factores determinantes de las mujeres universitarias en la incorporación a la población económicamente activa (PEA) ya que en nuestro país, uno de los fenómenos más interesantes de la era moderna es el papel de las mujeres en el mercado de trabajo y su nivel de escolaridad, de manera que con la diversificación de los tipos de familias y el incremento en el nivel educativo de la población femenina, la relación de mujeres y empleo evoluciona. Es importante seguir indagando sobre este tema, por eso el estudio brindará información relevante que facilite el desarrollo de futuras investigaciones a profundidad y se pueda planificar proyectos económicos, sociales, culturales y políticos que brinden una igualdad de oportunidades con perspectiva de género, reforzando la función desempeñada por las mujeres en este sector y en su contribución al desarrollo social en general.

El énfasis de los distintos organismos mundiales es el de considerar a la mujer como protagonista importante para redefinir las políticas y para disminuir los niveles de pobreza, donde junto al hombre sean artífices del desarrollo de los países. A pesar de las desigualdades presentes entre los países las metas están dirigidas en general a incrementar la igualdad de derechos, la asignación equitativa de oportunidades y responsabilidades, en definitiva, un real estado de bienestar en las naciones.

Debemos educar a mujeres que sean conscientes de sus propias capacidades especiales femeninas y no se dejen llevar por una competencia falsa con los hombres, ni traten de imitar su comportamiento. Necesitamos urgentemente mujeres bien formadas en los campos de la ciencia, de la política, de la economía y de la educación para que el

mundo, que está dominado unilateralmente por los hombres, experimente una corrección necesaria. Una corrección que esté marcada por la empatía, la preocupación por la vida y por la familia, y no solamente por el lucro, la imagen y el éxito profesional. Nuestro país ha avanzado con respecto a este tema, pero todavía existen muchas oportunidades para seguir mejorando.

Es importante recalcar que hay que fortalecer a las mujeres mediante la educación, para que encuentren su propio camino, después de una consideración exacta de los valores, sin que se vean forzadas por imágenes falsas de la sociedad, a asumir un papel que no corresponde a su vocación específica.

Las mujeres trabajan porque son conscientes de las cualidades que poseen, y que suponen un beneficio para la sociedad, o bien, simplemente porque tienen que contribuir a la estabilidad económica de la familia. Las mujeres que se entregan conscientemente y con desprendimiento personal al servicio del trabajo y de la familia, deben ser protegidas por la sociedad. Esta doble "carga" amplía sus horizontes, y las sensibiliza frente a la situación actual de la sociedad. Estas experiencias y visiones pueden ayudar a cambiar las estructuras sociales de nuestro mercado laboral. Pero tenemos que insistir en que la prioridad es el bien de la familia y de los propios hijos. Es un error fatal pensar que se puede ser eficaz para el bien común, sin acoger y amar a los seres humanos que están en contacto directo con nosotras. Sólo en una cooperación comprensiva de los sexos se pueden superar los retos del futuro en un mundo globalizado. Solamente en el reconocimiento de la riqueza que Dios ha dado como Creador al hombre y a la mujer, se puede mejorar la situación de este mundo.

Los factores de índole social, familiar y cultural han sido los principales indicadores de discriminación hacia las mujeres en las universidades. Corresponderá a las mujeres y los hombres eliminar esas barreras para incursionar a un mundo laboral, académico, de investigación, etc. en condiciones distintas a las que se tiene en la actualidad. Ésta no es una tarea fácil, de ahí que tenemos que emprender de manera conjunta, autoridades,

personal académico, estudiantes, Estado y la sociedad en su conjunto las reformas necesarias para cambiar los escenarios docentes y de investigación.

Ante esta situación, considero que los estudios sobre la educación, en especial, los relacionados con el nivel superior, deben hacer visible el quehacer académico de las mujeres en la docencia, la investigación y en la toma de decisiones. De manera que se tengan al alcance estadísticas que permitan conocer la situación actual de la mujer; además se deberían diseñar políticas educativas internas en los centros de ecuación, que lleven a potenciar el papel de las mujeres, así mismo, poder contar con beneficios laborales que les permita desarrollarse en el mercado de trabajo y puedan cumplir con las responsabilidades dentro del hogar.

Es importante que se impulsen los estudios sobre las mujeres en las propias instituciones educativas: por facultades, departamentos, áreas, etc., así como también, que las universidades y autoridades permitan la inserción de cursos sobre estas temáticas, además que las instituciones destinen recursos para llevar a cabo este tipo de investigaciones al interior de las universidades. Todos estos elementos contribuirán sin lugar a dudas, a construir instituciones educativas a nivel superior más equitativas.

4.1.2 Problema de investigación y justificación de su importancia

Las múltiples funciones que se atribuyen tradicionalmente a la educación y a la formación, combinadas con el énfasis que dan nuestras sociedades modernas a los cambios constantes (políticos, económicos, ambientales, tecnológicos y sociales) exigen inevitablemente que el aprendizaje se convierta en una función permanente. En este sentido la importancia del aprendizaje a todo lo largo de la vida activa, pasará a ocupar cada vez más un lugar prioritario en los planes de los particulares, de los países y de la comunidad internacional, de manera más concreta en la actualidad, tratándose de un cuestión de “supervivencia”. [Gálvez, 2001]

Existen varias razones para estudiar la educación superior desde una perspectiva de género, teniendo en cuenta que la razón más conocida que se encuentra en nuestro país es el problema de un currículo oculto que perpetúa la segregación en la educación superior y en el mercado laboral, a excepción de la Universidad de Costa Rica que realiza estudios de equidad de género cada cierto tiempo para esta institución.

En los últimos años se ha resaltado la importancia de considerar como elemento del desarrollo de toda nación, la igualdad de oportunidades y participación del hombre y la mujer en los distintos sectores sociales. La educación es una de las principales áreas donde debe buscarse dicha equidad, pues constituye un instrumento importante que posibilita el acceso a otros recursos, mediante los cuales es posible lograr un mejor bienestar. Cualquier desigualdad en las oportunidades educativas limita la contribución que la ecuación tiene en el desarrollo del país.

El problema de investigación del presente estudio es conocer la importancia de algunos factores determinantes de las mujeres universitarias en la incorporación a la Población Económicamente Activa (PEA) del país. Se utiliza la información disponible en el archivo de datos de la Encuesta de Hogares de Propósitos Múltiple de julio 2004, la cual indica que la integración de la mujer universitaria al trabajo denominado "productivo" y, por consiguiente, a la población económicamente activa (PEA), es el resultado de relevantes transformaciones socioeconómicas y de la consolidación del modo de producción capitalista en Costa Rica, por lo que se busca una igualdad de condiciones para que dichas profesionales puedan tener las condiciones necesarias para formar parte del mercado laboral, minimizando, al punto que se desaparezcan, las diferencias y las razones que en ocasiones estas profesionales deben enfrentar y por las que se les dificulta ingresar a la PEA, a pesar de su alto nivel educativo. [Brennes, 2003]

Es importante recalcar que una cuarta parte de estas mujeres no forman parte de la PEA, por lo que nuestro trabajo se desarrollará en conocer ¿qué factores inciden y cuál es su importancia en la incorporación a la PEA de las mujeres con nivel universitario? El

supuesto teórico del que se partió y que posibilitó el análisis es que las relaciones entre la formación universitaria y el mercado de trabajo no pueden ser aisladas del modo de vida y de producción de una estructura social determinada. Por esta razón el aprovechamiento de los graduados por parte de los sectores productivos, se encuentra enmarcado en un complejo sistema de relaciones sociales, económicas, políticas e ideológicas que no pueden soslayarse si se pretende un abordaje integral del mercado de trabajo de los profesionales. Puede verse entonces que la igualdad de las mujeres en el acceso a la educación representa grandes beneficios para las familias y la sociedad en su conjunto.

4.1.3 Objetivos

Objetivo General

Identificar la importancia y los factores que inciden para que las mujeres con nivel universitario se incorporen a la PEA, empleando un modelo de regresión logística con la información recopilada en la Encuesta de Hogares de Propósitos Múltiple de julio 2004, de tal manera que se puedan comprender algunas de las razones del por qué existe un porcentaje representativo de mujeres universitarias que no se incorporan al mercado laboral.

Objetivos específicos

Considerando el objetivo general se propone los siguientes objetivos específicos:

- Analizar la situación actual de las mujeres costarricenses en la incorporación a la PEA y su nivel educativo.
- Caracterizar a las mujeres universitarias según las principales variables que aparecen en la Encuesta de Hogares de Propósitos Múltiples del año 2004.

- Identificar aquellos factores estadísticamente significativos, así como la magnitud y dirección de su influencia sobre la incorporación a la PEA de la mujer universitaria en Costa Rica.
- Utilizar el método de regresión logística como enfoque alternativo al clásico análisis descriptivo, para analizar el fenómeno de la mujer universitaria en la incorporación a la PEA.

4.1.4 Variables del análisis para ajustar el modelo y sus relaciones

En este apartado, se determinan las variables que van a ser escogidas para el estudio, las cuales serán descritas y presentada su codificación. Siempre debe considerarse la suficiencia del tamaño muestral para el número de covariables que se desea incluir en el modelo: modelos excesivamente grandes para muestras con tamaños muestrales relativamente pequeños implicarán errores estándares grandes o coeficientes estimados falsamente muy elevados (sobreajuste). En general se recomienda que por cada covariable se cuente con un mínimo de 10 individuos por cada evento de la variable dependiente con menor representación [Peduzzi, 1994].

Con el fin de alcanzar los objetivos propuestos en este estudio, fue necesario visitar el Instituto Nacional de Estadística y Censos (INEC), y se requirió el procesamiento y análisis de la Encuesta de Hogares y Propósitos Múltiples del año 2004. Consecutivamente se procedió a tomar los datos de las mujeres universitarias que se registraron en dicha encuesta para ese año en estudio, implementando el uso de los datos de todas las personas que cumplían con las características antes mencionadas. Esta información fue recolectada sin ningún contratiempo. Además de las variables que incluía la base de datos, se procedió a realizar nuevas variables que fueran de interés dentro del estudio. Se estableció que el análisis debería basarse sobre diecisiete variables, las mismas que se describen en los siguientes párrafos.

Los objetivos relacionados con la explicación de un fenómeno físico o social pueden lograrse recogiendo y analizando los datos. Así al realizar la investigación de algún fenómeno, se debe recoger observaciones de diferentes variables. El método por medio del cual se realiza el análisis de observaciones simultáneas sobre muchas variables es llamado Análisis Multivariado, a partir de este punto se utilizó una regresión logística.

El primer aspecto por considerar para la realización del estudio empírico se refiere a la selección de las variables que se deben incluir en el modelo por desarrollar, tanto la dependiente, que define el fenómeno cuyo comportamiento se trata de explicar, como las independientes o explicativas de dicho fenómeno. Con base en los aspectos teóricos de las técnicas aplicadas en el estudio, la variable dependiente o respuesta que modeliza el fenómeno que se analiza es dicotómica, siendo sus modalidades la condición económicamente activa e inactiva de la unidad en estudio, que se han codificado, respectivamente, con los valores uno y cero. Por lo que se refiere a las variables independientes o predictoras que permiten explicar el comportamiento de la variable dependiente, se han seleccionado teniendo en cuenta los factores que las mujeres universitarias consideran para medir la incorporación a la población económicamente activa.

Se ajustó un modelo de regresión logística multivariado, utilizando como variable de respuesta la condición económicamente activa de la mujer universitaria, la cual se transformó en dicotómica con el propósito de que existiera la ausencia o presencia en la pertenencia a la población económicamente activa (PEA). El motivo de utilizar esta variable es porque permite clasificar a la población entre económicamente activa y económicamente inactiva, según sea el tipo de actividad principal que realizan las personas en cuestión, y, para esta investigación en particular, nos enfocamos en las mujeres con nivel de instrucción universitaria.

En la construcción de la función Z para el modelo de regresión logística, se seleccionó un subconjunto de variables independientes que más información aportaba sobre

las probabilidades de pertenecer al grupo establecido por los valores de la variable. De manera se utilizó el método de selección "hacia atrás", utilizando los criterios basados en el estadístico P>z para la selección y eliminación de variables. Por lo tanto, para las variables incluidas de regresión logística, es decir para cualquier variable independiente X_j, seleccionada, si $ß_j$ es el parámetro asociado en la ecuación de regresión, el estadístico de P>z permite contrastar la hipótesis nula: H_o: $ß_j = 0$

La interpretación de dicha hipótesis es que la información que se perdería al eliminar la variable X_j, no es significativa. Si el p-valor asociado al estadístico de P>z es menor que α se rechazará la hipótesis nula al nivel de significación α. Bajo este punto de vista, en cada etapa del proceso de selección de variables, la candidata a ser eliminada será la que presente el máximo p-valor asociado al estadístico P>z.

En estadística, la regresión logística es un modelo de regresión para variables dependientes o de respuesta binomialmente distribuidas. Es útil para modelar la probabilidad de un evento que ocurre como función de otros factores. Es un modelo lineal generalizado que usa como función de enlace la función logit. La regresión logística es usada extensamente en las ciencias médicas y sociales. Otros nombres para regresión logística usados en varias áreas de aplicación incluyen modelo logístico, modelo logit, y clasificador de máxima entropía. El modelo de regresión logística puede escribirse como:

$$\log \frac{p}{1 - p} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$$

donde p es la probabilidad de que ocurra el evento de interés (en este caso es pertenecer a la PEA). Dado el valor de las variables independientes, podemos calcular directamente la estimación de la probabilidad de que ocurra el evento de interés de la siguiente forma:

$$\hat{p} = \frac{e^{suma}}{1 + e^{suma}}; \quad donde\ suma = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \ldots + \hat{b}_k x_k$$

Dentro de esta investigación se tiene que la Z en el modelo de regresión logística por utilizar está definida de la siguiente manera:

$$\mathbf{Z} = \beta_0 + \beta_1 * \text{zona} + \beta_2 * \text{aponse} + \beta_3 * \text{relpar} + \beta_4 * \text{edad} + \beta_5 * \text{anoes} + \beta_6 * \text{titul} + \beta_7 * \text{edcony} + \beta_8 * \text{nivicony} + \beta_9 * \text{anescony} + \beta_{10} * \text{titcony} + \beta_{11} * \text{actcony} + \beta_{12} * \text{acthog} + \beta_{13} * \text{hij} + \beta_{14} * \text{hijact} + \beta_{15} * \text{estciv} + \beta_{16} * \text{lningt} + \beta_{17} * \text{miemb} + \beta_{18} * \text{carte} + \beta_{19} * \text{cmedamb} + \beta_{20} * \text{ceconom} + \beta_{21} * \text{cfimaes} + \beta_{22} * \text{csalud} + \beta_{23} * \text{csocial} + + \beta_{24} * \text{cderec} + \beta_{25} * \text{cdocen} ++ \beta_{26} * \text{ingen}$$

El conjunto de variables independientes utilizadas se determinó con base en criterios derivados del estudio de algunos antecedentes citados en las referencias de este trabajo, así como a criterios e inquietudes propios, estas fueron consideradas preliminarmente para ajustar el modelo final por medio del programa estadístico STATA (Ver anexo d), de tal manera señalo a continuación las variables por estudiar y su operacionalización:

Cuadro 4.1

COSTA RICA: Operacionalización variables independientes del modelo inicial

Variables independientes		Descripción	Tipo	Códigos
VARIABLES PERSONALES	relpar	Relación de parentesco que tienen las mujeres universitarias con el Jefe de Hogar.	Cualitativa	1 = Jefe 0 = No Jefe
	edad	Edad en años cumplidos de las mujeres universitarias en estudio.	Cuantitativa	
	estciv	Estado civil de las mujeres universitarias	Cualitativa	1 = Con cónyuge 0 = Sin cónyuge
VARIABLE ACADÉMICA	anoes	Años de escolaridad de las mujeres universitarias en estudio.	Cuantitativa	
	titul	Título universitario de las mujeres universitarias en estudio	Cualitativa	1 = Con título 0 = Sin título
	carte	Carrera en Artes de las mujeres universitarias en estudio	Cualitativa	1 = Con carrera 0 = Sin carrera
	cmedamb	Carrera en Medio Ambiente de las mujeres universitarias en estudio	Cualitativa	1 = Con carrera 0 = Sin carrera
	ceconom	Carrera en Ciencias Económicas de las mujeres universitarias en estudio	Cualitativa	1 = Con carrera 0 = Sin carrera
	cfimaes	Carrera en Física, Matemáticas y Estadística	Cualitativa	1 = Con carrera 0 = Sin carrera
	casalud	Carrera en Ciencias de la Saludo de las mujeres universitarias en estudio	Cualitativa	1 = Con carrera 0 = Sin carrera
	csocial	Carrera en Ciencias Sociales de las mujeres universitarias en estudio	Cualitativa	1 = Con carrera 0 = Sin carrera
	cderc	Carrera en Derecho de las mujeres universitarias en estudio	Cualitativa	1 = Con carrera 0 = Sin carrera
	cdocen	Carrera en Docencia de las mujeres universitarias en estudio	Cualitativa	1 = Con carrera 0 = Sin carrera
	cingen	Carrera en Ingenierías de las mujeres universitarias en estudio	Cualitativa	1 = Con carrera 0 = Sin carrera

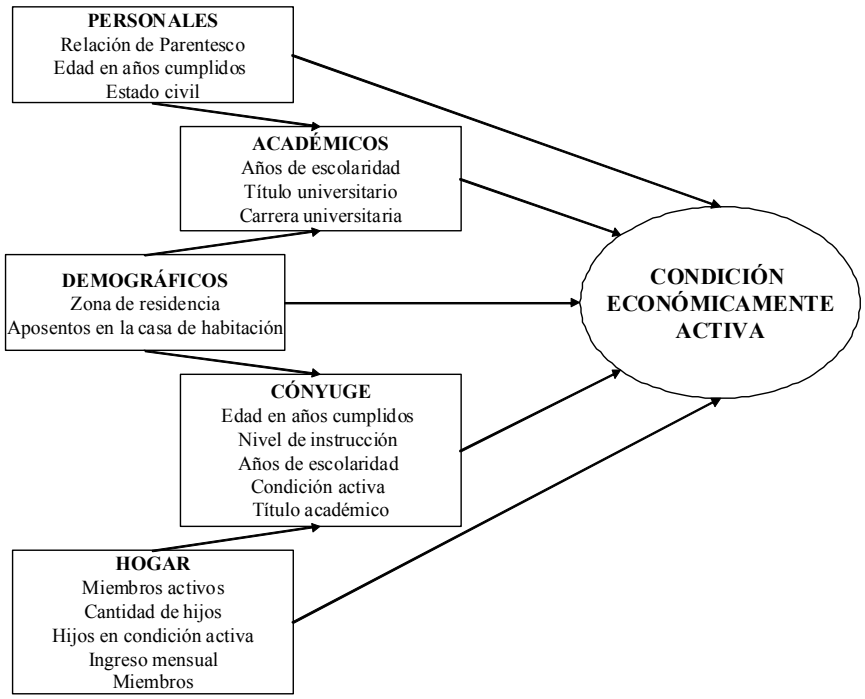
Variables independientes		Descripción	Tipo	Códigos
VARIABLES DEMOGRÁFICAS	zona	Zona de residencia de las mujeres universitarias en estudio	Cualitativa	1 = Urbana 0 = Rural
	aponse	Aposentos en la casa de habitación de las mujeres universitarias en estudio	Cuantitativa	
VARIABLES DEL HOGAR	acthog	Cantidad de miembros del hogar en condición económicamente activa	Cuantitativa	
	hij	Cantidad de hijos de las mujeres universitarias en estudio	Cuantitativa	
	hijact	Cantidad de hijos en condición económicamente activa de las mujeres universitarias en estudio	Cuantitativa	
	lningt	Ingreso mensual total del hogar	Cuantitativa	
	miemb	Cantidad de miembros en el hogar de las mujeres univesitarias	Cuantitativa	
VARIABLES CONYUGE	edcony	Edad en años cumplidos del cónyuge de las mujeres univesitarias	Cuantitativa	
	nivicony	Nivel de instrucción del cónyuge de las mujeres univesitarias	Cualitativa	1 = Universitario 0 = No universitario
	anescony	Años de escolaridad del cónyuge de las mujeres univesitarias	Cuantitativa	
	actcony	Condición económicamente activa del cónyuge de las mujeres	Cualitativa	1 = Activa 0 = No activa
	titulcony	Título universitario del cónyuge de las mujeres universitarias	Cualitativa	1 = Sin título 0 = Con título

Muchas de las variables de las cuales se ha obtenido la información son de tipo cualitativo. Para poder realizar el análisis estadístico de ellas se tuvo que codificar mediante criterios específicos que cumplieran con las necesidades del estudio y posteriormente para el análisis de regresión logística. Mediante el esquema anteriormente citado sobre la codificación, los coeficientes de las nuevas variables reflejarán el efecto de las categorías representadas respecto al efecto de la categoría de referencia.

La condición económicamente activa de la mujer universitaria se ve relacionada con factores explicativos que se ilustran en el Diagrama 4.1a, el cual está formado por cinco áreas de interés: aspectos personales, académicos, demográficos, del cónyuge y del hogar, y estos entre sí, están interrelacionados y se indican en el esquema por las flechas que los unen.

Diagrama 4.1a

Modelo teórico de la interrelación entre los factores asociados con la condición económicamente activa de la mujer universitaria



Existen situaciones personales que afectan a las mujeres universitarias en la incorporación a la PEA, como es el hecho de que sean las Jefas de Hogar. La probabilidad de que sean las proveedoras en sus hogares es muy alta: son responsables de su hogar y buscan el trabajo para poder brindar lo necesario para el resto de la familia, a esto se suma también el hecho de tener cierta edad, por lo cual su tiempo para concebir hijos ha finalizado y esto le permite poder continuar con sus labores. A su vez, al tener mayores funciones y responsabilidades dentro de su hogar, busca un mejor nivel económico que está relacionado con la búsqueda de un mejor nivel de instrucción, con el fin de alcanzar mejores puestos labores que le sean bien retribuidos económicamente, de manera que la cantidad de años de escolaridad estará estrechamente relacionada con el título universitario según el grado y la carrera que haya estudiado.

La realidad demográfica, económica y social hace que la PEA cambie de acuerdo con estas circunstancias y necesidades de cada país. La tendencia ha sido de un incremento continuo de mujeres con inobjetables logros académicos que refleja su incorporación a la Población Económicamente Activa y la gradual modificación de pautas culturales y sociales. Es claro que los aspectos demográficos van relacionados con la situación del cónyuge. Existen zonas residenciales en las cuales la probabilidad de que las mujeres convivan con un hombre es alta, y estas a su vez afectan la condición económicamente activa de dicha población femenina ya que, dependiendo de estas características, ellas no se ven en la necesidad de trabajar, o viceversa.

Con relación a la división del trabajo por género, es importante señalar que tal y como se da en las sociedades patriarcales, provoca que especialmente las mujeres realicen una doble o triple jornada de trabajo. Esto quiere decir que aunque están incorporadas al mercado laboral, social y culturalmente, la sociedad les asigna las responsabilidades reproductivas, de ahí que las características de su cónyuge juegan un papel muy importante y por ende las de su hogar, como la cantidad de hijos y miembros dentro de éste.

La estructura por edad y sexo de la población y su dinámica brindan la posibilidad de visualizar cuales son las necesidades de la población. De ahí la importancia de la relación del nivel de instrucción y su condición económica, en busca de una mejor calidad de vida. Una característica del cónyuge, como los años de escolaridad, afecta si la mujer se incorpora a la PEA o no; ya que si este tuviera un grado universitario completo, tiene la ventaja de que puede aspirar a mejores trabajos y por ende a un mejor salario, de modo que en sus hogares la necesidad de que otra persona tenga un ingreso mensual es menor versus una persona que tenga menos nivel de instrucción. Por ello la situación del cónyuge se relaciona a su vez con los aspectos del hogar ya que, en ocasiones, de esto depende la cantidad de hijos y el ingreso mensual total.

4.2 ESTADO DE LA SITUACIÓN

La incorporación de la mujer al mercado laboral se puede considerar uno de los grandes logros del pasado siglo. No sólo dignifica el papel de las mujeres, sino que demuestra que en este campo son iguales a los hombres. Las mujeres han tenido la oportunidad de avanzar en el terreno profesional y han demostrado evidentes capacidades, como por ejemplo la entrega, la capacidad de organización o humanización de la empresa, estos factores son algunos de los elementos positivos que la mujer ha introducido con su trabajo.

Poco a poco, se van rompiendo los techos que les impiden acceder a puestos de responsabilidad, y hoy ya no suena extraño que una mujer opte a la Presidencia de un Gobierno, dirija una gran empresa multinacional o porte sobre su hombro un fusil de asalto. Los resultados electorales del 2005 constituyen un claro ejemplo de la situación expuesta anteriormente, pues la Asamblea Legislativa está integrada por el mayor número de diputadas en la historia patria, 22 diputadas que hoy integran el Congreso, es decir, aproximadamente un 39% de total de los miembros. [Departamento de Relaciones Públicas, Prensa y Protocolo, 2006]

Sin embargo, esta llegada masiva de la mujer a los puestos de trabajo, desde mediados de los años 80, no ha ido acompañada de los cambios necesarios en la estructura de la sociedad para que quedaran cubiertos los papeles que antes correspondían exclusivamente a las madres. Las labores del hogar, el cuidado y la educación de los niños, la atención a los ancianos, los enfermos y los discapacitados, o la gestión de la economía doméstica, son tareas encomendadas, durante siglos, a las mujeres de la casa. El hombre tiene un papel muy delimitado como proveedor. A él le corresponde ganarse el “pan con el sudor de su frente”, y esta labor parece dejarle exento de cualquier otra tarea, de tal manera que el trabajo desempeñado por las mujeres ha tenido una importancia vital, aunque su

contribución a la economía ha variado dependiendo de la estructura, las necesidades, las costumbres y los valores sociales. A pesar de todos los esfuerzos todavía siguen existiendo diferencias significativas entre géneros y se pretenden eliminar con el paso del tiempo. [Quirós, 1984: 51]

Esto ha hecho que en los países industrializados se estén produciendo una serie de cambios tímidos aún, pese a que incluyen una mayor proporción de mano de obra femenina en la fuerza de trabajo; una disminución de las cargas familiares (debido al menor tamaño familiar y a los avances tecnológicos que facilitan las tareas domésticas); mayor nivel cultural de las mujeres y un mayor nivel salarial, así como tareas de mayor responsabilidad para las mujeres, que se emplean por motivos económicos y personales.

La condición de desigualdad que enfrentan las mujeres en todo el mundo, respecto de los hombres, se ve reflejada en los distintos espacios de la vida social y ha sido motivo de organización y de lucha para algunas mujeres y de preocupación y atención para distintos sectores de la población, principalmente de aquellos que se posicionan en contra de la discriminación y en el respeto de todos los derechos como elementos fundadores de una sociedad justa y democrática. [Chinchilla, Lápiz y Segura, 2006]

Un aspecto que es interesante destacar es que las mujeres, en menor proporción que los hombres, pertenecen a la PEA. De tal manera se refleja esta situación que aún con mayor educación no todas acceden y/o quizá no siempre aspiran a entrar a la PEA, es decir el 91.8% de los hombres universitarios tienen condición económicamente activa, mientras que las mujeres, bajo estas mismas condiciones, sólo representan un 74.6%. Las mujeres que tienen niveles educativos de primaria y secundaria son las que tienen mayor participación del total de "condición económicamente activa" con un 37.0% y 34.1% respectivamente. Con esto se quiere mostrar que la necesidad de trabajar presiona para aceptar todo tipo de empleo e incorporarse a la PEA, a pesar de las serias dificultades que ha habido en los últimos años para la obtención de empleos. [INEC, 2000]

El Instituto de Estudios de la Mujer de la Universidad Nacional se inserta en un proceso latinoamericano, que se inicia a partir del interés de académicas de distintas disciplinas y se formaliza con el acuerdo de la reunión regional de la UNESCO de 1980, para impulsar la creación de instancias de estudio sobre la mujer y de género, que contribuyan en los procesos hacia la equidad de género entre mujeres y hombres, para la transformación misma de las universidades.

Además se han desarrollado diversas luchas sociales que han permitido la conquista de nuevas posiciones por parte de la mujer en diversos ámbitos de la práctica social. Muchas de esas luchas han terminado por plasmarse en acuerdos, convenios y programas internacionales, como por ejemplo: "Programa de la Agencia Latinoamericana de información sobre las mujeres, género y comunicación en América Latina y el Caribe" CEPAL, "Conferencia Regional sobre la mujer de América Latina y el Caribe", leyes nacionales, por ejemplo, "La ley de promoción de la igualdad social de la mujer", programas y proyectos, los cuales han tenido como objetivo la generación de condiciones para favorecer la participación de la mujer en espacios de importancia estratégica, entre los cuales están la Ciencia y Tecnología.

4.2.1 Las leyes de promoción de la igualdad social de la mujer

Son leyes que tutelan los derechos de las mujeres, y responden verdaderamente a los fundamentos de las legislaciones antidiscriminatorias que contemplan los derechos de las mujeres en otros ámbitos (más allá que su función reproductiva o familiar), tales como los relacionados con la economía, la política, la cultura, etc.

En Costa Rica se cuenta con algunas leyes de acciones afirmativas y de tutela exclusivamente para las mujeres:

La ley de promoción de la igualdad social de la mujer" más conocida por el nombre que llevaba el proyecto de "igualdad real de la mujer", presentada el 8 de marzo de 1988.

La discusión de esta ley constituyó uno de los debates nacionales de mayor trascendencia que se han dado en el país en torno a la desigualdad y discriminación de las mujeres. Esta ley es el primer paso importante, legislativa y jurídicamente, hacia la ejecución y cumplimiento de la CEDAW ("Convención sobre la eliminación de todas las formas de discriminación contra la mujer") en Costa Rica, para pasar de la igualdad formal a la igualdad real. [Arroyo, 1999: 6-9]

La Ley de Igualdad contempla un conjunto de medidas para promover la equivalencia real entre hombres y mujeres en todos los ámbitos de la vida. Propone acciones para favorecer el acceso de la mujer al empleo, facilitar su carrera profesional y la conciliación de la vida laboral y familiar; mejora las prestaciones por maternidad y reconoce un permiso de paternidad de ocho días, independiente del que corresponde a la madre. Además, promueve una composición equilibrada entre hombres y mujeres en las listas electorales y en los consejos de administración de las empresas. No es una ley para las mujeres, sino para los hombres y para las mujeres, "porque a todos dignifica", y "contribuye decisivamente a hacer una sociedad más justa y equilibrada". [Arroyo, 1999: 6-9]

4.2.2 El sistema de educación superior

Antes de 1973, la única institución de educación superior existente era la Universidad de Costa Rica. Según los datos disponibles, esta institución atendió entre 1955 y 1970 una población estudiantil compuesta por alrededor de un 63% de hombres y 37% de mujeres. El acelerado crecimiento de la población estudiantil que se dio durante ese periodo fue en términos relativos similar en la población femenina y la masculina, manteniéndose así la composición por sexo. [UCR, 1960-1961, 1970)

En 1973 iniciaron lecciones la Universidad Nacional y el Instituto Tecnológico de Costa Rica, quienes junto con la Universidad de Costa Rica atendieron ese año una población de 30,125 estudiantes, de los cuales un 41,2% eran mujeres, incrementándose así,

la participación de éstas. [Brennes, 2003] Ya en 1976 inició labores la Universidad Autónoma de Centro América, la única institución universitaria privada en esa fecha, y en 1977, la Universidad Estatal a Distancia. Para 1979, se tienen registros de matrícula, que no incluyen esta última institución, los cuales muestran una participación femenina aún mayor: 46,7% (47,8% en las estatales y 31,1% en las privadas).

Conforme se crearon más instituciones, se fue haciendo más difícil tener una medición completa de la matrícula universitaria del país y aún más su composición por género. La Oficina de Planificación de la Educación Superior (OPES) ha hecho esfuerzos por conocer la participación de hombres y mujeres en la educación superior universitaria, pero al no tener injerencia sobre las universidades privadas, sus logros se limitan al sector estatal. Esta es una limitación importante dado el crecimiento de la educación superior privada, la cual pasa de representar un 22% de la matrícula universitaria total en 1987 a quizás más del 50% en los últimos años. [Brennes, 2003]

Con el fin de llenar este vacío de información, realicé algunas estimaciones con base en los datos censales, y se observa cómo en el último Censo del año 2000 las universitarias tienen mayor participación con respecto al total, siendo un 51.6% del conjunto de universitarios, comportamiento que varió según los años anteriores, en donde los hombres eran los que asumían el mayor volumen. (Ver Anexo Cuadro A 4.2)

Para 1990 se cuenta con información de matrícula por sexo en las Universidades Estatales y parte de las privadas (56% de la matrícula de éstas). De acuerdo con los datos, en el sector estatal en ese año se observa una participación de géneros bastante equitativa (49,4% de mujeres); en el sector privado dicha equidad se observa en 1994. Podría decirse entonces que es en la primera mitad de la década de los noventa cuando las mujeres logran igualarse a los hombres en cuanto a participación en la educación superior, y que a partir de ahí empiezan a superarlos pero en forma moderada.

En los últimos años se han realizado algunas investigaciones sobre las condiciones que propician las desigualdades de género en la Educación Superior del país, y señalan lo siguiente: [Proyecto Estado de la Nación, 2002]

- En la sociedad costarricense se sigue reproduciendo la tradicional asignación de roles de sociedades patriarcales, en donde los hombres son responsables de la vida pública y las decisiones, mientras que las mujeres tienen a su cargo la reproducción, la educación y el cuido de la familia.
- Las prácticas sexistas en el aula (procesos que limitan el desarrollo de las potencialidades integrales debido al sexo de cada persona y, por ende, de su grupo de iguales) tienden a reforzar estructuras educativas discriminatorias y sus efectos sobre las mujeres.
- Los contextos familiares y los mensajes de los medios de comunicación contribuyen a reproducir el régimen de dominación de género e inciden en los procesos que conducen a la elección de carreras, profesiones u oficios.
- Existe segmentación por sexo en el mercado laboral costarricense al presentarse ocupaciones típicamente femeninas y masculinas. Según el indicador de segregación por grupos ocupacionales, casi todas las mujeres que trabajan deberían cambiar de ocupación, para que se dé una estructura ocupacional paritaria.
- Los varones eligen carreras universitarias por indicación directa de los padres de forma que les garantice un futuro económico más favorable, situación que es consustancial para las mujeres.
- La mujer tradicionalmente ha elegido carreras relacionadas con las funciones hogareñas, que tienen poco reconocimiento salarial. La presencia de éstas es minoritaria en carreras que enfatizan el uso de la matemática y en carreras de las ciencias experimentales y tecnológicas.
- Existe una demarcación de carreras que conlleva a una "feminización" o "masculinización" de las mismas. Es mal visto que un varón acceda a una carrera supuestamente orientada para las mujeres.

4.2.3 Visión internacional

Es importante recalcar el proceso y el rol que posee la población femenina en la actividad económica de otros países desarrollados ya que, según los datos de la Organización Internacional del Trabajo, a medida que los países se van industrializando, las mujeres mejoran su categoría profesional.

El empleo de mujeres en países industrializados como Europa, Estados Unidos y Japón es muy similar. Antes de 1990 la participación de la mujer en Alemania Occidental (ahora parte de la reunificada República Federal de Alemania), era del 38%, y del 55% en Suecia. En España el porcentaje es mucho menor, debido a la tardía incorporación de la mujer al mercado laboral. En casi todos los países industrializados existe una legislación relativa a la igualdad de oportunidades y a la protección de la mujer en el trabajo. La negociación colectiva se utiliza con más frecuencia en Europa que en Estados Unidos para mejorar las condiciones laborales de las mujeres. [Mujica, 2006]

Las políticas de empleo en la Europa del Este y en los países de la Unión de Repúblicas Socialistas Soviéticas (URSS) con regímenes comunistas partían de la creencia de que la mujer tenía tanto el derecho como el deber de trabajar. En 1936 la Constitución soviética señalaba que no se podía legislar en contra de la igualdad de la mujer. La URSS y sus aliados promulgaron leyes a favor de la protección de menores, la educación, la salud y las actividades lúdicas. [Mujica, 2006]

Estimaciones en las décadas de 1970 y principios de 1980 señalan que el 85% de las mujeres soviéticas entre 20 y 55 años trabajaban fuera de casa; en la Alemania del Este el número de mujeres asalariadas superaba el 80%. Aunque participaban más en el mercado laboral que las mujeres de Occidente, las trabajadoras de Europa del Este también se ocupaban de tareas que requerían poca calificación y casi siempre en cargos de menor responsabilidad que la de los hombres. Por ejemplo, en Bulgaria 78% de los trabajadores del sector textil eran mujeres, pero sólo 25% contaban con la categoría de ingenieros; para

la Unión Soviética (U.S.) estas cifras eran 74% y 40% respectivamente. Aunque en la U.S. no se fomentaba el empleo a tiempo parcial, 50% de las mujeres casadas trabajaban sólo una parte de la jornada. [Mazzei, 2006]

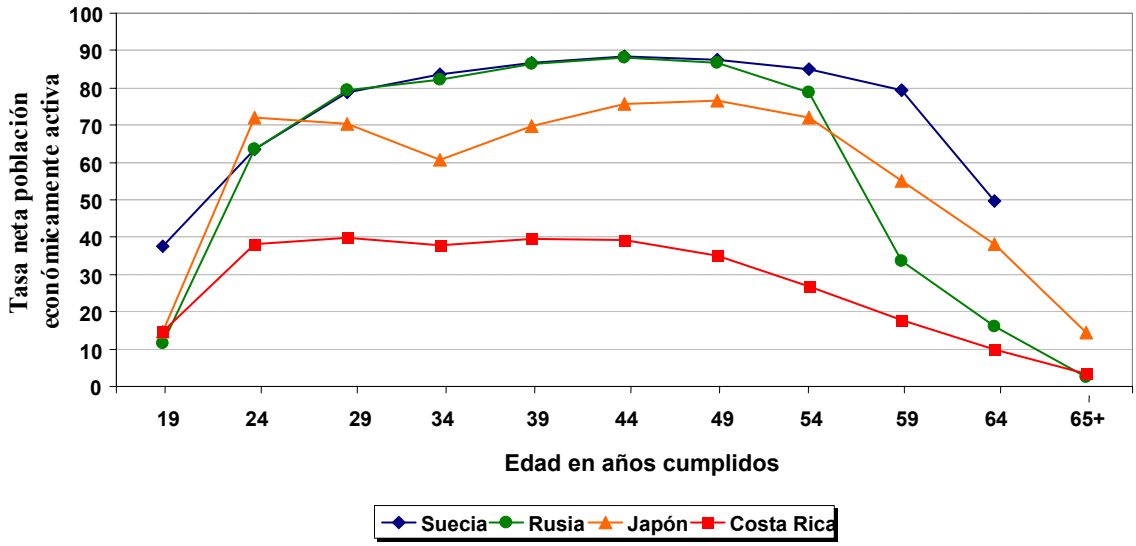
De esta manera, en países desarrollados para el año 2001, la tasa neta de la población económicamente activa es de 55.7% a diferencia de 64.2% en países menos adelantados, como los que se encuentran en América Latina y el Caribe que representan un 42.2%. En cuanto a hombres se refiere, esta región tiene una de las tasas netas de la población económicamente activa más bajas del mundo con un 52%. (Ver Anexo Cuadro A 4.3)

Los estados comunistas defendían que debía pagarse igual salario a igual trabajo, pero eran pocas las mujeres que alcanzaban lo más alto del escalafón. Sin embargo, la exactitud de estas cifras ha sido puesta en duda tras la caída de los regímenes comunistas en Europa y Euro Asia, aunque quizá sea cierto que las mujeres de estos países disfrutaban de una mayor igualdad salarial y un número superior de oportunidades que las mujeres occidentales. Sin embargo, es interesante observar la evolución de la situación cuando desaparezcan las industrias estatales y los sistemas de seguridad social en los países de Europa central y del Este. [Mazzei, 2006]

Cabe destacar que la curva por edades de la tasa neta de población femenina, económicamente activa en Costa Rica, presenta la misma tendencia que los países industrializados, claro está que los porcentajes son mucho menores. (Ver Gráfico 4.1)

Gráfico 4.1

MUNDIAL: Tasa neta de la población femenina económicamente activa en 2000



Fuente: Estadísticas de Trabajo (LABORSTA), (BA) Registro de fuerza laboral tomado de la base de datos.

Entre los países occidentales, Suecia es el único país que ha logrado una mayor igualdad laboral para las mujeres. Durante las dos últimas décadas los ingresos salariales medios de las mujeres han pasado del 66% al 87% de los ingresos de los hombres. Al mismo tiempo, el Gobierno sueco emprendió una reforma de los libros de texto, la educación de los padres, la protección de menores, las políticas de impuestos y la legislación relativa al matrimonio y al divorcio para fomentar la igualdad de la mujer en el mercado laboral, al tiempo que se reconocen las necesidades específicas de las madres trabajadoras. Se crearon programas de ayuda y asesoría para las mujeres que se reincorporaban, tras un periodo de maternidad, al mercado de trabajo.

Otros países europeos han analizado el modelo sueco, y algunos están adaptando los programas de ayuda a sus políticas de bienestar, aunque el costo económico de la

adaptación del sistema de bienestar sueco es un obstáculo importante para conseguir estos cambios. Japón, el país más industrializado de Oriente, conserva algunas de sus tradiciones hacia la mujer. La participación de las mujeres en el mercado laboral es algo menor que en los países occidentales, pero la mujer suele dejar su trabajo de forma concluyente cuando tiene hijos, a pesar de que el alto porcentaje de educación en Japón hace que exista un elevado número de mujeres con estudios superiores. Se ha creado una legislación relativa a la igualdad de oportunidades para garantizar y fomentar el empleo de las mujeres en tareas diferentes a las reservadas; según la tradición a la mano de obra femenina (empleos mal pagados, entre los que predominan las tareas de secretaría y administración) pero el promedio de mujeres que desempeñan altos cargos, tanto en el sector privado como en la administración pública, sigue siendo menor que el registrado en algunos países occidentales.

En Corea del Sur, Singapur y Taiwan, y en otras economías jóvenes en el plano industrial del Sureste asiático, se han creado (gracias al desarrollo económico) nuevas oportunidades laborales para las mujeres. En Corea del Sur la presencia de la mujer en el mundo laboral es más reducida que en Japón; en los demás países de esta zona la presencia femenina es aún menor. Las actitudes paternalistas tradicionales, la importancia de la familia en las diferentes religiones confucionistas y el predominio del Islam[45] en algunas zonas tienden a disminuir el estatus y la presencia de la mujer en el mercado laboral. Sin embargo, el crecimiento económico ha permitido que las mujeres puedan desempeñar cargos y cobrar salarios que nunca antes habrían podido imaginar. Además, estos países son los primeros interesados en impedir que las limitaciones de la tradición reduzcan su potencial creación de riqueza. Con esto se reafirma que no necesariamente la industrialización ha mejorado la situación laboral. En estos países ha existido interés por las mujeres pese a los roles fuertemente tradicionales, ya que a ellas se les remunera con salarios más bajos.

---

[45]/ Islam: es una religión monoteísta basada en el Corán, libro sagrado, que creen que fue dictado por Dios (árabe: Allāh) a Muhammad (español: Mahoma) a través del arcángel Gabriel.

Tomando en cuenta países con características similares a las de nosotros en cuanto a política, economía y cultura se refiere, podemos observar que para el 2004 Costa Rica tiene 28% de la población femenina económicamente activa, a diferencia de otros países centroamericanos (CAM) con porcentajes menores, como lo son Guatemala y Honduras con un 25% y 26% respectivamente, mientras que el desarrollo en países como El Salvador con un 32% y Nicaragua con 34% mantienen más altos estos valores. Claro está que se debería de hacer un análisis más profundo sobre los factores que influyen, ya que somos el tercero en CAM en cuanto a esta variable, pero tenemos una mejor economía al posicionarnos en un 2do lugar con un 26%. [FLACSO, 2002: 47-51]

4.2.4 Situación nacional

4.2.4.1 El entorno socioeconómico

Costa Rica muestra una distribución de la población por sexo bastante equilibrada. El Censo de Población efectuado en el año 2000 reflejó porcentajes de 49.9% y 50.1% para hombres y mujeres respectivamente. Este equilibrio se mantiene en los diferentes grupos de edad. La esperanza de vida de la mujer costarricense es más elevada que la de los hombres, mostrando una diferencia de entre cuatro y cinco años en el periodo 1991-2002. Para este último año el valor fue de 80.2. [INEC, 2000]

Los censos de población de 1973, 1984 y 2000 muestran una distribución según nivel educativo de la población femenina. En la cual se ha venido incrementando los porcentajes en niveles desde primaria completa hasta educación superior. Del total de la población femenina en 1973, el 79,75% correspondía a un nivel de primaria, mientras que en el Censo del 2000 se puede observar que este nivel de educación ya no tiene tanta participación con en años anteriores siendo un 59,35% del total, en la cual las mujeres universitarias han alcanzado mayores porcentajes, y para ese año está representando un

12.89% del total, lo que indica que año tras año son más las mujeres a nivel superior. (Ver Cuadro 4.2)

Cuadro 4.2
COSTA RICA: Frecuencia relativa de la población femenina mayor de 25 años
SEGÚN: Nivel de educación
1973, 1984 y 2000.

Nivel de educación	Femenino		
	1973	1984	2000
Total	**100,00%**	**100,00%**	**100,00%**
Primaria	79,75%	66,37%	59,35%
Secundaria	16,15%	26,16%	27,76%
Univers / ParaUniver	4,09%	7,47%	12,89%

FUENTE: Instituto Nacional de Estadística y Censos. Censos 1973, 1984 y 2000.

La distribución de la fuerza de trabajo muestra una concentración de mujeres en los servicios sociales y personales, en el comercio y en las industrias manufactureras, durante el periodo 1993-1998. El porcentaje de mujeres en estas categorías fue de alrededor de 45%, 23% y 20%, respectivamente. Los hombres se ubican principalmente en "agricultura, caza, silvicultura y pesca" aunque el porcentaje en esta rama de actividad ha presentado una disminución; también en comercio, servicios sociales y personales e industrias manufactureras, pero en menor proporción que las mujeres.

La tendencia de los hogares con jefatura femenina está en aumento, tanto en la zona urbana como en la zona rural. Por otra parte, estos hogares son más afectados por la pobreza que aquellos cuyo jefe es un hombre. Esta brecha de pobreza por sexo de la jefatura de hogar se ha profundizado en los últimos años. Es importante también destacar

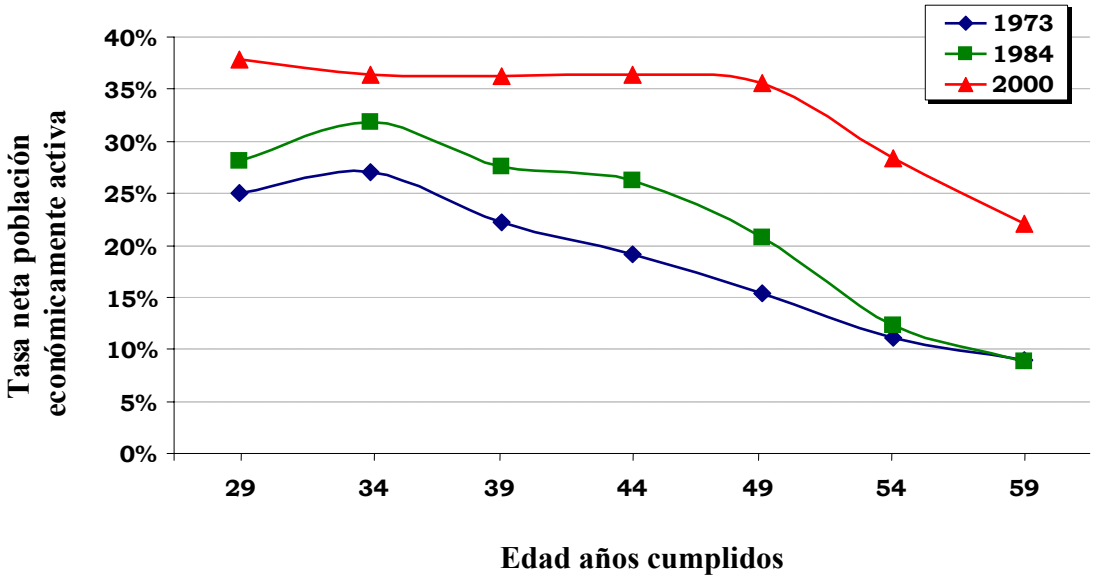
que conforme pasa el tiempo, son más mujeres las que tienen un nivel de educación universitaria, siendo en 1973 el 48.3% del total de la población universitaria, mientras que el Censo del 2000 muestra que las universitarias alcanzan un mayor volumen que los hombres, representando el 51.6% del total. [INEC, 2000]

La mayor parte de los universitarios, sin distinción de sexo, logran incorporarse al mercado laboral. No obstante, hay diferencias en el porcentaje de hombres y de mujeres que trabajan una vez que concluyen su carrera universitaria. Hoy en día, cada vez más gente logra alcanzar niveles educativos más altos; de modo que la oferta de profesionales ha crecido mientras la demanda se estanca o retrae, lo que indica que no es la falta de escolaridad lo que limita las posibilidades de trabajar, sino que la economía no genera la suficiente cantidad de empleos, aún para la franja de los más capacitados que son los que poseen mayor empleabilidad. Esta situación se refleja en la cantidad de universitarios que, según el Censo del 2000, creció aproximadamente en un 183% con respecto al Censo de 1984, donde las mujeres alcanzaron un mayor porcentaje que los hombres, es decir un 205% y 163% de aumento respectivamente.

Con el pasar de los años se observa cómo la participación neta de la mujer universitaria incorporada en la PEA está creciendo con respecto al total de mujeres activas, donde uno de los puntos máximos son las mujeres de 40 a 44 años que corresponde a un 36.4% para el Censo del 2000, mientras que para 1984 y 1973 es 26.2% y 19.2% respectivamente. (Ver Gráfico 4.2)

Gráfico 4.2

COSTA RICA: Distribución relativa de las mujeres universitarias económicamente activas con respecto al total de mujeres económicamente activas. 1973, 1984 y 2000.



Fuente: Instituto Nacional de Estadística y Censos, Censos 1973, 1984 y 2000.

4.2.4.2 Situación actual de la PEA

Junto con los cambios ocurridos en Costa Rica en materia económica y demográfica durante las últimas décadas, se han gestado importantes transformaciones en la sociedad, como por ejemplo, los cambios demográficos han corrido a la par de una más amplia participación de las mujeres en la vida social, política y económica, y particularmente en el empleo.

El nivel educativo de la población femenina y masculina censada es similar; en ambos sexos se han venido incrementando los porcentajes en niveles desde primaria completa hasta educación superior. Para el Censo 2000 se muestra que la mayor participación a nivel superior y secundaria completa corresponde al sexo femenino con un

15.8% (hombres -0.7% de diferencia) y 10.1% (hombres -1.3% de diferencia) respectivamente. (Ver Anexo Cuadro A 4.4)

La tasa neta total de participación de las mujeres en la fuerza de trabajo se incrementó de 18,6% en 1973 a 27,0% en el 2000; la de los hombres, por el contrario disminuyó durante el periodo. Sin embargo, la participación de las mujeres se ve afectada por el desempleo, que es mayor en esta población que en la masculina. De tal manera la rama ha caído en general, pero las mujeres han incrementado su participación.

La distribución de la fuerza de trabajo por rama de actividad muestra una concentración de mujeres en los servicios sociales y personales, en el comercio y en las industrias manufactureras; de 1973 al 2000 disminuyó el porcentaje de éstas en la primera categoría, pero continúa siendo bastante alto. Los varones se ubican principalmente en "agricultura, caza, silvicultura y pesca", aunque también el porcentaje de éstos en dicha rama de actividad ha venido disminuyendo; además, en comercio, servicios sociales y personales e industrias manufactureras. La mayor intervención de las mujeres en la vida económica ha respondido a varios factores y momentos en los últimos 30 años, mientras que la de los hombres se mantiene estable a lo largo de los años. [FLACSO, 2000]

Los miembros de los grupos familiares más pobres tuvieron que trabajar más en contextos laborales precarios y con salarios cada vez más bajos. Sobrevivir se convirtió en una tarea que requería estrategias de intensificación del trabajo de los miembros de las familias. Las mujeres adultas y los varones jóvenes aumentaron su presencia en los mercados laborales, a cambio de salarios empobrecidos pero de creciente importancia para las economías doméstico-familiares.

Fue la década del aumento masivo de las mujeres en el empleo y de obstáculos crecientes para la permanencia de los niños en las escuelas. Sin duda, la familia actuó como un amortiguador, pero los costos fueron altos: los patrones de consumo cambiaron, las dietas se deterioraron, la gente tuvo que trabajar más pero comía peor y menos, muchos

niños tuvieron que dejar la escuela y las relaciones de género, según muchos estudios, sufrieron mayores conflictos y roces violentos [FLACSO, 2000].

En los últimos 30 años la tasa neta de la población femenina en la actividad económica ha crecido a más del doble, según datos censales del año 1973 al 2000 hubo un crecimiento del 244%, mientras que del 1984 al 2000 fue de 119%. (Ver Anexo A 4.1). La participación económicamente activa de las mujeres crece fuertemente en edades jóvenes, entre los 15 y los 25 años, para dejar de crecer a partir de esa edad, mientras la de los hombres sigue aumentando hasta los 30 años y se mantiene a tasas muy altas (sobre el 90%) hasta los 60 años. Esta diferencia es producto de la dedicación exclusiva de muchas mujeres a labores domésticas y de procreación.

Sin embargo, la tasa neta de la población femenina económicamente activa no cae fuertemente a partir de los 25 años, incluso en el Censo del 2000 hay un pequeño repunte conforme avanza la treintena. Está indicando que, además de aquellas que regresan al campo laboral, hay una proporción importante de mujeres que ya no abandonan la actividad laboral por emparejamiento o maternidad.

El crecimiento acelerado de la actividad laboral a edades tempranas es mayor en las zonas rurales que en las urbanas, ello es más notable en los hombres que en las mujeres. La tasa neta de la población económicamente activa en la zona urbana es de 45.7% versus la zona rural que es de 42.6%, según el Censo 2000. A lo largo del tiempo, la mujer abarca más participación de la población económicamente activa, llegando a un 28.5% (388,610 mujeres) del total de la población económicamente activa en el Censo del 2000 (1.364.468 personas), en la cual bajo esta misma fuente, se incrementó la población económicamente activa en un 70% con respecto al del Censo realizado en el año de 1984.

Esto representa un reto para las instituciones públicas, para la equidad en las oportunidades de empleo, salarios y prestaciones que afectan no sólo a la mujer trabajadora, sino a todas las familias. La estructura familiar se ha tenido que adecuar a las necesidades

del aparato productivo, y ello se ha reflejado en su composición tradicional; sin embargo, estos cambios no han sido asumidos plenamente por toda la sociedad. Hay que desarrollar una nueva cultura que permita, desde la familia, conformar una plena equidad con actitudes más democráticas y de respeto a los derechos de las personas.

En la actualidad existe una población que, a pesar de obtener el título universitario, no se incorpora al mercado laboral, y corresponde a un 25.5% del total de mujeres en este nivel superior. Además de que el nivel de Instrucción Superior es el que mantiene una tasa neta de población económicamente activa mayor que la otras con un 68%, en comparación de un 33% en ningún nivel, 41.2% en Primaria y 45% en Secundaria. (Ver Cuadro 4.3)

Cuadro 4.3
COSTA RICA: Tasa neta de participación femenina universitaria
SEGÚN: Título obtenido
2004 (Para población femenina de 25 a 29 años)

Título	Total	No activa	Activa
Universitaria Total	**1.469**	**25,5%**	**74,5%**
Bachillerato	532	29,1%	70,9%
Licenciatura	546	16,3%	83,7%
Posgrado	140	16,4%	83,6%
Profesorado	62	46,8%	53,2%
Técnico, Perito o Diplomado	189	41,8%	58,2%

FUENTE: Idem cuadro 4.1

Casi todos los estudios que analizan los datos del mercado de trabajo coinciden en señalar que el nivel de estudios influye en la situación laboral de una persona, tanto en el nivel de actividad como en el nivel de desempleo.

4.2.4.3 Evolución de la incorporación laboral de mujeres graduadas

A lo largo de los años se muestra como los costarricenses van superando las barreras de la educación y cada año la cantidad de población va en tendencia creciente. En este sentido la población universitaria pasó de representar un 3.0% de la población total en 1973 a un 5.6% en 1984, y según datos del censo 2000 representa un 10.5% de la población total de Costa Rica.

Con respecto a la población femenina mantiene la misma tendencia que el total, es decir, para el 2000 la participación de las universitarias fue de 10.8% del total de mujeres, para 1984 fue de un 5.6% y finalmente para 1973 fue sólo un 2.9%. (Ver Anexo Cuadro A 4.5). En el Censo de 1984 el 52.2% de la población universitaria era liderada por hombres, pero para el último Censo del 2000 se muestra como se invierten los porcentajes y son las mujeres las que tienen la mayor participación en el total de universitarios, con un 51.6%. (Ver Anexo Gráfico A 4.2)

La evolución de la población femenina universitaria se incrementa entre los 25 y los 39 años, para un total de 88,192 mujeres, creciendo aproximadamente en un 166.7% con respecto al Censo de 1984. De esta manera se puede notar que en nuestros tiempos la estudiante se preocupa cada día más por tener mejores conocimientos a nivel educativo, por llegar hasta a la universidad y por equipararse con el hombre. (Ver Anexo Gráfico A 4.3)

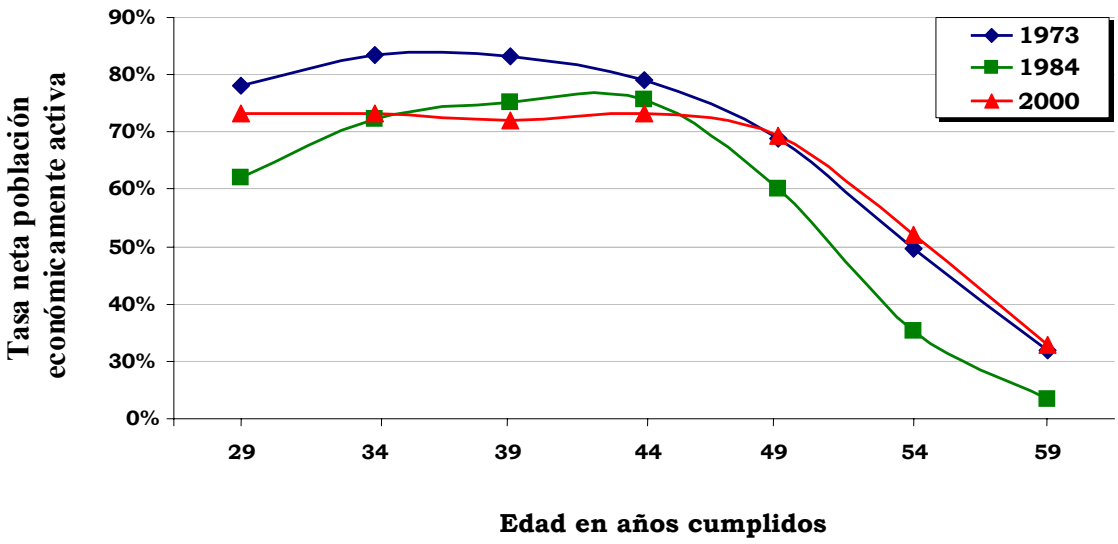
Al igual que la población femenina universitaria aumenta entre los 25 y 39 años, también se puede notar que la tasa de actividad entre estas edades son las que tienen porcentajes de mayor valor a lo largo de sus vidas. Conforme aumenta la edad en las mujeres, se empieza a resaltar que pierden participación de la vida económicamente activa, teniendo un descenso a partir de los 40 años, aspecto similar que se da en otros en censos.

Para el Censo de 1973 se muestra un 76.6% de las mujeres universitarias entre los 25 y 59 años de edad que ingresan a la PEA, alcanzando el mayor valor entre las fechas en estudio, ya que en el Censo siguiente de 1984 fue de un 65.7% y seguidamente el del año

2000 fue un 69.2%, además para este último censo cabe destacar que entre los 30 y los 44 años las mujeres universitaria que ingresaron al PEA tuvieron una participación menor que en otros años siendo un 73.3% del total de estas mujeres, mientras que en año anteriores se alcanzó un 75.6% en 1984 y un 83.5% en el 1973. (Ver Gráfico 4.3)

Gráfico 4.3

COSTA RICA: Tasa neta de población femenina universitaria económicamente activa entre los 25 y 59 años edad. 1973, 1984 y 2000



Fuente: Idem gráfico 4.2.

Como dato relevante se observa que en el Censo 2000, antes de los 45 años las mujeres universitarias económicamente activas presentan indicadores menores a los censos anteriores ya que el punto más alto que se pudo alcanzar fue de 73.3% representado por las mujeres entre 25-29 años y en las de 40-44 años, mientras que estas últimas alcanzaron un 75.6% y 79% en 1973 y 1984 respectivamente.

Un 68.9% de las mujeres corresponden a universitarias económicamente activa, un porcentaje mayor que al de 1984 con un 65.4% pero menor que hace 17 años con 76%. En donde la zona más afectada es la rural ya que año tras años son más mujeres universitarias las que no se incorporan al PEA. Son 14.724 mujeres para el Censo del 2000. Situación muy diferente a la que muestra la zona rural, pero aún sigue existiendo un 41.1% que no son económicamente activas. (Ver Anexo Gráfico A 4.4)

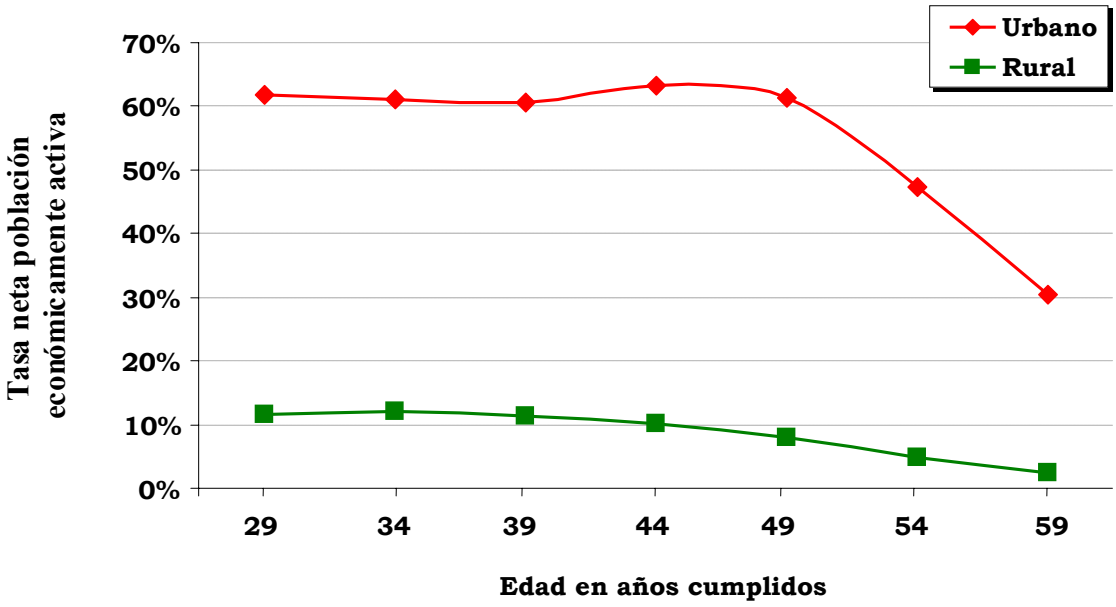
Los problemas de conciliación de vida laboral y familiar que se plantean en las zonas rurales, se relacionan con los que se presentan en zonas urbanas. La vida en el campo tiene algunas implicaciones negativas. En primer lugar, en un entorno donde la media de edad es de 55 años y el nivel de cultura es muy diferente, es más frecuente encontrar en los varones ideas más machistas y tradicionales. Otro grave problema habitual en la zona rural es la falta de infraestructuras.

En las localidades más pequeñas, muchas veces no hay un colegio, porque se comparte un único centro con varios pueblos de la zona. Y es aún menos frecuente la existencia de guarderías para los más pequeños. Los puntos negativos que implica la vida en la zona rural, se matizan con aspectos positivos con los que no se puede contar en las ciudades. En los poblados rurales se suele vivir con un concepto de familia ampliada, donde son muchos los que comparten tareas como la crianza de los niños o el cuidado de los ancianos necesitados.

Además, en los pueblos es frecuente que las distancias sean muy cortas, de modo que una madre puede aprovechar realmente permisos como el de lactancia. En general, el sentimiento de comunidad (inexistente en las ciudades donde prima el individualismo) es una tranquilidad para la madre que, incluso aunque tenga que trabajar durante muchas horas, sabe que siempre habrá alguien dispuesto a ayudar a sus niños si así lo requieran o necesitaran. (Ver Gráfico 4.4)

Gráfico 4.4

COSTA RICA: Evolución de la tasa neta de población femenina universitaria económicamente activa entre 25 y 59 años de edad según la zona de residencia. Censo 2000



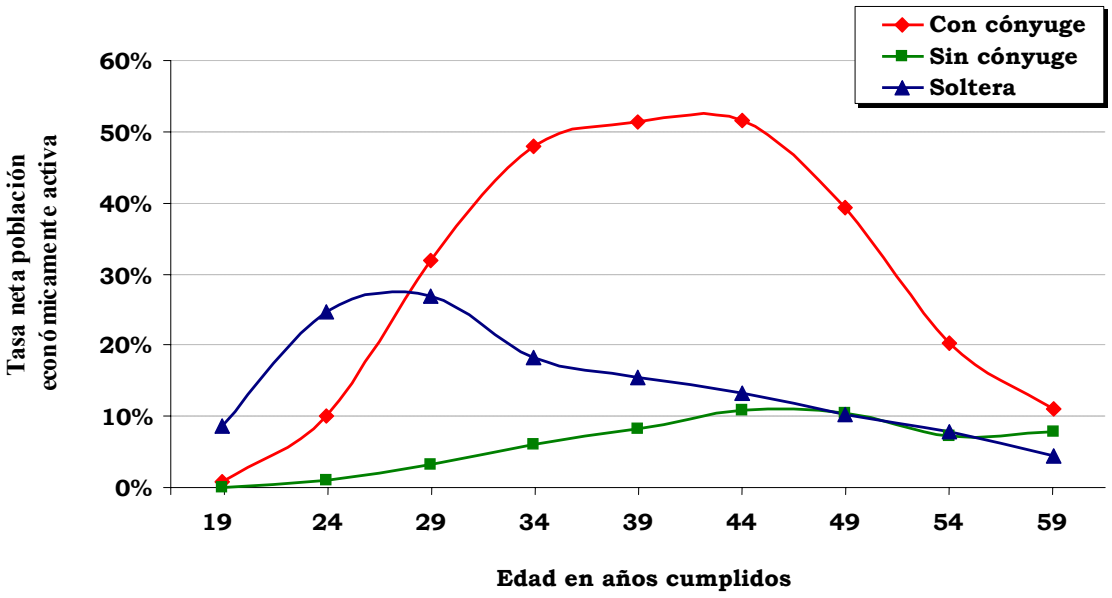
Fuente: Instituto Nacional de Estadísticas y Censos, Censo 2000.

Según el Censo 2000, las mujeres universitarias que son económicamente activas y con cónyuge son las que tienen mayor participación con respecto al resto, siendo un 39.2% de la población femenina con las características anteriormente citadas. Con un 19.6% se reflejan las mujeres que son solteras y por último un 10.1% las que no tienen cónyuge ya sea por separación, divorcio o viudez. (Ver Anexo Gráfico A 4.5). A lo largo de todo el período analizado las tasas de las mujeres sin cónyuge superan claramente a las tasas de las que tienen cónyuge, aunque también es destacable que el cambio más importante se da en las curvas de actividad de las mujeres con cónyuge, y especialmente de las más jóvenes. Pero si bien son cada vez más numerosas las mujeres con cónyuge y con hijos que desean trabajar en el mercado, lo cierto es que las diferencias entre las tasas de actividad laboral de

las que tienen cónyuge y las que no lo tienen siguen siendo muy importantes. Es importante señalar que la categoría "Sin cónyuge" corresponde a la mujer universitaria que estuvo casada o en unión libre y actualmente no está con el cónyuge. Cabe destacar que para el Censo 2000, el INEC subestimó la incorporación de las mujeres universitarias a la población económicamente activa, por ello en el siguiente gráfico la condición de "Soltera" muestra una tendencia decreciente. (Ver Gráfico 4.5)

Gráfico 4.5

COSTA RICA: Tasa neta de población femenina universitaria económicamente activa entre los 25 y 59 años de edad según el estado conyugal. 2000



Fuente: Idem gráfico 4.4.

Por otra parte, podemos analizar que el 52.6% de las mujeres universitarias que no son jefas de hogar, son las que tienen mayor participación en la PEA, mientas que las que sí lo son, sólo tiene un 16.3%, es decir, 23.944 mujeres en total tienen estas condiciones. (Ver Anexo Gráfico A 4.6) Sin embargo, los espacios laborales que las mujeres han logrado abrirse, principalmente a base de estudio, no llegan todavía a contrarrestar la tendencia

hacia la feminización de la pobreza, que se caracteriza básicamente por la mayor presencia de jefatura femenina en los hogares pobres. (Ver Anexo Gráfico A 4.7)

A mayor nivel de educación, menor cantidad de hijos, en la cual la población de mujeres activas para el Censo del 2000 se mantiene con familias de 1 a 2 hijos en el mayor de los casos, correspondiendo a un 16.7% del total de mujeres, seguido por las que no tienen hijos con un 9.9%, es decir, 27.297 mujeres. (Ver Anexo Cuadro A 4.6)

Con estos análisis, es importante considerar que las mujeres graduadas forman parte de la base de recursos humanos esenciales de sus respectivos países. Por eso, tienen derecho a las mismas oportunidades que sus colegas masculinos en lo referente al acceso a la enseñanza superior y a las carreras profesionales. Por otra parte, es necesario considerar la reforma de la educación superior como una prioridad; debería contraerse un firme compromiso de dotar a las mujeres con todas las competencias de gestión necesarias para contribuir a la renovación general de este sector de la educación. Y además la feminización de las funciones de dirección requiere ser analizada y definida con mayor claridad.

4.3 ABORDAJE METODOLÓGICO

En este apartado se presentan los métodos y los procedimientos seguidos para realizar esta investigación. La metodología consiste de una serie de pasos para el logro de los objetivos del estudio y en los cuales se emplea la información recopilada en la Encuesta de Hogares de Propósitos Múltiples de Julio de 2004.

La primera etapa de esta investigación se hizo mediante la revisión bibliográfica, y documental tanto en Internet como en los centros de documentación, se recopiló textos en torno al tema de la población femenina, mujeres universitarias, incorporación de la mujer en la población económicamente activa y la aplicación de modelos logísticos. Esto permitió la construcción de los antecedentes del estudio.

4.3.1 Variable a predecir y uso de modelo de regresión para predecir

Con el fin de alcanzar los objetivos propuestos en este estudio, se requirió el procesamiento y análisis de la Encuesta de Hogares y Propósitos Múltiples del año 2004. El STATA fue el paquete estadístico utilizado en este trabajo, tanto para la preparación, manejo y procesamiento de los datos como para el desarrollo de los modelos ajustados bajo la metodología de regresión logística. El motivo principal de su uso fue la utilidad que tuvo como herramienta de trabajo, además es fácil de utilizar, presenta rapidez y homogeneidad de comandos en plataformas, tiene portabilidad de los datos entre plataformas y ha sido de gran uso en investigaciones.

STATA es un miembro de una generación de programas para procesamiento estadístico de datos, que puede ser utilizado en diferentes plataformas como DOS, WINDOWS, MACINTOSH y UNIX con ligeras variaciones. Para computadoras que trabajen en DOS, existe Small, Pseudos-intercooled e Intercooled Stata, siendo esta última

la que se utilizará en esta investigación en su versión 4.0 para Windows. [Chavarría, Molina y Zamora, 1997]

Entre las características más sobresalientes de este paquete se puede mencionar un eficiente manejo de toda la memoria que el computador posee, poco espacio requerido para su instalación, su rapidez en la ejecución de distintas rutinas que lo conforman, además de que implementa una amplia variedad de procedimientos estadísticos. [Chavarría, Molina y Zamora, 1997]

Con el uso de este paquete estadístico y el archivo de datos en estudio se ajustó un modelo de regresión logística multivariado, utilizando como variable de respuesta la "condición económicamente activa" de la mujer universitaria, la cual se transformó a una variable dicotómica. El motivo de utilizar esta variable es porque permite clasificar a la población entre económicamente activa y económicamente inactiva, según sea el tipo de actividad principal que realizan las personas en cuestión, y para esta investigación en particular nos enfocamos en las mujeres con nivel de instrucción universitaria.

Dentro del estudio se utilizó como herramienta estadística el Modelo de Regresión Logística, ya que es la técnica para el estudio de la relación entre una o más variables independientes (X1, X2, X3....Xn) y una variable dependiente de tipo dicotómica.

4.3.2 Justificación del modelo de regresión más apropiado

El problema de la clasificación es uno de los primeros que aparece en la actividad de la investigación y constituye un proceso consustancial con casi cualquier actividad humana, de tal manera que en la resolución de problemas y en la toma de decisiones la primera parte de la tarea consiste precisamente en clasificar el problema o la situación, para después aplicar la metodología correspondiente y que en buena medida dependerá de esa clasificación. Cuando hablamos de clasificar a un sujeto en un grupo determinado, a partir de los valores de una serie de parámetros medidos u observados, y esa clasificación tiene un

cierto grado de incertidumbre, resulta razonable pensar en la utilización de una metodología probabilística, que nos permita cuantificar esa incertidumbre.

Las respuestas binarias o dicotómicas surgen en muchos campos de estudio, y esta investigación es un ejemplo de ello, de modo que el análisis de regresión logística a diferencia de las otras herramientas estadísticas es usado para investigar la relación entre este tipo de respuestas y un conjunto de variables explicatorios o factores de riesgo. Definiendo como variable dicotómica aquella que solo admite dos categorías que definen opciones o características mutuamente excluyentes u opuestas. [Hosmer y Lemeshow, 1989]

De todos es sabido que este tipo de situaciones se aborda mediante técnicas de regresión. Sin embargo, la metodología de la regresión lineal no es aplicable ya que ahora la variable respuesta sólo presenta dos valores (nos centraremos en el caso dicotómico), como puede ser presencia / ausencia en la PEA.

Desde el punto de vista estadístico podemos distinguir dos enfoques diferentes al problema de la clasificación. En el primero de ellos los grupos están bien definidos y se trata de determinar un criterio para etiquetar cada individuo como perteneciente a alguno de los grupos, a partir de los valores de una serie limitada de parámetros. En este caso las técnicas más utilizadas se conocen con el nombre de análisis discriminante, aunque como veremos existen otras posibles alternativas, tales como la utilización de la regresión logística. El segundo enfoque corresponde a aquel caso en el que a priori no se conocen los grupos y lo que precisamente se desea es establecerlos a partir de los datos que poseemos.

En el análisis discriminante estudiamos las técnicas de clasificación de sujetos en grupos ya definidos. Partimos de una muestra de N sujetos en los que se ha medido variables cuantitativas independientes, que son las que se utilizarán para tomar la decisión en cuanto al grupo en el que se clasifica cada sujeto, mediante el modelo matemático

estimado a partir de los datos. Dentro del análisis discriminante nos encontramos a su vez con dos enfoques diferentes, uno que denominaremos predictivo y otro explicativo.

En el análisis discriminante predictivo se trata de estimar a partir de los datos unas ecuaciones que aplicadas a un nuevo sujeto, para el que se determinan los valores de las diferentes variables, pero del que se desconoce a qué grupo pertenece, nos proporcionen una regla de clasificación lo más precisa posible. Se trata pues de formular un algoritmo por el que se pueda determinar a qué grupo pertenece una nueva observación. En el análisis discriminante predictivo es importante cuantificar con qué precisión se clasificará a un nuevo sujeto.

A diferencia del anterior, en el análisis discriminante descriptivo estamos más interesados en las variables empleadas para diferenciar los grupos, en las variables explicativas, y lo que deseamos es determinar cuáles de esas variables son las que más diferencian a los grupos, cuales son importantes y cuales no, para efectos de clasificar los sujetos.

El enfoque seguido presenta a este tipo de modelos como una forma alternativa de llevar a cabo un Análisis Discriminante sin tener que recurrir a las hipótesis de normalidad y homocedasticidad, ni exigir que las variables clasificadoras utilizadas sean cuantitativas.

De manera que el principal inconveniente del análisis discriminante tradicional radica en que supone que los grupos pertenecen a poblaciones con distribución de probabilidad normal multivariante para las variables explicativas X1 a Xp, con igual matriz de varianzas y covarianzas. Por ello no debiera incluirse en el modelo variables que no cumplieran esa condición, lo que no permite, por ejemplo, la utilización de variables cualitativas.

Sin embargo, en el modelo de regresión logística, se estima la probabilidad de un suceso en función de un conjunto de variables explicativas y en la construcción del mismo no hay ninguna suposición en cuanto a la distribución de probabilidad de esas variables, por

lo que pueden intervenir variables no normales y variables cualitativas. De esta forma podemos considerar la regresión logística como una alternativa al análisis discriminante. Además la interpretación del resultado de aplicar una ecuación logística es más intuitiva al tratarse de un valor de probabilidad.

Como consecuencia de las explicaciones presentadas en párrafos anteriores, la técnica de regresión logística es similar al análisis discriminante, aunque con ciertas diferencias que constituyen una serie de ventajas con respecto al análisis discriminante y que son el motivo para su utilización dentro de esta investigación:

- La variable dependiente o respuesta también presenta dos categorías, pero en este caso representan la ocurrencia y no ocurrencia del acontecimiento definido por la variable, codificándose con los valores uno y cero, respectivamente. Por lo que se refiere a las variables independientes o explicativas, no se establece ninguna restricción, pudiendo ser cuantitativas, tanto continuas como discretas, y categóricas, con dos o más modalidades.
- Definida la variable dependiente como la ocurrencia o no de un acontecimiento, el modelo de regresión logística la expresa en términos de probabilidad, utilizando la función logística para estimar la probabilidad de que ocurra el acontecimiento o de que un individuo elija la opción uno de la variable dependiente, dados determinados valores de las variables explicativas, mediante la siguiente formulación:

$$\pi_i = \frac{e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\cdots+\beta_p x_{pi}}}{1+e^{\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\cdots+\beta_p x_{pi}}} = \frac{1}{1+e^{-(\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+\cdots+\beta_p x_{pi})}} \quad ,, \; i=1,2\ldots n$$

siendo πi = P = (y =1). Puesto que el modelo anterior no es lineal respecto a las variables independientes, se considera la inversa de la función logística, que es el logit o logaritmo de la odds o ventaja de que un suceso ocurra, definiéndose ésta como el cociente entre la

probabilidad de que ocurra un acontecimiento y la probabilidad de que no ocurra, que es su complementaria, como puede observarse en la siguiente expresión:

$$\text{logit}(\pi_i) = \ln\left[\frac{\pi_i}{1-\pi_i}\right] = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} \quad ,, \quad i = 1, 2 \ldots n$$

La formulación anterior facilita la interpretación del modelo y de sus coeficientes, que reflejan, de este modo, el cambio en el logit correspondiente a un cambio unitario en la variable independiente considerada.

- La capacidad predictiva del modelo de regresión logística se valora mediante la comparación entre el grupo de pertenencia observado y el pronosticado por el modelo, que clasifica a los individuos en cada grupo definido por la variable dependiente en función de un punto de corte establecido para las probabilidades predichas, a partir de los coeficientes estimados y del valor que toman las variables explicativas para cada individuo.
- La elección del mismo en detrimento de otros, tales como el análisis discriminante o la regresión lineal, se fundamenta en las características de las variables independientes y la variable dependiente. Así, según indica Sánchez (2000; 431) "el análisis discriminante serviría para abordar situaciones como las descritas; sin embargo, la posibilidad de que coexistan variables independientes de naturaleza cuantitativa y categórica viola la asunción de normalidad multivariante". Además, como este mismo autor afirma, la regresión logística "no sólo solventa las dificultades planteadas por el análisis discriminante, sino que también suple las limitaciones del modelo de regresión lineal respecto a la naturaleza dicotómica de la variable dependiente. Así pues, el modelo de regresión logística es un procedimiento por medio del cual se intenta analizar las relaciones de asociación entre una variable dependiente dicotómica (binaria o dummy) "Y" y una o varias variables independientes (regresores o predictores) "Xn...".

Con base en todo lo anterior, se aplicó la regresión logística a la muestra de 1039 mujeres universitarias procedente de la Encuesta de Hogares de Propósitos Múltiples de julio 2004, y se utilizó para ello el programa STATA y el archivo que contiene los datos está identificado como "bdmujp.dat". Para su ajuste se siguió el método condicional de introducción de las variables consideradas. El objetivo primordial que resuelve esta técnica es el de modelar cómo influye en la probabilidad de pertenecer o no a la Población Económicamente Activa (PEA), habitualmente dicotómico, con respecto a las otras variables en estudio que no son necesariamente dicotómicas.

Se trata de explicar una variable discreta, denominada respuesta (explicada, dependiente, etc) en función de otras que se denominan covariables (explicativas, independientes, etc), estas últimas pueden ser tanto discretas como continuas. El estudio se restringirá al caso en que la variable respuesta sea dicotómica. Por tanto, la regresión logística será adecuada para estudiar la presencia o ausencia de una característica o el resultado, según los valores de un conjunto de covariables.

El modelo de regresión logística dará lugar a una ecuación que permitirá estimar la probabilidad de que ocurra un suceso en el que se está interesado, en función de los valores que toman las variables que se consideran como explicativas. Los parámetros del modelo se estiman mediante el método de máxima verosimilitud. Desde el punto de vista del cálculo efectivo, hay que utilizar algoritmos numéricos sin fórmula general cerrada, a diferencia de lo que ocurría con el método de mínimos cuadrados ordinarios en regresión lineal.

4.3.3 Enfoque del modelo logístico

4.3.3.1 Modelo

Esta investigación es principalmente descriptiva. Se utilizan varias fuentes de información, tanto nacionales como internacionales, con el fin de tener una visión general, aunque no extensiva, sobre el tema en cuestión. A la vez permite conocer las características y factores que influyen en la mujer universitaria para poder incorporarse o no a la población económicamente activa del país.

Consecutivamente se analizó y procesó la información recopilada con el propósito de seleccionar las variables más adecuadas a los objetivos definidos, y que existiera una factibilidad de uso dentro de la base de datos por utilizar en la investigación. En primera instancia, en esta investigación se utilizó una fuente de datos correspondiente a la Encuesta de Hogares de Propósitos Múltiples del año 2004, que fue brindada por el Instituto Nacional de Estadística y Censos, con el propósito de diseñar la base de datos necesaria para dicho estudio con las personas que son objetivo en la investigación.

En un modelo de regresión es necesaria una cantidad representativa de datos para realizar la estimación, de modo que dentro de esta investigación existe suficiente número de observaciones para ser desarrolladas por esta técnica. A partir de esta base de datos, se procedió a dividir en dos grupos la totalidad de los mismos en forma aleatoria por medio de la herramienta estadística SPPS, y representarán una muestra de trabajo y una muestra de comprobación.

El propósito de utilizar la división en la totalidad de la muestra, es para que en la primera submuestra se pueda verificar la bondad del ajuste del modelo. Se utilizará una técnica estadística para ajustar el modelo a una de las partes y, posteriormente, se aplicará los resultados a la otra parte restante. Los resultados logrados con el uso de la primera submuestra son esenciales para formular el modelo teórico de los factores determinantes

que inciden en que las mujeres con nivel universitario se incorporen a la PEA (población económicamente activa) y se procede a seleccionar las variables independientes del modelo por ajustar.

Para obtener los resultados del modelo en estudio se siguió el siguiente proceso:

- Se ajustó un modelo que contenía todas las variables independientes y la variable de respuesta (condición económicamente activa). A partir de este momento se obtuvo el logaritmo natural de la razón de verosimilitud Log Likelihood (se utiliza para estimar los coeficientes de un modelo logístico de regresión, en el que se calcula la probabilidad de que ocurra un determinado suceso).
- Luego se ajusta el modelo para obtener la estimación del Log Likelihood. A efecto de evaluar la significancia de la variable agregada, se calcula la diferencia entre la devianza (-2 Log Likelihood) del modelo completo y del modelo de un reducido.
- Finalizando este proceso se obtienen las probabilidades asociadas a cada "diferencia de devianza" denominada "G" con base en una Chi-cuadrado, y aquellas variables independientes que resultaron ser significativas fueron consideradas para cada modelo final.

La regresión logística no hace supuestos sobre la distribución de las variables independientes. En la regresión logística [Hosmer y Lemeshow, 1989] para una única variable independiente X, el modelo de regresión logística toma la forma:

$$\ln(p/q \mid x) = \alpha_0 + \alpha_1 X$$

o, para simplificar la notación, simplemente:

$$\ln(p/q) = \alpha_0 + \alpha_1 X$$

donde ln significa logaritmo neperiano, α_0 y α_1 son constantes y X una variable que puede ser aleatoria o no, continua o discreta. Este modelo se puede fácilmente generalizar para k variables independientes:

$$\ln (p/q) = \alpha_0 + \alpha_1 X_1 + \ldots + \alpha_k X_k$$

Por simplicidad, se inicia por el modelo simple, extendiéndose después al modelo múltiple.

Es más ventajoso o más conveniente plantear el modelo con el logaritmo de odds que en lugar de plantearlo simplemente con la probabilidad de éxito o con el odds ya que el logaritmo de odds llamado también como razón de posibilidades, razón de ventajas, razón de oportunidades, razón de momios, razón de probabilidades, es el cociente entre la probabilidad de que un evento suceda y la probabilidad de que no suceda. Otra ventaja es que el campo de variación de ln(p/q) es todo el campo real (de $-\infty$ a ∞), mientras que, para p el campo es sólo de 0 a 1 y para p/q de 0 a 1. Por lo tanto, con el modelo logístico no hay que poner restricciones a los coeficientes que complicarían su estimación. Por otro lado, y más importante, en el modelo logístico los coeficientes son fácilmente interpretables en términos de independencia o asociación entre las variables.

4.3.3.2 Supuestos

El modelo de regresión logística no se adjudica linealidad en la relación entre las variables independientes y la dependiente, no requiere que las variables se distribuyan normalmente, no asume homocedasticidad y, en general, tiene requisitos menos exigentes. Este modelo tiene ventajas sobre el análisis discriminante al no requerir supuestos como los citados anteriormente, que en muchos casos son difíciles de cumplir. El éxito de la regresión logística puede establecerse por las tablas de clasificación, las que muestran la correcta e incorrecta clasificación de la variable dependiente.

4.3.3.3 Limitaciones

Las limitaciones de la Regresión Logística residen en que los parámetros del modelo se calculan usando una estimación de máxima verosimilitud. Estas solo son válidas cuando para cada combinación de variables independientes tenemos un número suficientemente grande, posiblemente esta condición sea violada. Puede solucionarse el problema agrupando en categorías donde tenga sentido hacerlo.

4.3.3.4 Resultados

Se dice que un proceso es binomial cuando sólo tiene dos posibles resultados: "éxito" y "fracaso", siendo la probabilidad de cada uno de ellos constante en una serie de repeticiones. A la variable número de éxitos en *n* repeticiones se le denomina variable binomial. A la variable resultado de un sólo ensayo y, por tanto, con sólo dos valores: 0 para fracaso y 1 para éxito, se le denomina binomial puntual.

Un proceso binomial está caracterizado por la probabilidad de éxito, representada por p (es el único parámetro de su función de probabilidad), la probabilidad de fracaso se representa por q y, evidentemente, ambas probabilidades están relacionadas por p+q=1. En ocasiones, se usa el cociente p/q, denominado "odds", y que indica cuánto más probable es el éxito que el fracaso, como parámetro característico de la distribución binomial aunque, evidentemente, ambas representaciones son totalmente equivalentes.

Los modelos de regresión logística son modelos de regresión que permiten estudiar si una variable binomial depende, o no, de otra u otras variables (no necesariamente binomiales): Si una variable binomial de parámetro p es independiente de otra variable X, se cumple $p=p|X$, por consiguiente, un modelo de regresión es una función de p en X que a través del coeficiente de X permite investigar la relación anterior.

Una de las características que hacen tan interesante la regresión logística es la relación que éstos guardan con un parámetro de cuantificación de riesgo conocido en la

literatura como "odds ratio" (aunque puede tener traducción al castellano, se renuncia a ello para evitar confusión ya que siempre se utiliza la terminología inglesa).

El odds asociado a un suceso es el cociente entre la probabilidad de que ocurra frente a la probabilidad de que no ocurra:

$$odds = \frac{p}{1-p}$$

siendo ***p*** la probabilidad del suceso.

Si las variables explicativas tienen carácter dicotómico,, el coeficiente ß de la ecuación logística está directamente relacionado con el odds ratio de tener la presencia a no tenerlo. Esto es, el ratio

Odds Ratio = exp (ß)

es una medida que cuantifica el efecto que ejerce sobre la probabilidad de aparición del suceso el hecho de que se posea el factor correspondiente respecto a no poseerlo suponiendo constantes las restantes variables del modelo.

Cuando la variable explicativa es numérica, el odds ratio cuantifica el cambio en la probabilidad de aparición del suceso cuando se pasa de un valor de la variable explicativa a otro, permaneciendo constante el resto de las variables.

Resta señalar que cuando el coeficiente b de una variable dada es positivo se obtiene un odds ratio mayor que uno. Por tanto, la variable explicativa en cuestión es un factor potenciador de la aparición del suceso. Por el contrario, un coeficiente negativo implicaría que el odds ratio es menor que la unidad, de manera que la variable ejerce un efecto obstaculizador sobre la aparición del suceso.

4.3.3.5 Bondad de ajuste

Siempre que se construye un modelo de regresión es fundamental, antes de pasar a extraer conclusiones, el corroborar que el modelo calculado se ajusta efectivamente a los datos usados para estimarlo.

La utilización de cualquier modelo ajustado para realizar inferencias requiere analizar su bondad de ajuste. Tomando en consideración que el proceso de construcción y ajuste son satisfactorios, además del modelo se incluyen las variables que deben estar, de modo que el correspondiente paso a seguir es analizar la efectividad para describir la variable respuesta y su poder predictivo.

De tal manera, las opciones propuestas y utilizadas en esta investigación para la medición de la bondad de ajuste en el modelo de regresión logística son las siguientes:

1. Tabla de Clasificación: el modelo clasifica cada caso en el grupo que tiene la probabilidad pronosticada más alta; la comparación en una tabla de contingencia de los casos según grupo observado y predicho permite establecer el porcentaje de casos clasificados correctamente en forma global y para cada una de las categorías de la variable dependiente. Es un método para evaluar la capacidad discriminatoria del modelo. No obstante, pese a que se podría obtener una clasificación global no muy alta, es posible que el modelo sea correcto. Cuando los grupos son de tamaño desigual, los casos serán clasificados más probablemente en los grupos grandes, independientemente de que tan bien el modelo ajuste. Pese a que la tabla de clasificación proporciona información interesante, por sí misma dice poco de lo apropiado del ajuste del modelo a los datos. [Ramírez, 2000: 58]

2. Área bajo la curva ROC: Una curva ROC enfrenta en un sistema de ejes la sensibilidad (en el eje y), al complementario de la especificidad (en el eje x). El procedimiento consiste en determinar las correspondientes tablas de clasificación de puntos de corte de $P(Y=1|X)$ crecientes (0.1, 0.2, … , 0.8, 0.9, 1), y determinar a

partir de ellas las correspondientes sensibilidades y especificidades. Si estuviésemos interesados en el punto de corte óptimo para predecir Y, es decir, el punto de corte que ofreciera mayor sensibilidad y especificidad, recurriríamos al análisis gráfico de la curva, seleccionando como punto de corte aquél que correspondiera con el punto de inflexión de la curva ROC. Otra forma sería analizar una gráfica en la que se representaran para cada punto de corte (en el eje x) su sensibilidad y especificidad (en el eje y); el punto de corte óptimo coincidiría con aquél en el que se cruzaran las dos curvas.

Diferentes modelos nos ofrecerán diferentes curvas ROC. La comparación entre modelos respecto a la capacidad predictiva de los mismos puede hacerse comparando la forma de las curvas y el área bajo las mismas; las mejores curvas serán aquellas con área más próxima a la unidad. Como regla general, un área de 0.5 implica ausencia de discriminación; entre 0.7 y 0.79 es una discriminación aceptable; entre 0.8 y 0.89 es excelente; 0.9 ó superior es una discriminación excepcional. El área bajo la curva suele estar implementada en los paquetes estadísticos más usados, pero es fácil de calcular de la siguiente forma:

$$(U_{MW}) / (n_0 \times n_1)$$

donde U_{MW} es el estadístico U de Mann-Whitney y n_0 y n_1 son el número de sujetos con y=0 e y=1.

Es interesante resaltar que un modelo puede tener una mala bondad de ajuste dada por los estadísticos vistos anteriormente, pero tener una buena capacidad de discriminación. Y viceversa, modelos con una buena bondad de ajuste pueden ser malos predictores. [Abraira, 1996]

3. Prueba Pearson Chi-cuadrado: Esta estadística a menudo se emplea para establecer la discrepancia entre los valores observados y los esperados. Valores altos de X^2 de

Pearson indicarían que el modelo no ajusta adecuadamente. La hipótesis nula en este caso sería que el modelo se ajusta a los datos. [Ramírez, 2000: 58]

4. Prueba Hosmer-Lemeshow: La evaluación del grado de coincidencia entre las probabilidades observadas y ajustadas en todo el rango de los valores de las probabilidades se le identifica como calibración del modelo. Una prueba común para este propósito es la de Hosmer-Lemeshow, consistente en dividir los casos en diez grupos aproximadamente iguales basados en la probabilidad estimada de ocurrencia del evento de interés (deciles de riesgo). La hipótesis nula establece que no existe diferencia entre los valores observados y los pronosticados, lo que implica que el modelo ajusta adecuadamente [Ramírez, 2000:59]

5. Comportamiento de residuos de Pearson: utilizado para establecer qué tan bien un modelo se ajusta a los datos observados. Basándose en las probabilidades predichas y las frecuencias esperadas (empleando el modelo ajustado) puede producirse tablas de varias dimensiones con las covariables del modelo incluyendo las frecuencias observadas y las esperadas para cada celda de las variables involucradas. El residuo de Pearson es la diferencia entre el valor observado y pronosticado dividido por una estimación de la desviación estándar. Las celdas con valores superiores a 2 en valor absoluto deben examinarse para tratar de identificar posibles razones por las cuales el modelo podría no ajustar bien. [Ramírez, 2000: 59]

6. P-seudo R2: es un estadístico que pretende simular la función que tiene R^2 en la regresión lineal múltiple, por lo que su valor pretende expresar la cantidad de variabilidad (esto es, de varianza) que es explicada por el modelo; a mayor valor de R^2 mejor sería el modelo. No obstante no se recomienda en general su uso para analizar la bondad de ajuste del modelo. Pero sí podrían ser de utilidad para la comparación de modelos durante la etapa de selección del mejor modelo y para el análisis de multicolinealidad. La de interpretación más directa sería la R^2 de

Nagelkerke, que puede tomar valores entre 0 y 1, y por tanto vendría a valorar el porcentaje de variabilidad explicado por nuestro modelo. [Abraira, 1996]

7. Devianza del modelo. Es una medida del grado de diferencia entre las frecuencias observadas y predichas por el modelo de la variable dependiente, de forma que a mayor devianza, peor es el modelo. Su cálculo es -2 veces el logaritmo neperiano de la verosimilitud del modelo. La devianza nos puede orientar durante la etapa de selección del modelo final. Idealmente el modelo final, el mejor modelo, debería tener la menor devianza de los modelos analizados. [Abraira, 1996].

8. Razón de verosimilitud. El estadístico que se usa es G, que es -2 veces el logaritmo neperiano del cociente entre la verosimilitud del modelo con el conjunto de p covariables introducidas en el mismo y la del modelo sólo con la constante (o más fácil la diferencia entre las devianzas del modelo saturado y el modelo sólo con la constante). Este estadístico sigue una distribución χ^2 con p grados de libertad. Si este estadístico alcanza significación estadística indica un buen ajuste, quiere decir que uno o más de los coeficientes de las covariables introducidas en el modelo es distinto de 0. [Abraira, 1996]

9. Multicolinealidad [Montgomery y Peck, 2002]: significa la existencia de una relación perfecta o exacta entre las variables explicativas de un modelo de regresión. En la actualidad se incluye en la multicolinealidad el término de error estocástico. Representándose de la siguiente forma: $\lambda 1\ x1 + \lambda 2\ x2 + \lambda k\ xk + \nu i = 0$. La multicolinealidad así referida se refiere solamente a relaciones lineales entre variables x. No elimina las relaciones no lineales existentes entre ellas. Se supone que en un modelo clásico de regresión lineal no hay multicolinealidad debido a que: si la multicolinealidad es perfecta los coeficientes de la regresión de las variables x son indeterminados y sus errores estándar son infinitos. Si la multicolinealidad es menos que perfecta los coeficientes de regresión poseen grandes errores estándar, lo que hace que los coeficientes no pueden ser estimados con gran precisión. Por lo

general no existe una relación perfecta entre las variables, pero si puede haber una alta multicolinealidad, situación en la cual es posible la estimación de los coef de regresión β2 y β3.

La solución a la multicolinealidad no es fácil:

- Puede intentarse eliminar la variable menos necesaria implicada en la colinealidad, a riesgo de obtener un modelo menos válido.
- Se puede intentar cambiar la escala de medida de la variable en conflicto (es decir, transformarla), para evitar sacarla del modelo, si bien no siempre se encuentra una transformación de forma directa. Algunas transformaciones frecuentes son el centrado respecto de la media, la estandarización o la creación de variables sintéticas mediante un análisis previo de componentes principales (que es otro tipo de análisis multivariado). Estas transformaciones por el contrario hacen al modelo muy dependiente de los datos actuales, invalidando su capacidad predictiva.
- También se puede recurrir a aumentar la muestra para así aumentar la información en el modelo, cosa que no siempre nos será posible.

Fuentes de Multicolinealidad:

- El método de recolección de información empleado, es decir muestras obtenidas en un rango limitado de valores.
- Restricciones sobre el modelo o en la población que es objeto de muestreo.
- Especificación del modelo.
- Un modelo sobredeterminado. Es decir que posee más variables explicativas que el Nº de observaciones.

La prueba de multicolinealidad se realiza con el propósito de analizar si dos o más de las covariables del modelo mantienen una relación lineal, por lo que se continuó con su detección por medio de regresiones auxiliares, en la cual el primer paso que se siguió fue desentendernos por el momento de la variable dependiente y realizar sendos modelos en los que una de las covariables actuará como variable dependiente y las restantes covariables como variables independientes de aquella. A cada uno de esos modelos se les calculo su R^2 (dispersión total o medida de ajuste), luego la tolerancia R^2 (1- R^2) y factor de inflación de la varianza (FIV) al inverso de la tolerancia (1/(1- R^2)). Cabe destacar que la multicolinealidad es una característica de las muestras no de la población.

Cuando existe estrecha relación entre covariables la tolerancia tiende a ser 0, y por tanto FIV tiende al infinito. Por lo tanto, como regla general debe preocupar tolerancias menores de 0.1 y FIV mayores de 10, concluyendo que bajo esta condición ninguna de las covariables dentro del modelo presenta una colinealidad significativa, a excepción de la variable miembros en la cual su tolerancia es menor a lo establecido en regla general, concluyendo que existe un a multicolinealidad moderada, mostrando una mínima correlación entre covariables.

4.3.4 Fuente de datos

4.3.4.1 Encuesta [INEC, 2004: 18-25]

Los datos se refieren a temas sobre la actividad económica de la población: empleo, desempleo, subempleo e ingresos; y sobre los niveles de pobreza de los hogares. La encuesta investiga además, otras variables sociodemográficas, tales como edad, sexo, estado conyugal, educación y migración.

Por ser de propósitos múltiples, cada año se incorporan módulos especiales de investigación. En esta oportunidad, se incluyeron los siguientes: Vivienda y Servicios, al que se le incorporaron preguntas sobre el sistema de recolección de basura, uso de servicio

doméstico pagado y demanda de financiamiento para compra de vivienda, lote o mejoras. Así mismo se incluyeron preguntas sobre los Regímenes de Pensiones y un módulo especial sobre el Uso del Tiempo.

La Encuesta de Hogares de Propósitos Múltiples, realizada en julio de cada año, tiene como objetivos principales los siguientes: (a) mantener un flujo continuo de estadísticas relacionadas con la fuerza de trabajo, el empleo, el desempleo, el subempleo y los ingresos, y de otras variables socioeconómicas necesarias para el establecimiento de políticas y para la formulación de planes orientados al desarrollo económico y social del país, así como para la evaluación de sus efectos; (b) proveer información periódica, sistemática y oportuna en los períodos intercensales referente, entre otras, a las variables mencionadas; y (c) servir de fuente de información a instituciones gubernamentales, universitarias o de investigación, interesadas en temas relativos a la población y el empleo.

La población de estudio de la Encuesta de Hogares de Propósitos Múltiples está definida como el conjunto de todas las viviendas particulares existentes en el país y de sus ocupantes, que son residentes habituales en esas viviendas. Se excluye del estudio a la población residente en las viviendas colectivas (cárceles, conventos, asilos, residencias colectivas para estudiantiles y trabajadores, hospitales y hoteles).

Los marcos muestrales de viviendas son los instrumentos utilizados para la selección de muestras de viviendas. Para seleccionar las muestras de las EHPM, entre 1987 y 1998 se empleó un marco muestral de viviendas construido en 1986 con la cartografía y la información de los Censos Nacionales de Población y Vivienda de 1984 (MMV-86).

Los dominios de estudio para los cuales se pueden obtener estimaciones con un nivel de confianza conocido son los siguientes:

- Región Central urbana
- Región Central rural
- Resto del país urbano

• Resto del país rural

Además la muestra permite obtener estimaciones de algunas características de la población, para las otras regiones de planificación: Pacífico Central, Chorotega, Brunca, Huetar Atlántica y Huetar Norte.

El diseño muestral corresponde a un diseño probabilístico de áreas, estratificado y bietápico. Es de áreas debido a que las probabilidades de selección están asociadas a los segmentos censales los cuales son áreas geográficas debidamente delimitadas; es estratificado porque para la distribución y selección de la muestra se definieron doce estratos de interés - cada región de planificación dividida por zona urbana y rural - con la finalidad de tener una mejor representatividad de estas áreas y aumentar así la eficiencia relativa del diseño; es bietápico ya que en una primera etapa se seleccionan segmentos censales o Unidades Primarias de Muestreo (UPM), y en una segunda etapa se seleccionan viviendas o Unidades Secundarias de Muestreo (USM) dentro de las UPM seleccionadas en la primera etapa.

El diseño muestral establece que la muestra sea autoponderada dentro de cada estrato. Por ello, como los segmentos se seleccionaron con igual probabilidad, para mantener la autoponderación se toma una fracción fija de viviendas en la segunda etapa de selección: un cuarto (1/4) en los segmentos urbanos y un tercio (1/3) en los segmentos rurales.

Con respecto a la confiabilidad de los datos, se señala que la muestra utilizada en la encuesta es una de todas las muestras posibles del mismo tamaño que podrían haberse seleccionado utilizando el mismo diseño muestral; cada muestra proporciona una estimación del valor poblacional que se desea conocer. El error de muestreo representa una medida de la dispersión o variabilidad de las estimaciones de todas las muestras posibles con respecto a ese valor poblacional que se desea estimar y es de hecho una medida de la precisión de las estimaciones.

Debido a que se selecciona sólo una muestra de todas las muestras posibles, el error de muestreo no es posible calcularlo directamente sino que se estima por medio del error estándar que es una medida del nivel de precisión de las estimaciones de la encuesta.

A diferencia de un censo, los datos de una muestra están sujetos a los errores de muestreo, los cuales se presentan debido a que la investigación estadística se hace sólo en una parte representativa de la población.

Para analizar las estimaciones, es importante tomar en consideración que las celdas con valores pequeños representan características poco frecuentes en la población y, por tanto, los resultados deben ser tratados con cuidado.

4.3.4.2 Reprocesamiento requerido para la construcción del archivo de trabajo

La fuente de datos que se utilizó fue la información contenida en la Encuesta de Hogares Julio 2004, suministrada por el Instituto de Estadística y Censos (INEC) y estaba conformada de 43,779 observaciones con 179 variables contenidas en un archivo en formato ASCII que luego se transformó a la extensión de EXCEL, para su utilización en el trabajo en estudio, en la cual sólo se estudió las mujeres con un nivel de educación universitaria. Además de esta misma institución del gobierno (INEC) se hizo uso de los Censos de 1973, 1984 y 2000 para realizar el estudio sobre la situación del país.

Con el fin de poder realizar el análisis de dicho estudio y cumplir los objetivos establecidos, fue necesario poder transformar la base de datos original de la Encuesta de Hogares Julio 2004 ya que esta por sí sola no incluía algunas variables que eran de interés, además de que el formato establecido no era el idóneo para poder hacer el estudio utilizando la regresión logística.

Lo primero que se realizó fue detectar las variables de interés para el estudio y agruparlas en una sola hoja de Excel con el fin de que no se confundieran con el resto. Mediante este proceso se obtuvo una base de datos de trabajo identificada como

"bdmuj.dat" y que quedó conformada por 2.088 observaciones de las 43,779 originales y 179 variables. Definiendo las variables, se procedió a la codificación de las mismas, lo cual se hizo por medio de la hoja electrónica EXCEL utilizando el comando "VLOOKUP", en donde se tuvo una hoja con las codificaciones y otra con la base de datos sin codificar, de manera que a cada número se le asignó el término al que representaba dicho código.

Se procedió a realizar dos nuevas variables llamadas "años de escolaridad". Una de ellas correspondía a la Jefatura de Hogar y la otra, para su respectivo cónyuge, y consistía en la cantidad de años de estudio dependiendo de su nivel de instrucción, básicamente en la ecuación acumulada desde 0, que indica que no tuvo educación, hasta los 19 años en educación formal, correspondiente al octavo año de universidad. (Ver Anexo Cuadro A 4.7)

Luego se inició la selección de las variables con respecto al primer modelo que se utilizaría para el estudio, de tal manera que se modificó la base de datos original para poder obtener algunas características de los miembros de la familia y que cada Jefe de Hogar se pudiera analizar como una sola línea, es decir, se tuvo que utilizar las características del cónyuge y los hijos del Jefe de Hogar en la misma línea y no en otra diferente como aparecía en la base original utilizando los "IF anidados" de la hoja electrónica EXCEL. Para este caso las variables que se reorganizaron fueron las siguientes:

- Sexo del cónyuge
- Edad del cónyuge
- Nivel de instrucción del cónyuge
- Años de escolaridad del cónyuge
- Condición de actividad del cónyuge
- Ingreso mensual del cónyuge
- Número de miembros en condición económicamente activa en el hogar
- Cantidad de hijos
- Cantidad de hijos en condición económicamente activa

Cabe destacar que la variable correspondiente al Ingreso Total Mensual del Hogar, tuvo que ser transformada a logaritmo natural para simular el ingreso permanente. Se procedió con la construcción de las variables indicadoras[46] (Ver Anexo d) en donde se utilizó la hoja electrónica de Excel para recodificarlas.

Se asumieron como valores fuera de consideración para efectos de este estudio, aquellos que no respondieron o no sabían (NS/NR) y no aplicaban (NA). Además, no se consideró a los menores de doce años y todas aquellas observaciones que presentaban alguna inconsistencia.

La elección de estas variables se realizó una vez concluida la revisión bibliográfica acerca de este tema, lo cual sugirió que estas eran las más apropiadas para alcanzar los objetivos específicos propuestos en la investigación en estudio. Para aumentar la eficiencia y la velocidad de procesamiento de la información, se procedió a diseñar una matriz de datos construida únicamente con las variables y observaciones de interés, tomando en cuenta un filtro que indique a sólo mujeres con nivel educativo universitario, por los objetivos del estudio, de modo que se excluyó de la base de datos inicial la siguiente información:

- Todas las observaciones con al menos una declaración de "missing".
- Todas las variables no pertinentes al estudio.
- Todas aquellas variables para las cuales se construyeron variables indicadoras.

El total de la base de datos está conformado de 2,078 mujeres que cumplen con las características del objetivo del estudio, en la cual por su tamaño se procedió a dividir aleatoriamente este total en dos muestras de 50% de las observaciones cada una. De manera que la segunda muestra convalidará la primera. Ambas muestras están en formato STATA guardadas con el nombre "bdmujs.dat" y "bdmujp.dat".

[46] Variables indicadoras: Asumen el código "0" (cero) cuando la característica está ausente y "1" cuando la característica está presente.

Debido a que el estudio sólo toma en cuenta las variables y definiciones empleadas por la EHPM, podrían no haberse analizado algunas variables que expliquen los modelos estudiados, como por ejemplo, grado de endeudamiento del individuo, el aspecto cultural del Jefe de Hogar, grado de satisfacción con respecto a su condición de actividad, etc.

4.3.4.3 Selección de muestra de datos del archivo

A partir de esta base de datos, se procedió a dividir en dos grupos la totalidad de los datos, realizándola en forma aleatoria por medio de la herramienta estadística SPPS, en la cual representarán a una muestra de trabajo y una muestra de comprobación. Se utilizó una muestra complementaria para la validación del modelo, en la cual incluye el 50% de la población en estudio, ambas submuestras son del mismo tamaño (1,039 mujeres universitarias), es decir, con objeto de validar el modelo se va a seleccionar el 50% de las observaciones, el restante 50% (1.039 observaciones) se empleará para ver si el comportamiento clasificatorio del modelo de regresión logística es correcto.

Esta es una alternativa para aproximar la bondad del ajuste, ya que si el modelo que se diseñó con la primera muestra resulta en valores similares de las mediciones estadísticas para el otro grupo de validación, implica que su ajuste es apropiado y tendría validez. Este procedimiento es particularmente útil cuando el modelo se utilizará con fines predictivos. [Hosmer y Lemeshow, 1989].

4.4 RESULTADOS

4.4.1 Ajuste del modelo

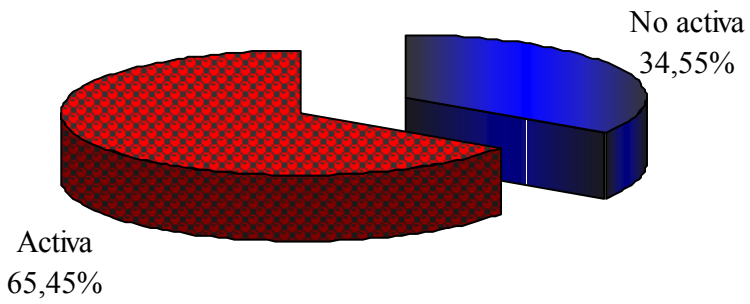
4.4.1.1 Proceso: Descripción de variables y sus características

El INEC define a la Población Económicamente Activa (PEA) de la siguiente manera: son personas de 12 años o más de edad que trabajaron al menos una hora en la semana de referencia o que, sin hacerlo, buscaron trabajo en las últimas cinco semanas. [EHPM, 2004]

La base de datos en estudio muestra que las mujeres universitarias que pertenecen a la población económicamente activa representan el 65.45% del total, versus el 34.55% que no lo están. (Ver Gráfico 4.6).

Gráfico 4.6

Frecuencia relativa para la condición económicamente activa de las mujeres universitarias (n = 1.039)



Fuente: Instituto Nacional de Estadística y Censos. Encuesta de Hogares de Propósitos Múltiples, julio 2004.

Las variables independientes que se utilizaron para ajustar el modelo con el propósito de alcanzar los objetivos establecidos en esta investigación fueron seleccionados según la información correspondiente a la situación actual de las mujeres universitarias haciéndose mención en el punto 3.2 de este documento, además del alcance de las variables disponibles dentro de la base de datos de la Encuesta de Hogares de Propósitos Múltiples para el año 2004 y la construcción de nuevas variables utilizando esta misma fuente, de esta manera se procedió a definir y analizar 26 variables en total. Estas variables son de mucha importancia en la investigación de las mujeres universitarias y la selección de este conjunto de variables se ha realizado basándose en el criterio objetivo y fundamento en estudios anteriores a este. De esta manera la elección de estas variables se realizó una vez concluida la revisión bibliográfica acerca de este tema, la cual sugirió que ellas eran las más apropiadas para alcanzar los objetivos específicos propuestos en esta investigación.

Para la elaboración de la estadística descriptiva del conjunto de datos observados, se consideró que el número de casos sobre los que obtuvimos los datos es mil treinta y nueve, de manera que se procederá a mostrar las estadísticas descriptivas para cada una de las variables.

- **Relación de parentesco con el jefe de familia – relpar -:** es una variable cualitativa que se convirtió en dicotómica en donde 1 corresponde a "Jefe de Hogar" y 0 es "No Jefe de Hogar", y representa la relación de parentesco que tienen las mujeres universitarias en estudio con el Jefe de Hogar. En el gráfico de paste, la frecuencia de esta variable nos damos cuenta que, la relación de parentesco que predomina entre las mujeres universitarias es la de no ser jefe de hogar, representando un 80.37%. es decir 835 de las personas en estudio no son jefas de hogar.

  De esta mayoría de mujeres universitarias que no son Jefes de Hogar, se tiene que 528 de ellas tienen una condición económicamente activa (63,23%), mientras que las restantes 307 mujeres no están incorporadas a la PEA. Es

importante destacar que no todas las mujeres universitarias que son Jefes de Hogar (19.63%) están en condición económicamente activa, de hecho sólo 152 mujeres cumplen con esta condición, es decir el 74,51% de las mujeres que son Jefes de hogar, caso inverso a las restantes 52 mujeres que no se mantienen económicamente activas.

Por otro lado, la mayoría de las mujeres universitarias que no son Jefes de Hogar viven en la zona urbana representando un 54,38% del total de mujeres en esta situación. Este comportamiento es similar en el caso de las otras mujeres universitarias que son jefes de hogar sólo que su porcentaje de participación en la zona urbana es mayor, representado un 67.16% del total de mujeres en esta condición. (Ver Gráfico 4.7)

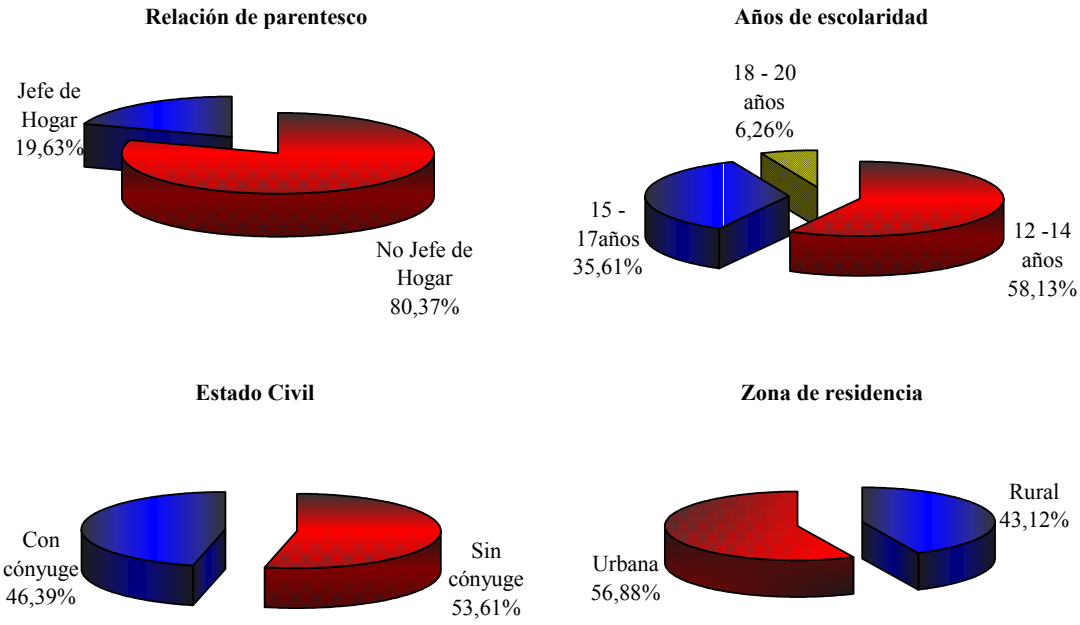
- **Estado civil –estciv-:** es una variable nominal que se transformó en dicotómica con el propósito de que 1 representara a las mujeres con cónyuge y 0 corresponde sin cónyuge. En el Gráfico 4.7, la frecuencia de esta variable se puede observar que el 53.61% de las mujeres universitarias en estudio no tienen cónyuge, y se relaciona con la variable edad analizada anteriormente, ya que la mayoría de las mujeres en estudio son jóvenes adultas (16 a 25 años), por lo cual existe una alta probabilidad de que todavía no tengan cónyuge. Por otro lado el 65.35% del total de las mujeres universitarias que no tienen cónyuge están en condición económicamente activa, mientras que las restantes 193 mujeres en estudio no se incorporaron en la PEA. En cambio del total de las mujeres universitarias que si tienen cónyuge el 65.56% si pertenecen a la PEA, mientras que las restantes 166 mujeres en estudio están económicamente inactivas. Cabe destacar que la mayoría de estas mujeres universitarias sin cónyuge viven en la zona urbana, representando el 57.45% del total en esta condición, esta zona también es de preferencia en las mujeres universitarias que si tiene cónyuge. (Ver Gráfico 4.7)

- **Cantidad de años de escolaridad –anoses-:** es una variable continua que se construyó a partir de otra variable correspondiente al nivel de instrucción en la cual se codificó según el último grado que aprobó (Ver anexo 3.11). Representa la cantidad de años en institutos de educación que han cursado las mujeres universitarias, tomando en cuenta su último año aprobado de estudios. En el gráfico de frecuencia de esta variable se puede observar que la cantidad de años de escolaridad que tiene mayor participación es entre 12 y 14 años, siendo un 58.13% del total de la muestra, de manera que el mínimo valor observado es de 12 años y el máximo es de 22 años, el promedio de los datos observados es de 14 años, la varianza de los datos de esta variable es de 3.18, lo cual indica que no existe una alta variabilidad de los datos. (Ver Anexo Cuadro A 4.8) El 76.04% de las mujeres universitarias en condición económicamente inactiva han tenido entre 12 y 14 años de escolaridad, mientras que en las mujeres universitarias incorporadas en la PEA la mayor participación se da entre las mujeres con más de 15 años de escolaridad representando el 51,32% de las mujeres en esta condición. Tanto en la zona urbana como en la rural la mayor participación de las mujeres universitarias pertenece a las que tienen entre 12 y 14 años de escolaridad, siendo un 62.5% y 54.82% respectivamente. (Ver Gráfico 4.7)

- **Zona de residencia –zona-:** esta es una variable que se transformó en dicotómica en donde 1 corresponde a la zona Urbana y 0 es la zona Rural, y representa la zona residencial de las mujeres universitarias en el momento de realizar la encuesta. La zona de residencia Urbana es la que más predomina dentro del estudio, reflejando un 56.89% de las mujeres universitarias, mientras que el restante 43.12% viven en la zona rural. (Ver Gráfico 4.7)

Gráfico 4.7

COSTA RICA: Frecuencias relativas de la población femenina universitaria según Relación de parentesco, Años de escolaridad, Estado civil y Zona de residencia. Julio 2004

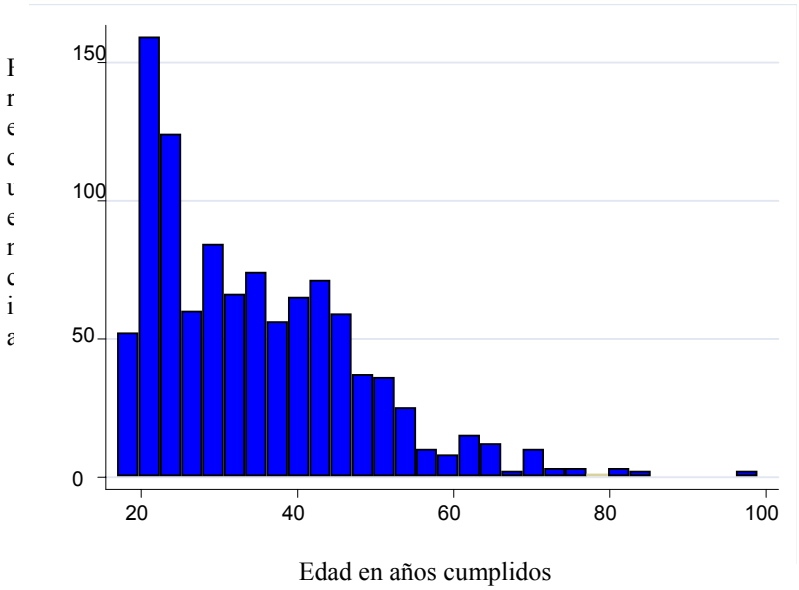


- **Edad – edad -:** es una variable continua que indica la edad en años cumplidos de las mujeres universitarias en estudio. En el histograma de frecuencia de esta variable se puede observar que la edad de las personas en estudio hacen el 80.37% entre los 16 y 45 años. Siendo el rango de 16 a 25 años es más significativo con un 32.24% de la muestra total, reflejando que las mujeres en estudios son relativamente jóvenes adultas en las cuales acaban de finalizar el grado universitario. Por otro lado el 85.29% de las mujeres en condición económicamente activa tiene entre 16 y 45 años de edad, mientras que el 71% de las mujeres universitarias que no están incorporadas a la PEA pertenecen al rango de esta edad. El mínimo valor observado

es de 16 años y el máximo es de 85 años, el promedio de los datos observados es de 35 años, la varianza de los datos de esta variable es de 174.62, lo cual indica que existe una alta variabilidad de los datos. (Ver Anexo Cuadro A 4.8)

Gráfico 4.8

COSTA RICA: Frecuencia absoluta de la población femenina universitaria según la edad (en años cumplidos). Julio 2004



- **Cantidad de aposentos en la casa de habitación –aponse-:** esta es una variable discreta que indica la cantidad de aposentos que existen dentro de la casa de habitación de las mujeres universitarias en estudio. En el gráfico de frecuencia de esta variable se puede observar que en la mayoría de las casas donde habitan las mujeres universitarias, posee entre 3 y 5 aposentos, representando un 75.71% del total de la muestra. Por otro lado el 78.55% de las mujeres universitarias que no están incorporadas a la PEA tienen entre 3 y 5 aposentos en sus casas, mientras que el 73.38% de las mujeres universitarias en condición económicamente activas tienen este mismo rango de aposentos. Con respecto a la zona de residencia, se tiene que

las mujeres universitarias que tiene entre 3 y 5 aposentos en su casa, corresponden a un 77.23% del total de la zona rural y un 73.60% del total en la zona urbana. El mínimo valor observado es de 1 aposento y el máximo es de 8, el promedio de los datos observados está dado por 3.12 aposentos y la varianza de los datos de esta variable es de 0,88 lo cual indica que no existe una alta variabilidad de los datos. (Ver Anexo Cuadro A 4.8 y Gráfico 4.10)

- **Cantidad de miembros en condición económicamente activa (acthog):** es una variable discreta que se construyó a partir de la variable "condición económicamente activa" y la agrupación por hogar, la cual representa la cantidad total de miembros en el hogar en la que su condición económica es activa. En el gráfico de frecuencia de esta variable se puede analizar que el 74.40% de las mujeres universitarias tiene menos de 2 miembros del hogar que pertenecen a la población económicamente activa. Los resultados de esta variable se ven relacionadas con las conclusiones de las anteriores, ya que en la mayoría son jóvenes de menos de 25 años en la cual un alto porcentaje no tiene cónyuge, de esta manera la cantidad de personas en condición económicamente activa es menor que las mujeres universitarias en condiciones diferentes. Por otro lado el 86.90% del total de las mujeres universitarias que no se han incorporado a la PEA tienen menos de 2 miembros activos en el hogar, mientras que del total de las que si están en condición económicamente activas representan un 67.80% con menos de 2 activos en el hogar. La cantidad de aposentos que está de moda son 2, existiendo una desviación estándar del 1,06 y mostrando que su rango mínimo es 1 y el máximo es 8 personas en condición económicamente activas dentro del hogar. (Ver Anexo Cuadro A 4.8 y Gráfico 4.10)

- **Cantidad de hijos (hij):** es una variable discreta que se construyó a partir de la variable "relación de parentesco con el jefe de hogar" y la agrupación por familia, que corresponde a la cantidad total de hijos en el hogar. En el histograma de

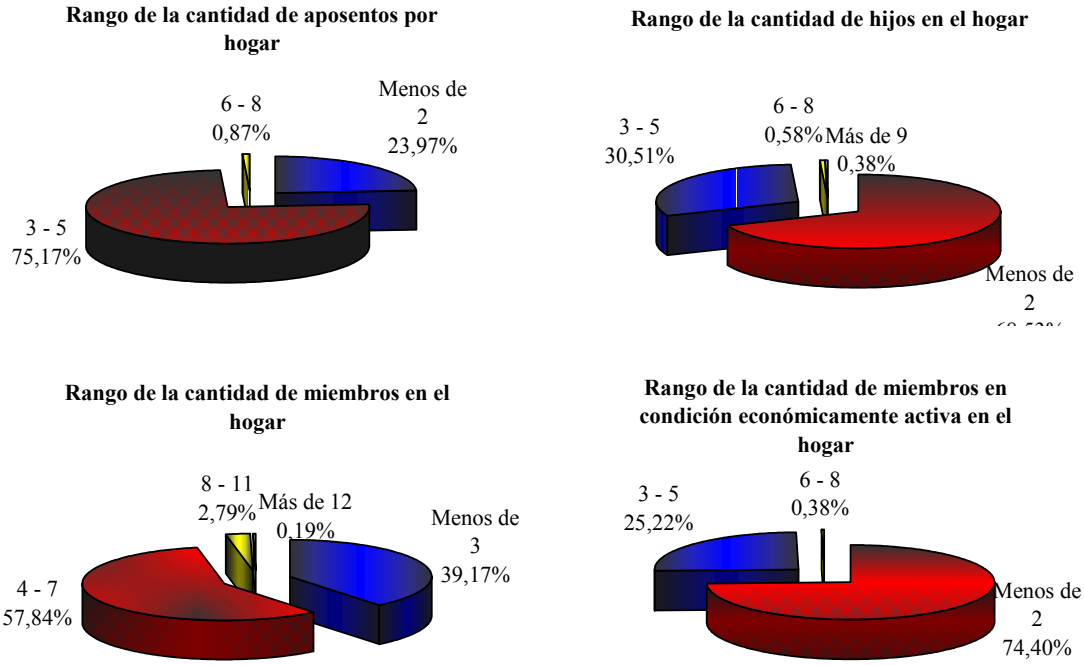
frecuencia de esta variable se puede observar que el 68.53% de las mujeres universitarias tiene menos de 2 hijos, caso similar que se da en el resto de mujeres con otro nivel de instrucción, siendo este monto el promedio de hijos dentro de un hogar. Por otro lado se tiene que el 79.59% del total de las mujeres universitarias en condición económicamente activa tienen menos de 2 años, mientras que el 64.62% del total de las mujeres universitarias que no están incorporadas en la PEA tienen menos de 2 hijos. Con respecto a la zona en que viven las mujeres universitarias, se tiene que la mayor participación es entre las que tienen menos de 2 hijos, correspondiente a un 63.33% del total de mujeres universitarias que viven en la zona rural y un 72.41% del total de las que viven en la zona urbana. Si se realiza un cruce con la variable estado civil, se podrá observar que el 61.22% del total de las mujeres universitarias sin cónyuge tienen menos de 2 hijos, mientras que el porcentaje es mayor con las que tiene cónyuge, representado un 76.97% del total en este estado civil. El rango mínimo corresponde a mujeres sin hijos y el máximo a 10 hijos en total. La varianza muestra que no existe una alta variabilidad en los datos.(Ver Anexo Cuadro A 4.8 y Gráfico 4.10)

- **Cantidad de miembros (miemb):** es una variable discreta que se construyó a partir de la agrupación del hogar y que representa la cantidad de miembros en el hogar. En el histograma de frecuencia de esta variable se puede observar que el 57,84% de las mujeres universitarias integran hogares de 4 a 7 miembros. Este es el rango más significativo, seguido de los que tienen menos de 3 miembros que corresponde a un 39.17% del total de la muestra. Por otro lado el 58.77% del total de las mujeres universitarias que no están incorporadas a la PEA viven con 4 a 7 miembros dentro del hogar, mientras que con un porcentaje similar (57.35%) del total de mujeres universitarias en condición económicamente activa habitan con 4 a 7 miembros dentro de su respectivo hogar. De igual forma, existe una misma proporción entre el total de las mujeres universitarias que no tienen cónyuge y las que si, siendo un 59.96% y 55.39% respectivamente. El mínimo valor observado es

de 1 persona y el máximo es de 12 personas; el promedio de los datos observados es de aproximadamente 4 miembros en el hogar y la varianza de los datos de esta variable es de 11,36, lo cual indica que existe una alta variabilidad de los datos. (Ver Anexo Cuadro A 4.8 y Gráfico 4.10)

Gráfico 4.10

COSTA RICA: Frecuencias relativas de la población femenina universitaria según: Cantidad de aposentos por hogar, Cantidad de hijos en el hogar, Cantidad de Miembros en el hogar y Cantidad de miembros en el hogar en condición económicamente activa. Julio 2004

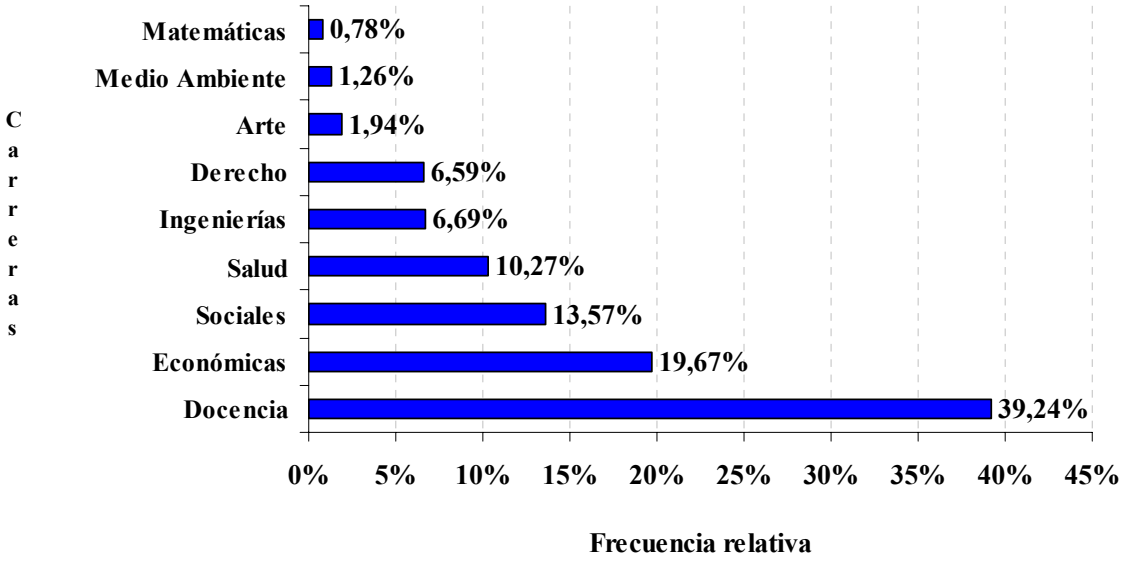


- **Título universitario –titul-:** esta es una variable discreta que se transformó en dicotómica en donde 1 corresponde a con título y 0 sin título, y representa si la mujer universitaria obtuvo o no el título universitario. El obtener el título universitario es el que más predomina dentro del estudio, reflejando un 62,17% de las mujeres universitarias. (Ver Gráfico 4.11)

- **Carreras universitarias:** con respecto a esta variable discreta, se tuvo la necesidad de dividirlas en 9 variables más correspondiendo a cada una de las carreras universitarias con el fin de hacerlas dummyes: Artes, Medio Ambiente, Ciencias Económicas, Física, Matemáticas y Estadística, Ciencias de la Salud, Ciencias Sociales, Derecho, Docencia e Ingenierías. De tal manera cada una de ellas se codificó, en donde 1 corresponde a la presencia de esa carrera en específico y 0 la ausencia de la misma, de modo que en forma general las carreras que tienen mayor participación fueron las relacionadas con Docencias, Ciencias Económicas y Ciencias Sociales, representado aproximadamente el 72,48% del total de la muestra. (Ver Gráfico 4.11)

Gráfico 4.11

COSTA RICA: Frecuencias relativas de las mujeres universitarias según las carreras universitarias. Julio 2004

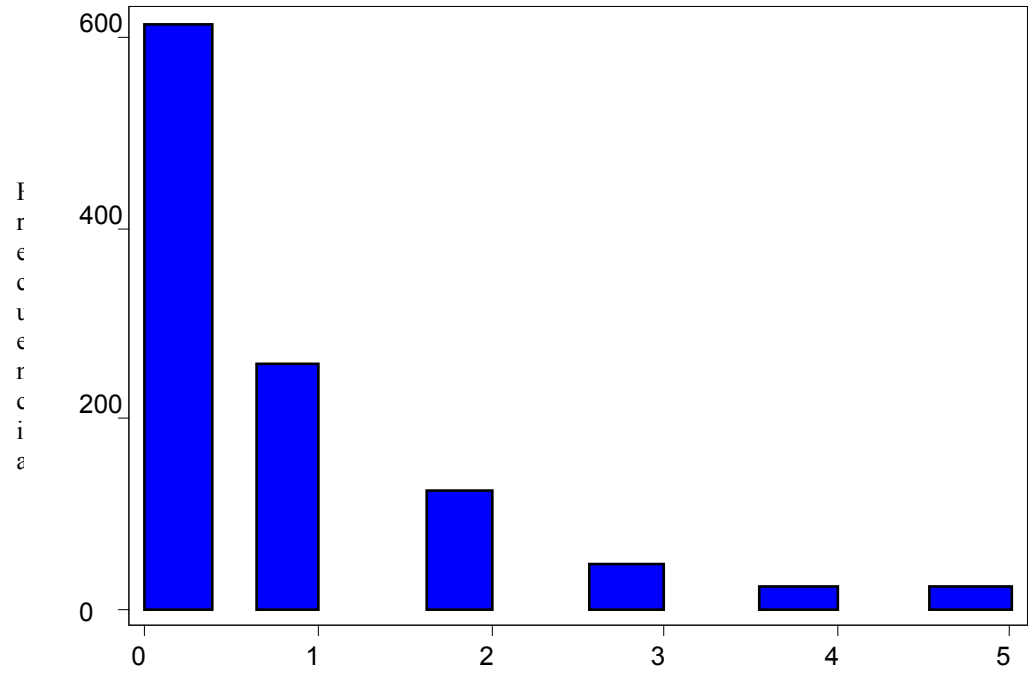


- **Cantidad de hijos en condición económicamente activa (hijact):** es una variable discreta que se construyó a partir de dos variables que corresponden a la condición

económicamente activa y la relación de parentesco, agrupándolos según la familia. Esta variable indica la cantidad de hijos en total que su condición económica es activa dentro del hogar. Siguiendo con los resultados que se han mostrado y son característicos de las mujeres universitarias dentro de este estudio, se puede añadir que la mayoría no tiene hijos en condición económicamente activa, representado un 59.10% del total de la muestra. El rango mínimo corresponde a ningún hijo en condición económicamente activa y el rango máximo es de 5 hijos en dicha condición. (Ver Gráfico 4.12)

Gráfico 4.12

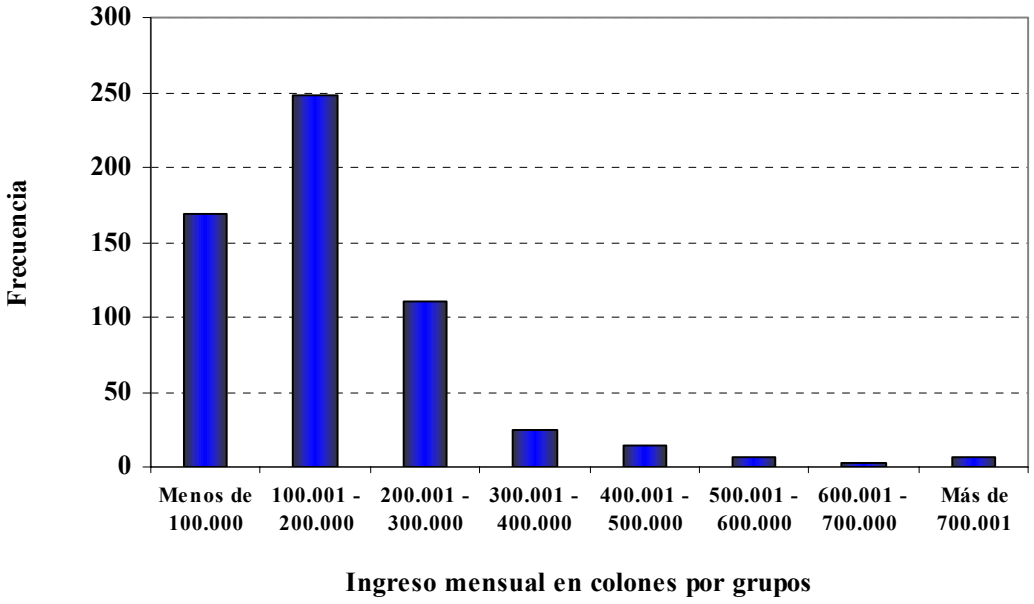
COSTA RICA: Frecuencia absoluta de la población femenina universitaria según la cantidad de hijos en condición económicamente activa. Julio 2004



- **LN ingreso mensual total (lningt):** es una variable continua que muestra el logaritmo natural del ingreso total de hogar en forma mensual. En el histograma de frecuencia de esta variable se puede observar que 156 mujeres universitarias aseguran percibir más de 700.000 colones en el ingreso mensual total del hogar, representado un 18.16% del total de la muestra; de esta manera observamos que este es el mayor rango dentro de la investigación. Con respecto a los demás rangos, no se diferencia en gran medida con el anterior, y existen proporciones muy similares entre las mujeres universitarias que indicaron que el ingreso mensual total del hogar está entre 100.001 y 300.000 colones. El mínimo valor observado es de 25.000 colones y el máximo es de 5.197.802 colones. El promedio de los datos observados es de 460.626 colones. (Ver Anexo Cuadro A 4.8 y Gráfico 4.13)

Gráfico 4.13

COSTA RICA: Frecuencia absoluta de la población femenina universitaria según el ingreso mensual total del hogar. Julio 2004



La información correspondiente a características del cónyuge resultó ser escasa, siendo insignificativas para el estudio, por ello la descripción de estas variables será en base al total de respuestas suministradas dentro de la base de datos de la Encuesta de Hogares de Propósitos Múltiples de julio 2004:

- **Edad (edcony):** es una variable continua en la que muestra la edad en años cumplidos del cónyuge de las mujeres universitarias en estudio. De las 29 mujeres universitarias que contestaron la edad de su cónyuge, se tiene que el 58.62% señalaron que son de edades mayores a 40 años, en la cual el rango mínimo que se definió fue de 25 años y el de máximo rango fue de 54 años de edad. Cabe destacar que las mujeres universitarias que dijeron tener cónyuge fueron 482 del total de la muestra, a pesar de que sólo el 6.01% dio respuesta en esta variable.

- **Nivel de instrucción (nivicony):** es una variable dicotómica que representa el nivel educativo del cónyuge de las mujeres universitarias en estudio. En donde 1 representa que es universitario y 0 no es universitario. Dentro de los resultados se tiene que el 47.51% del total las mujeres universitarias que comentaron tener cónyuge, concluyeron que su cónyuge no tiene título universitario.

- **Cantidad de años de escolaridad (anescony):** es una variable continua que se construyó a partir de la variable nivel de instrucción y se codificó según la cantidad de años de estudio, definido por el último nivel de instrucción que aprobó. El rango mínimo en los años de escolaridad del cónyuge fue de 6 mientras que el máximo fue de 17, siendo valores un poco más bajos a los definidos por las mujeres universitarias. Cabe destacar que la cantidad de años de escolaridad que tuvo mayor frecuencia fue la de 11 años, es decir en secundaria.

- **Condición económicamente activa (actcony):** esta variable dicotómica en la que representa que 1 es económicamente activo y 0 no lo es, muestra la condición activa del cónyuge de las mujeres universitarias en estudio. Con respecto a esta

variable se tiene que el 97.35% de las personas que contestaron, afirmaron que de su cónyuge pertenece a la población económicamente activa, es decir 26 personas del total en esta condición.

- **Título universitario (titulcony):** es una variable dicotómica en donde 1 = Con título universitario y 0 = sin título universitario, del cónyuge de las mujeres universitarias en estudio. Con respecto a esta variable se tiene que el 32.55% de las personas que contestaron, afirmaron que de su cónyuge pertenece a la población económicamente activa, es decir 17 personas del total en esta condición.

4.4.1.2 Precisión del ajuste

En un modelo de regresión logística, se tiene que contrastar su significación global. Esto se hace mediante las pruebas de ajuste global del modelo, en la cual en las técnicas estadísticas que se utilizaron dieron como consecuencia que el ajuste del modelo resultó bastante satisfactorio dentro de esta investigación.

Es importante señalar que con sólo el uso de una de estas mediciones era suficiente para determinar si el modelo ajusta bien a los datos, sin embargo se consideró conveniente el resultado de las siete técnicas para corroborar su respuesta unánime de que el modelo propuesto ajusta muy bien con los datos. Posteriormente a este desarrollo, se realizó la aplicación del modelo ajustado a una muestra de casos adicionales en la cual permitió confirmar el poder predictivo del modelo ajustado, resultando también satisfactorio. Por lo tanto, en los párrafos siguientes se explicará el uso y los resultados de cada una de estas mediciones para la bondad de ajuste.

El modelo nulo como punto de partida tiene un valor en el modelo completo de LL es de -474.36437, asignando a todos los grupos una probabilidad igual al promedio de la población. La introducción de los factores explicativos citados aumenta el modelo nulo a LL = -471.1920, con un valor para la Chi-cuadrado altamente significativo en el cual las frecuencias observadas y teóricas coinciden completamente (χ^2 =0). Por lo tanto, el valor

del Likelihood disminuye sensiblemente entre ambas instancias y tendiende a cero de manera que el modelo predice bien los datos. (Ver Anexo Cuadro A 4.10)

El Pseudo R^2 es de 59,65%, representado el porcentaje que el modelo ajustado explica la probabilidad de que una mujer universitaria esté en condición económicamente activa, es decir el valor del pseudo R^2 indica que el modelo explica un 59,65% más que el modelo nulo. Este valor se considera apropiado para que el modelo ajuste bien a los datos.

Modelo reducido: Regresión logística

Análisis de variancia

Estimados Logit	Número de obs =	1039
	LR chi2(7) =	397.16
	Prob > chi2 =	0.0000
Log verosimilitud = -471.19201	Pseudo R2 =	0.5965

conact	Coef.	Desv Estánd	z	P>z	[95% Conf.	Interval]
relpar	1.215183	.2449121	4.96	0.000	.7351637	1.695201
edad	-.0177421	.0065792	-2.70	0.007	-.030637	-.0048471
anoes	.4409987	.0553212	7.97	0.000	.3325712	.5494262
acthog	1.921149	.151011	12.72	0.000	1.625173	2.217125
hij	.1963485	.1083713	1.81	0.070	-.0160553	.4087522
hijact	-.3785737	.1363711	-2.78	0.006	-.6458561	-.1112913
miemb	-.61587	.1021784	-6.03	0.000	-.816136	-.415604
_cons	-6.427418	.811707	-7.92	0.000	-8.018335	-4.836502

Otra forma para medir la bondad de ajuste es la Prueba Pearson Chi-cuadrado. Dentro de este estudio esta estadística es altamente significativa, como se observa en los resultados del programa STATA, indicando que el modelo se ajusta apropiadamente ya que

su probabilidad es 0. Sin embargo, la prueba de "Pearson chi-square"no es la más recomendada, ya que las probabilidades estimadas para cada patrón de covariable pueden mantenerse muy bajas aún cuando el tamaño de la muestra se incremente, por esta razón se prefirió utilizar también la prueba de Chi2 – Hosmer y Lemeshow, para evaluar el adecuado ajuste del modelo en todos sus aspectos dentro de la regresión logística, por lo que el valor obtenido para la prueba implica que no existe evidencia estadística para rechazar la hipótesis nula de que el ajuste del modelo es bueno.

Modelo reducido: Regresión Logística

Tabla de contingencia para la prueba Hosmer-Lemeshow

Grupo	Prob	Obs 1	Exp 1	Obs 0	Exp 0	Total
1	0,2004	9	11,8	95	92,2	104
2	0,3788	23	30,8	82	74,2	105
3	0,4973	45	44,9	58	58,1	103
4	0,6357	59	60,4	45	43,6	104
5	0,7377	78	71,4	26	32,6	104
6	0,8131	88	81,4	16	22,6	104
7	0,8667	91	87,8	13	16,2	104
8	0,9164	97	92,8	7	11,2	104
9	0,9588	92	97,7	12	6,3	104
10	1,0000	98	101,0	5	2,0	103

Número de observaciones = 1,309

Número de grupos = 10

Hosmer-Lemeshow chi2(8) = 20.48

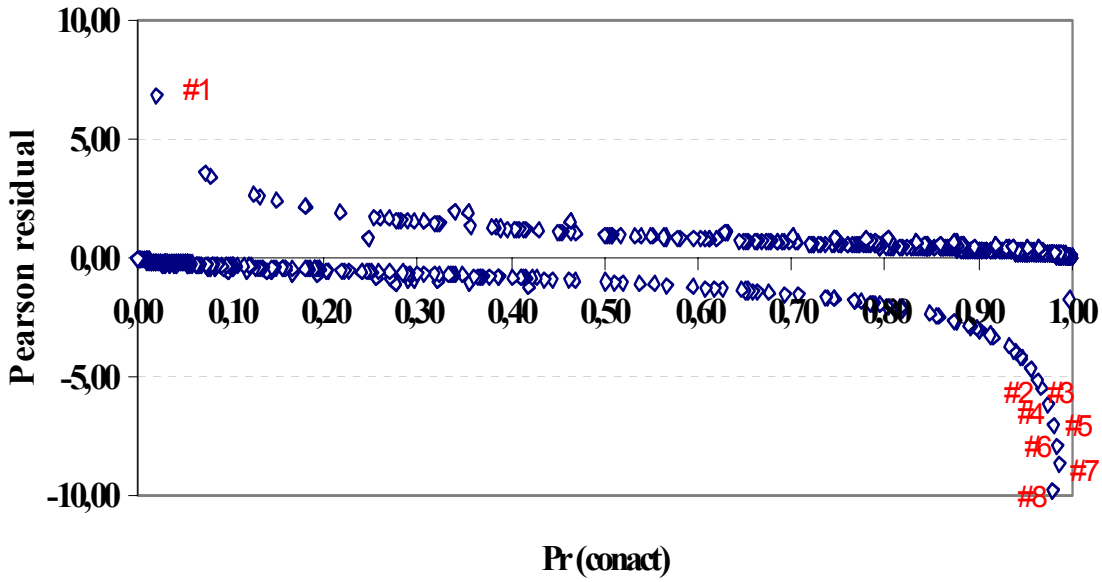
Prob > chi2 = 0.1813

Para tener un criterio aún más sólido y confiable sobre el correcto ajuste del modelo, se realizó la evaluación y el estudio de los residuos, por medio de un gráfico de los valores predichos versus los residuos, en la cual se puede apreciar que la mayoría de los puntos se encuentran en una franja alrededor de 0 a excepción de pocos puntos ubicados en la parte inferior derecha y superior izquierda. La ausencia de patrones claros de comportamiento de

los residuos es evidencia de lo apropiado del modelo. De esta manera a pesar de la dispersión obtenida, existe cierta tendencia de los puntos al valor cero. Esta tendencia se debe básicamente a que sólo hay 680 mujeres universitarias catalogadas en condición económicamente activa (presencia del fenómeno estudiado), es decir un 65.45% del total de la muestra en estudio. (Ver Gráfico 4.14)

Gráfico 4.14

Modelo reducido: Regresión logística
Análisis de residuos



El gráfico sugiere que los valores estimados y los residuos no presentan una fuerte asociación, al menos en un sentido lineal, esta evidencia permite asegurar, en buena medida, que las variables consideradas explícitamente en el modelo tienen gran influencia en la explicación de la condición económicamente activa, en cuanto a si su condición es activa o no. Del mismo gráfico es también clara la ocurrencia de nueve residuos "atípicos",

en el sentido de que estos se alejan de las bandas establecidas (-5 y 5) en base a una aproximación normal de los residuos, separándose en alguna medida, del resto de los puntos que representan cada diferente combinación de residuo y valor estimado (probabilidad estimada para ese patrón de covariable).

Los residuos "atípicos" se han numerado como #1, #2, #3 y así sucesivamente hasta el #8 y graficados en color rojo con el fin de facilitar su ubicación e identificación; el siguiente gráfico es el correspondiente patrón de covariable (o residuo) a los cuales corresponden:

- 1: Es la observación correspondiente a la mujer universitaria en condición económicamente activa, no es jefa de hogar, con 37 años de edad, en la cual han transcurrido 13 años de escolaridad, que vive en una casa con seis miembros de los cuales en dos tienen condición económicamente activa y es madre de 2 hijos, que no están en condición económicamente activa. Su probabilidad estimada (valor predicho) es de 0,02 y el residuo correspondiente es de -6,85.

- 2: Es la observación correspondiente a la mujer universitaria en condición económicamente no activa, no es jefa de hogar, con 23 años de edad, en la cual han transcurrido 14 años de escolaridad, que vive en una casa con seis miembros de los cuales en tres tienen condición económicamente activa y es madre de 2 hijos, que no están en condición económicamente activa. Su probabilidad estimada (valor predicho) es de 0,96 y el residuo correspondiente es de -5,16.

- 3: Es la observación correspondiente a la mujer universitaria en condición económicamente no activa, no es jefa de hogar, con 22 años de edad, en la cual han transcurrido 13 años de escolaridad, que vive en una casa con siete miembros de los cuales en cuatro tienen condición económicamente activa y es madre de un hijo, que no están en condición económicamente activa. Su probabilidad estimada (valor predicho) es de 0,97 y el residuo correspondiente es de -5,45.

- 4: Es la observación correspondiente a la mujer universitaria en condición económicamente no activa, no es jefa de hogar, con 62 años de edad, en la cual han transcurrido 14 años de escolaridad, que vive en una casa con cuatro miembros de los cuales en tres tienen condición económicamente activa y es madre de 5 hijos, en donde sólo 2 están en condición económicamente activa. Su probabilidad estimada (valor predicho) es de 0,97 y el residuo correspondiente es de -6,15.

- 5: Es la observación correspondiente a la mujer universitaria en condición económicamente activa, no es jefa de hogar, con 43 años de edad, en la cual han transcurrido 16 años de escolaridad, que vive en una casa con cuatro miembros de los cuales en dos tienen condición económicamente activa y es madre de 2 hijos, en donde sólo uno está en condición económicamente activa. Su probabilidad estimada (valor predicho) es de 0,98 y el residuo correspondiente es de -7,02.

- 6: Es la observación correspondiente a la mujer universitaria en condición económicamente no activa, no es jefa de hogar, con 27 años de edad, en la cual han transcurrido 15 años de escolaridad, que vive en una casa con cuatro miembros de los cuales en doss tienen condición económicamente activa y es madre de tres hijos, que no están en condición económicamente activa. Su probabilidad estimada (valor predicho) es de 0,98 y el residuo correspondiente es de -7,91.

- 7: Es la observación correspondiente a la mujer universitaria en condición económicamente no activa, jefa de hogar, con 48 años de edad, en la cual han transcurrido 18 años de escolaridad, que vive en una casa con cinco miembros de los cuales en tres tienen condición económicamente activa y es madre de 1 hijo, que no están en condición económicamente activa. Su probabilidad estimada (valor predicho) es de 0,99 y el residuo correspondiente es de -8,63.

- 8: Es la observación correspondiente a la mujer universitaria en condición económicamente no activa, no es jefa de hogar, con 27 años de edad, en la cual han

transcurrido 14 años de escolaridad, que vive en una casa con cuatro miembros de los cuales en tres tienen condición económicamente activa y es madre de 2 hijos, que no están en condición económicamente activa. Su probabilidad estimada (valor predicho) es de 0,98 y el residuo correspondiente es de -9,74.

Continuando con la evaluación sobre la eficiencia del modelo ajustado, se utilizó el cuadro de clasificación y el gráfico de ROC en el cual se emplea la siguiente tabla de clasificación y la curva ROC ("receiver operating characteristic").

Modelo reducido: Regresión Logística

Cuadro de Clasificación

CLASIFICIFICACIÓN		**Condición económica**		
		Activas	**Inactivas**	**TOTAL**
Condición económica	**Activas**	602	122	724
	Inactivas	78	237	315
	TOTAL	**680**	**359**	**1.039**

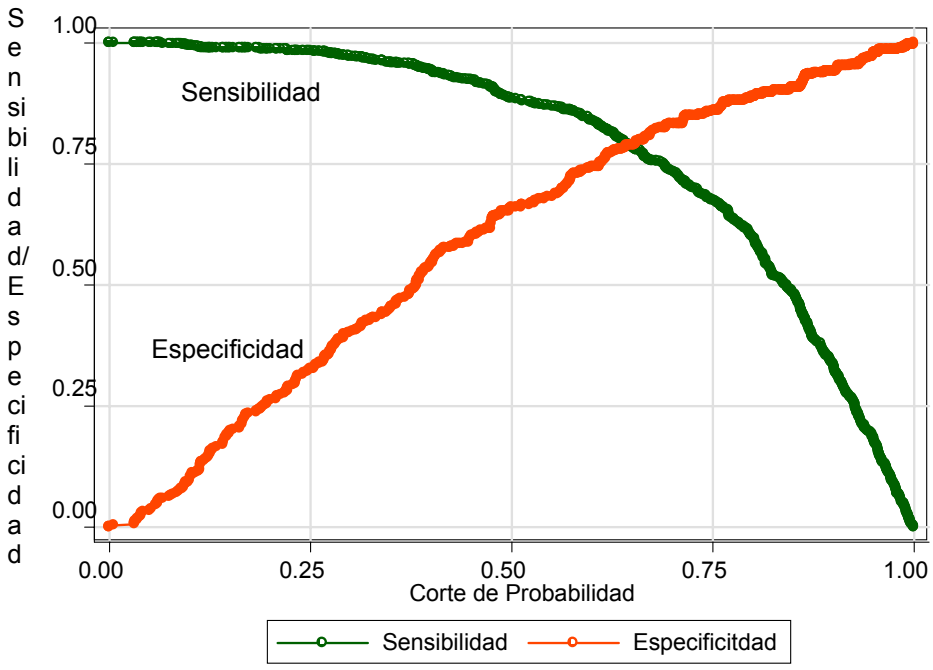
Sensibilidad	Pr(+| D)	88,53%
Especificidad	Pr(-|~D)	66,02%
Valor predictivo positivo	Pr(D| +)	83,15%
Valor predictivo negativo	Pr(~D| -)	75,24%
Clasificación correcta		**80,75%**

Los resultados de este análisis muestran que el modelo ajustado clasifica correctamente el 80,75% de las observaciones, siendo un porcentaje altamente satisfactorio para esta investigación. La sensitividad alcanzada es muy alta correspondiendo a un 88,53% del total de casos de mujeres universitarias en condición económicamente activa que son correctamente predichos por el modelo, y la especificidad es un poco más baja con un 66.02% de los casos, es decir del total de las mujeres universitarias en condición económicamente no activa son clasificados correctamente un 66,02%. Entre la sensibilidad y la especificidad existe un corte de probabilidad que es el cruce entre ambas curvas que es

aproximadamente de un 76%. (Ver gráfico 4.15) Es importante aclarar que la probabilidad de corte utilizada en este modelo para clasificar las observaciones como activas y no activas es de 0.5. Por lo que estos valores son bastantes satisfactorios para concluir que el modelo ajusta bien.

Gráfico 4.15

Modelo reducido: Regresión logística
Corte de Probabilidad

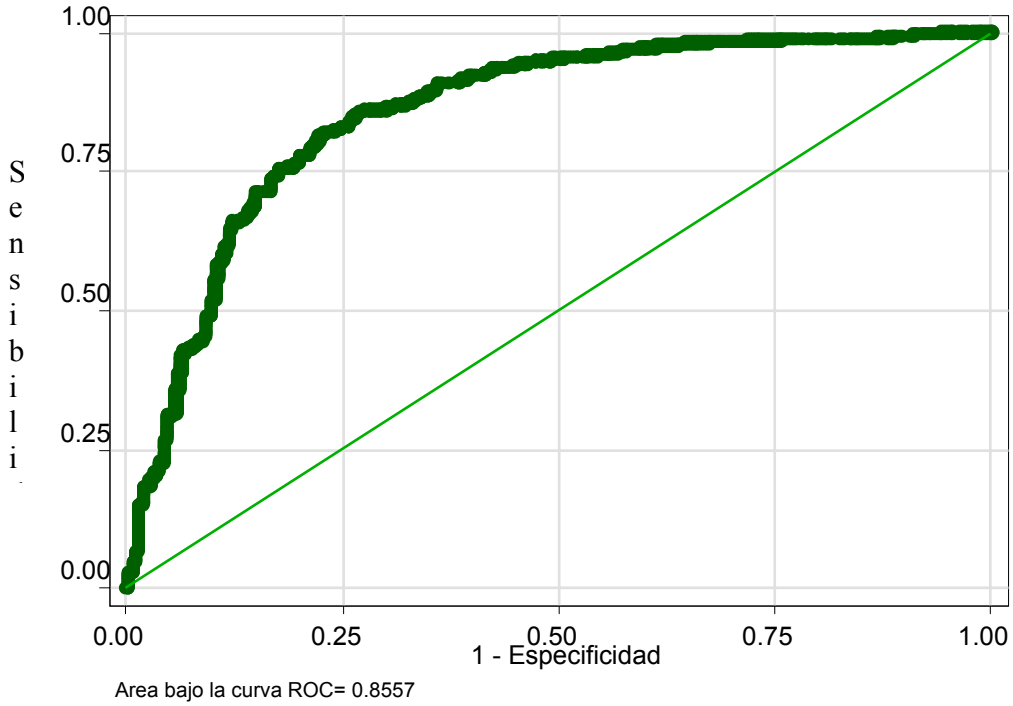


Por otro lado, también se concluye que del total de valores positivos, este modelo clasifica al 83,15%, es decir de un total de 724 mujeres universitarias activas, 602 lo eran certeramente, mientras que los valores negativos (condición económicamente inactiva) los clasifica en un 75,24%, de manera que para un total de 315 mujeres universitarias en condición económicamente inactiva, 237 son clasificadas correctamente. Demostrando que tanto estas conclusiones como las anteriores nos permiten determinar que el modelo se ajusta muy bien a los datos.

El gráfico ROC muestra que el área bajo la curva es muy alta (0,8557), representando lo efectivo que ajusta el modelo, es decir el nivel predicativo del modelo se ubica a la mitad entre un modelo sin ningún poder predicativo y uno perfecto. (Ver Gráfico 4.16)

Gráfico 4.16

Modelo reducido: Regresión logística
Gráfico ROC
(Característica de funcionamiento del receptor)



Hasta este punto se finaliza con la bondad de ajuste, de modo que se continuó con su validación por medio de una segunda submuestras con la misma cantidad de observaciones (1,039) que representa el 50% del total de los datos. Ambas muestras se seleccionaron aleatoriamente y fueron definidas al inicio de la investigación. De esta

manera la segunda base de datos muestra que el 64,1% de las mujeres universitarias están en condición económicamente activa, es decir 666 mujeres del total cumplen esta condición. Mientras que el restante 35,9% no se han incorporada a la PEA. Cabe destacar que esta tendencia es muy similar a los datos que se utilizaron con la primera muestra para ajustar el modelo en la cual el 65,45% representa a las mujeres universitarias en la misma condición económica, con una diferencia mínimo siendo mayor la base de datos utilizada para ajustar el modelo.

De esta manera se procederá a realizar una comparación del valor ajustado con el modelo y el valor real en la muestra, es decir a los datos de la segunda submuestra se van a validar con la fórmula que se señaló en el apartado anterior, sustituyendo los valores de las variables independientes para obtener el resultado de la variable Y (variable respuesta) que en este caso está asignada por la condición económicamente activa de las mujeres universitarias.

Según la siguiente tabla de clasificación, se puede observar que los resultados de este análisis muestran que el modelo ajustado clasifica correctamente el 79,21% de las observaciones, siendo un porcentaje altamente satisfactorio para esta investigación. La sensitividad alcanzada es muy alta correspondiendo a un 80,65% de casos reales de mujeres universitarias en condición económicamente activa que son correctamente predichos por el modelo, y la especificidad es un poco más baja con un 75,73% de los casos, representado el porcentaje de mujeres universitarias en condición económicamente inactiva clasificadas correctamentes. Por otro lado del total de mujeres universitarias incorporadas a la PEA, este modelo clasifica al 88,89% correctamente, es decir de un total de 666 mujeres universitarias activas, 592 lo eran certeramente, mientras son correctamente clasificadas el 61,93% de las mujeres universitarias en condición económicamente inactiva del total de estos casos, en otras palabras, para un total de 373 mujeres universitarias en condición económicamente inactiva, 231 son clasificadas correctamente. Demostrando que tanto

estas conclusiones como las anteriores nos permiten determinar que el modelo se ajusta muy bien a los datos.

Tabla de clasificación para validación del ajuste del modelo con la segunda muestra de datos

		Condición económica			Valor predictivo
		Activa	**Inactiva**	**Total**	
Condición económica	Activa	592	74	666	88,89%
	Inactiva	142	231	373	61,93%
	Total	**734**	**305**	**1.039**	
	Clasificación correcta				**79,21%**

Después de obtener estos resultados, se realizó la prueba de multicolinealidad por medio de regresiones auxiliares, de manera que no se observan tolerancias menores de 0.1 y FIV mayores de 10, concluyendo que bajo estas condiciones ninguna de las covariables dentro del modelo presenta una colinealidad significativa, mostrando una vez más que el modelo ajusta bien a los datos dentro de la investigación en estudio.

Detección de multicolinealidad mediante regresiones auxiliares

Variable dependiente	R2	Tolerancia	Factor de inflación de la varianza (FIV)
Relación de parentesco	0,4324	0,2454	1,7618
Edad	0,2679	0,1961	1,3659
Años de escolaridad	0,2284	0,1762	1,2960
Cantidad de activos en el hogar	0,6502	0,2274	2,8588
Cantidad de hijos	0,8024	0,1586	5,0607
Cantidad de hijos en la PEA	0,6931	0,2127	3,2584
Cantidad de miembros	0,2042	0,1625	1,2566

4.4.1.3 Desarrollo del modelo con el mejor ajuste de los datos

En el análisis de correlación lineal que se muestra en el Anexo Cuadro A 4.9, se puede apreciar que los valores observados en su mayoría son bajos, lo que significa que no existen altas asociaciones entre las variables en estudio que evidencien no ser consideradas dentro del análisis que se quiere validar, a excepción de cinco relaciones entre variables correspondiente a datos del cónyuge, las cuales muestran una correlación positiva fuerte, es decir, valores altos en este indicador, reflejando que sí existe prueba para no ser incluidas dentro del análisis, con lo que se evita la redundancia en la información. Las asociaciones que no se incluirá en el análisis son las siguientes:

- Años de escolaridad del cónyuge con la Edad del cónyuge, donde la asociación fue de un 0,93.
- Años de escolaridad del cónyuge con el Nivel de instrucción del cónyuge, donde la asociación fue de un 0,84.
- Título del cónyuge con los Años de escolaridad del cónyuge, donde la asociación fue de un 0,84.
- Condición económicamente activa del cónyuge con la Edad del cónyuge, donde la asociación fue de un 0,91.
- Condición económicamente activa del cónyuge con los Años de Escolaridad del cónyuge, donde la asociación fue de un 0,90.

Con el análisis anterior, se puede concretar que las variables independientes que se utilizarán para el estudio serán en total veintiuno, descartando “anescony”, “edcony”, “actcony”, “titcony” y “nivicony”. De esta manera la matriz de correlación con las variables restantes se muestra en el siguiente cuadro, en la cual se verifica que sus montos sean bajos, indicando que no existe una asociación entre variables.

Matriz de correlación con las variables que resultaron adecuadas en el modelo

	conact	*zona*	*aponse*	*relpar*	*edad*	*anoes*	*titul*	*acthog*	*hij*	*hijact*	*estciv*	*lningt*	*miemb*	*carte*	*cmedamb*	*ceconom*	*cfimaes*	*csalud*	*csocial*	*cderec*
conact	1,00																			
zona	0,05	1,00																		
aponse	-0,06	-0,05	1,00																	
relpar	0,09	0,10	-0,17	1,00																
edad	-0,06	0,12	-0,05	0,33	1,00															
anoes	0,26	0,09	-0,09	0,10	0,25	1,00														
titul	0,21	0,12	-0,17	0,18	0,34	0,61	1,00													
acthog	0,38	-0,05	0,28	-0,26	-0,24	0,00	-0,06	1,00												
hij	-0,02	-0,13	0,33	-0,17	-0,16	-0,10	-0,17	0,42	1,00											
hijact	0,19	-0,03	0,29	-0,11	-0,13	-0,02	-0,06	0,69	0,48	1,00										
estciv	0,00	-0,01	-0,15	-0,28	0,28	0,15	0,17	-0,04	-0,16	-0,34	1,00									
lningt	-0,02	-0,02	0,00	0,00	0,01	-0,03	0,08	-0,03	0,01	-0,03	0,04	1,00								
miemb	-0,06	-0,08	0,20	-0,17	-0,11	-0,02	-0,06	0,19	0,36	0,20	-0,05	-0,07	1,00							
carte	-0,05	-0,06	-0,02	-0,06	0,02	0,01	0,00	-0,02	-0,06	-0,04	0,07	0,00	-0,02	1,00						
cmedamb	0,00	-0,01	-0,02	-0,02	-0,04	-0,01	-0,02	0,02	0,02	-0,01	0,02	-0,01	0,00	-0,01	1,00					
ceconom	-0,03	0,04	0,03	0,03	0,00	-0,01	-0,02	-0,02	0,02	0,00	-0,04	0,03	0,00	-0,07	-0,08	1,00				
cfimaes	0,01	-0,06	-0,04	-0,01	0,02	0,08	0,03	-0,02	-0,01	-0,01	0,02	-0,02	0,00	-0,01	-0,01	-0,07	1,00			
csalud	0,00	-0,02	0,04	0,02	-0,03	-0,03	-0,04	0,03	0,07	0,05	-0,04	0,04	0,01	-0,04	-0,04	-0,22	-0,04	1,00		
csocial	-0,04	0,01	0,02	0,02	0,03	0,03	0,03	-0,07	-0,05	-0,03	-0,01	-0,06	-0,03	-0,04	-0,04	-0,23	-0,04	-0,12	1,00	
cderec	-0,01	-0,05	-0,02	-0,04	-0,03	-0,03	0,00	0,01	-0,02	-0,01	0,00	-0,07	-0,02	-0,02	-0,02	-0,09	-0,02	-0,05	-0,05	1,00
cdocen	0,07	0,03	-0,03	-0,03	0,06	0,05	0,05	0,03	0,00	0,00	0,07	0,03	0,05	-0,07	-0,08	-0,43	-0,07	-0,23	-0,23	-0,10
cingen	0,01	0,04	-0,08	0,03	-0,02	0,02	0,06	-0,03	-0,03	-0,02	0,00	-0,04	-0,03	-0,03	-0,03	-0,18	-0,03	-0,10	-0,10	-0,04

Una vez que se tiene el modelo inicial, se recurrió a su disminución de variables independientes hasta que el modelo más reducido siga explicando los datos. Para ello se utilizó al método de eliminación "hacia atrás", o a la selección de variables por mejores subconjuntos de covariables. De esta manera, se realizó el análisis con un modelo completo que muestra las veintiún variables independientes restantes. De este primer ajuste del modelo se observó que algunas de los coeficientes tienen valores p superiores al 5%, lo que permite excluir las variables con el fin de lograr un modelo que concuerde con los objetivos del estudio, ya que estas variables explicativas no originan una cantidad de información significativa al modelo. Por esa circunstancia no fueron incluidas las siguientes seis variables: “zona”, “aponse”, “titul”, “estciv”, “lningt”, “carte”, “cmedamb”, “ceconom”, “cfimaes”, “csalud”, “csocial”, “cderec”, “docen” y “cingen”. (Ver Anexo Cuadro A 4.10)

Estas variables que fueron desechadas están muy relacionadas en su mayoría con las condiciones de su hogar, es decir la zona de residencia, la cantidad de aposentos que tiene la casa de habitación y el ingreso total del hogar, además incluye aspectos personales de la mujer universitaria (titulo obtenido y estado civil) y la carrera universitaria que realizó. De manera que para continuar con el modelo reducido (siete variables), se analizarán tres variables que son características propias de las mujeres universitarias como lo son relación de parentesco, la edad en años cumplidos y los años de escolaridad, mientras que las restantes corresponden a situaciones del hogar: la cantidad de miembros en condición económicamente activa, el total de miembros, la cantidad de hijos y la cantidad de hijos en condición económicamente activa. El modelo debe ser aquél más reducido que explique los datos (principio de parsimonia), y que además sea clínicamente congruente e interpretable. Se debe tener en cuenta que un mayor número de variables en el modelo implicará mayores errores estándares.

Para construir el modelo que se desea desarrollar es necesario buscar minimizar el número de parámetros, de tal manera se continuó con la búsqueda de las variables que

mejor ajustan el modelo, utilizando un segundo modelo reducido que contenga solamente las variables que fueron significativas para sus coeficientes, de tal manera se mostrará a continuación:

$\mathbf{Z} = \beta_0 + \beta_1 *$ relpar $+ \beta_2 *$ edad $+ \beta_3 *$ anoes $+ \beta_4 *$ acthog $+ \beta_5 *$ hij $+ \beta_6 *$ hijact $+ \beta_7 *$ miemb

Remplazando los valores de los coeficientes podemos observar que el modelo de mejor ajuste, expresado en forma lineal (logit) corresponde a:

$\mathbf{Z} =$ -6,4274 + 1,1215* relpar – 0,0177* edad + 0,4410* anoes + 1,9211* acthog - 0,19635* hij - 0,3785* hijact – 0,6159* miemb

Modelo reducido: Regresión logística

Análisis de variancia

Estimados Logit	Número de obs =	1039
	LR chi2(7) =	397.16
	Prob > chi2 =	0.0000
Log verosimilitud = -471.19201	Pseudo R2 =	0.5965

conact	Coef.	Desv. Est.	z	P>z	[95% Conf.	Interval]
relpar	1.215183	.2449121	4.96	0.000	.7351637	1.695201
edad	-.0177421	.0065792	-2.70	0.007	-.030637	-.0048471
anoes	.4409987	.0553212	7.97	0.000	.3325712	.5494262
acthog	1.921149	.151011	12.72	0.000	1.625173	2.217125
hij	.1963485	.1083713	1.81	0.070	-.0160553	.4087522
hijact	-.3785737	.1363711	-2.78	0.006	-.6458561	-.1112913
miemb	-.61587	.1021784	-6.03	0.000	-.816136	-.415604
_cons	-6.427418	.811707	-7.92	0.000	-8.018335	-4.836502

Los valores del modelo obtenido con las variables que resultaron significativas es muy similar para los estadísticos que miden la bondad de ajuste. Usualmente el Pseudo R^2 es muy similar en ambas salidas de los modelos anteriormente señalados. Para este último modelo reducido podemos señalar que el Pseudos R^2 es de aproximadamente un 60%, de manera que el 59.65% de la variabilidad presente en la probabilidad de que una mujer universitaria sea económicamente activa está explicada por el modelo ajustado. Por otro lado se comparó los coeficientes de las variables restantes con las del modelo completo y se verificó que existe una mínima variación, de modo que esto reafirma que ninguna de las variables excluidas era importante para proporcionar el ajuste necesario del efecto de la variable que permanece en el modelo. Además la importancia de cada variable incluida en el modelo se verificó examinando la significancia de la razón de verosimilitud, lo cual involucra estimar el modelo con cada variable eliminada a la vez y verificar el cambio en el logaritmo de la verosimilitud, que para esta investigación, el logaritmo de la verosimilitud se ve afectado en pequeña escala, es decir, el cambio es mínimo, por lo que según ese criterio se puede concluir que el modelo es adecuado. Claro está, que es necesario utilizar otras medidas estadísticas para evaluar la bondad y la influencia de valores extremos en el ajuste.

Los resultados del modelo incluyen la desviación estándar de cada uno de estos coeficientes, por lo que se puede apreciar que la variable referente a la cantidad de miembros en condición económicamente activa es la que mantiene una desviación estándar superior al resto de las variables.

Cabe destacar que dentro de este mismo análisis no se encuentran casos en que el intervalo incluye el valor 0, por lo que no existe ningún indicador de que la influencia de esa variable es nula o casi nula en la predicción de la probabilidad de condición activa, por lo que si es significativa.

En el anterior cuadro, se puede observar que el valor del Chi2 (valor de la devianza respecto al modelo nulo), con 8 grados de libertad, resulta ser estadísticamente significativo

(valor “p” prácticamente nulo), lo cual permite concluir que las variables independientes incorporadas aportan a la explicación de la condición económicamente activa de las mujeres universitarias en estudio. A su vez con este análisis se puede expresar el modelo en forma lineal, utilizando el comando logit (transformación logito) en la cual se identifican los coeficientes del modelo.

Modelo reducido: Regresión Logística
Razones de ventaja (Odds Ratio)

Regresión Logística	Número de obs =	1,039
	LR chi2(7) =	397.16
	Prob > chi2 =	0.0000
Log likelihood = -471.19201	Pseudo R2 =	0.5965

conact	Razones Ventaja	Desviacion Estándar	z	P>z	[95% Conf.	Interval]
relpar	3.370909	.8255765	4.96	0.000	2.085823	5.447743
edad	.9824144	.0064635	-2.70	0.007	.9698276	.9951646
anoes	1.554259	.0859834	7.97	0.000	1.394549	1.732259
acthog	6.828801	1.031224	12.72	0.000	5.079297	9.180901
hij	1.216951	.1318825	1.81	0.070	.9840729	1.504939
hijact	.6848375	.093392	-2.78	0.006	.5242136	.8946781
miemb	.5401707	.0551938	-6.03	0.000	.4421368	.6599416

A partir lo anteriormente señalado, se puede notar que la significancia de la devianza, con respecto al modelo nulo, es sumamente alta reflejando con ello que las variables independientes consideradas en el ajuste son de relevancia para explicar en términos de la condición económicamente activa de las mujeres universitarias en estudio. Aquí mismo se

muestra los “odds ratios” (razones de ventaja), que nos permiten concluir con los siguientes resultados, si el resto de las variables se mantienen constantes:

- La Jefatura de Hogar está ligada al riesgo de que la mujer universitaria pertenezca a la población económicamente activa, resultando éste aproximadamente en un 337% mayor en las mujeres Jefas de Hogar.

- El coeficiente de la variable edad muestra que conforme aumenta los años de vida de la mujer universitaria, la probabilidad de pertenecer a la PEA se va reduciendo por cada cambio unitario.

- Los años de escolaridad de la mujer universitaria muestran que, ante cambios positivos en ellas, se modifica la razón de ventaja de pertenecer a la población económicamente activa. Esa razón de ventaja es de un 1.55, así que de acuerdo al aumento en los años de escolaridad de la mujer universitaria, el chance de tener una condición económicamente activa se incrementa por cada cambio unitario.

- Existe una asociación muy positiva entre la cantidad de miembros en el hogar en condición económicamente activa y el riesgo de incorporarse a la PEA. Es importante aclarar que esta asociación existe pero en términos de regresión logística de manera que conforme aumentan los miembros del hogar en condición económicamente activa, las probabilidades de la mujer universitaria de pertenecer a este grupo se van aumentando significativamente por cada cambio unitario.

- Conforme la cantidad de hijos aumenta, los chances de la mujer universitaria de pertenecer a la PEA se va incrementando por cada cambio unitario, es decir entre más hijos, más probabilidad de que la mujer universitaria tenga una condición económicamente activa.

- Si la cantidad de hijos en condición económicamente activa aumenta se tiene que las posibilidades de la mujer universitaria de pertenecer a la PEA se va reduciendo aproximadamente en un 32,52% por cada cambio unitario.

- La variable cantidad de miembros nos muestra que conforme aumenta, la probabilidad de la mujer universitaria de pertenecer a la PEA se va disminuyendo en aproximadamente un 46% por cada cambio unitario.

4.4.1.4 Conclusiones globales del ajuste del modelo

La tipología de la Encuesta de Hogares de Propósitos Múltiples del año 2004 adaptada a este estudio permitió definir inicialmente 26 variables independientes sobre la mujer universitaria en el país, basándose en 5 categorías, que corresponden a aspectos personales, demográficos, académicos, del cónyuge y del hogar. La variable respuesta dicotómica fue la condición económicamente activa de las mujeres en estudio.

El total de unidades de estudio (mujeres universitarias) que se obtuvo de la base de datos fue de un total de 2078 que se dividió en dos muestras seleccionadas aleatoriamente con la misma cantidad de personas (50%), con el objetivo de que la segunda muestra validara el modelo de mejor ajuste en los datos.

La técnica estadística que se utilizó fue la de regresión logística, ya que permite utilizar a la variable respuesta como dicotómica. El ajuste de modelo logístico para determinar los factores en el cual las mujeres universitarias se incorporan a la PEA, resultó muy aceptable y satisfactoria, al igual que la submuestra de validación y la evaluación de la multicolinealidad.

El proceso de ajuste del modelo indicó la posibilidad de omitir 19 variables y no incluirlas dentro del análisis, por dos motivos: en el primer caso porque su correlación era alta, lo que mostraba una asociación entre las variables y la segunda porque no eran lo suficientemente significativas para ser evaluadas dentro de la investigación. De manera que las variables independientes incluidas dentro de este modelo reducido y que resultaron significativas son: la relación de parentesco, la edad en años cumplidos, los años de escolaridad, la cantidad de miembros en condición económicamente activa en el hogar, la

cantidad de hijos, la cantidad de hijos en condición económicamente activa y la cantidad de miembros en el hogar. Por lo que las características del entorno (hogar) y las personales afectaron si la mujer universitaria se incorpora a la población económicamente activa.

El ajuste del modelo resultó bastante satisfactorio, denotando por un mejoramiento en la explicación de la variabilidad de la condición económicamente activa con relación al modelo empleado correspondiendo a un Pseudos R2 del 59,65%. La aplicación del modelo ajustado a una muestra de casos adicionales permitió confirmar el poder predictivo del modelo ajustado.

Como resultados importantes se tiene que las variables edad, cantidad de hijos en condición económicamente activa y la cantidad de miembros, mostraron que conforme avanza esta variable, la probabilidad de la mujer universitaria a pertenecer a la PEA se va reduciendo por cada cambio unitario, en la cual la restantes cuatro variables (relación de parentesco, cantidad de años de escolaridad, cantidad de miembros activos en el hogar, y la cantidad de hijos) muestran un comportamiento inverso, es decir conforme estos aumentan, el chance de que la mujer universitaria tenga una condición económicamente activa también se incrementará por cada cambio unitario.

El análisis de la bondad del ajuste utilizado es el adecuado y no existe evidencia estadística para desechar el modelo. Este modelo permite la clasificación correcta del 80,75% de los casos, valor bastante alto para el tipo de información que se analiza. La sensitividad es bastante elevada siendo un 88,53%, y la especificidad moderadamente alta con un 66,02%.

Como se señaló en la investigación, las brechas en la participación de hombres y mujeres en los distintos niveles de educación del país que favorecerían a los hombres hace varios años, no sólo se han venido reduciendo a través del tiempo, sino que han llevado a la mujer a mostrar tasas de cobertura en educación universitaria, superiores a las de los varones. A nivel universitario se gradúan más mujeres que hombres y aunque por muchos

años éstas han obtenido, en términos generales, un grado académico inferior al de los varones, se empieza a notar mayor equidad en este aspecto.

Un factor que parece tener peso en la desigualdad entre hombres y mujeres es el acceso a las diversas disciplinas y la asignación de roles que se continúa dando en nuestra sociedad: varones con rol de proveedores económicos y mujeres encargadas del cuidado del hogar. Lo analizamos como una posibilidad del por qué mujeres con niveles superiores en educación no se incorporan a la PEA, siendo muestra de una asignación social de roles fuertemente estructurados en la socialización.

4.5 REFLEXIONES FINALES

En nuestro país, la situación actual de la incorporación de las mujeres a la población económicamente activa ha sido placentera en el transcurso de los años, caso similar que sucede en otros países, pero aún se continua con la tendencia de que las mujeres en la PEA resultan en menor proporción que la población masculina, independientemente si tienen mayor nivel de educación, por lo que el 91.8% de los hombres universitarios tienen condición económicamente activa mientras que las mujeres bajo estas mismas condiciones sólo representa un 74.6%. [INEC, 2000]

Los censos de población de 1973, 1984 y 2000 muestran una distribución según nivel educativo muy similar entre la población femenina y la masculina. En ambos grupos se han venido incrementando los porcentajes en niveles desde primaria completa hasta educación superior. De esta manera se puede notar que para el Censo del año 2000 la población femenina en el nivel de educación primaria corresponde a un 49,63% del total de la población, mientras que en Secundaria y Universitaria o Parauniversitaria poseen la mayor participación de la población con un 51,61% y 51,58% respectivamente.

En la actualidad existe una población en la que a pesar de obtener el título universitario, no se incorpora al mercado laboral, correspondiendo a un 25.5% del total de mujeres en este nivel superior. Además de que el nivel de Instrucción Superior es el que mantiene una tasa neta de población económicamente activa mayor que la otras con un 68%, en comparación de un 33% en ningún nivel, 41.2% en Primaria y 45% en Secundaria.

De esta manera se desarrolló esta investigación, en la cual su principal objetivo es conocer la importancia de algunos factores determinantes en las mujeres de grado académico superior (universitarias), para formar parte de la población económicamente activa (PEA) del país, utilizando la información disponible en la base de datos de la Encuesta de Hogares de Propósitos Multiples de julio 2004, suministrada por el INEC, en la

cual inicialmente se utilizaron 26 variables independientes. Se ajustó un modelo de regresión logística multivariado, el objetivo primordial que resuelve esta técnica es el de modelar cómo influye en la probabilidad de pertenecer o no a la Población Económicamente Activa (PEA), habitualmente dicotómico, con respecto a las otras variables en estudio que no son necesariamente dicotómicas.

La base de datos en estudio muestra que las mujeres universitarias que pertenecen a la población económicamente activa representan el 65.45% del total, mientras que el 34.55% no están incorporadas a la PEA. El uso de varias técnicas estadísticas para medir el ajuste del modelo resultaron satisfactorias para los datos, dando como consecuencia las siguientes deducciones:

- El valor del pseudo R^2 indica que el modelo explica un 59,65% más que el modelo nulo.
- Prueba Pearson Chi-cuadrado de Bondad de Ajuste, en la cual la estadística es altamente significativa, su probabilidad es 0.
- El valor obtenido para la Chi2 Hosmer-Lemeshow implica que no existe evidencia estadística para rechazar la hipótesis nula de que el ajuste del modelo es bueno.
- En el análisis de los residuos se utilizó un gráfico que muestra a la mayoría de los puntos en una franja alrededor de 0 a excepción de pocos
- En la Tabla de Clasificación se muestra que los resultados de este análisis muestran que el modelo ajustado clasifica correctamente el 80,75% de las observaciones, siendo un porcentaje altamente satisfactorio para esta investigación. La sensitividad alcanzadas es muy alta correspondiendo a un 88,53% de casos reales de mujeres universitarias en condición económicamente activas que son correctamente predichos por el modelo y la especificidad es un poco más baja con un 66,02% de los casos,
- El gráfico ROC muestra que el área bajo la curva es muy alta (0.8557), representando lo efectivo que ajusta el modelo.

- La validación del modelo por medio de una segunda muestra, señala que el modelo ajustado clasifica correctamente el 79,21% de las observaciones, siendo un porcentaje altamente satisfactorio, a su vez, la sensitividad alcanzada es muy alta correspondiendo a un 88,89% de casos reales de mujeres universitarias en condición económicamente activas que son correctamente predichos por el modelo del total real de activas, la especificidad es un poco más baja con un 61,93% de los casos, representado el porcentaje de mujeres universitarias en condición económicamente no activa en el total de los casos. Por otro lado del total de valores positivos (activa), este modelo clasifica al 88,89%, mientras que los valores negativos (inactiva) los clasifica en un 61,93. Demostrando que tanto estas conclusiones como las anteriores nos permiten determinar que el modelo se ajusta muy bien a los datos.
- Con respecto a la multicolinealidad se observa que ninguna de las covariables dentro del modelo presentan un colinealidad significativa.

Los resultados más importantes obtenidos de los "Odds Ratios", detallan a continuación, siempre y cuando el resto de las variables se mantenga constante:

- La relación de parentesco, los años de escolaridad, la cantidad de miembros del hogar en condición económicamente activa y la cantidad de hijos hace que existan cambios positivos en las mujeres universitarias, mostrando que conforme aumentan estas variables, la probabilidad de que ellas pertenecezca a la PEA se va incrementando por cada cambio unitario.

Mientras que la edad, la cantidad de hijos activos en el hogar y la cantidad de miembros, hace que de acuerdo al aumento en esas variables, el chance que las mujeres universitarias tengan una condición económicamente activa se va reduciendo por cada cambio unitario.

Es importante señalar que la experiencia desarrollada en esta investigación es de gran importancia para los futuros estudios sobre la mujer, de esta manera en el transcurso

del desarrollo de esta tesis se pudo observar que no se tiene conocimiento de las existencias de estudios o investigaciones previas sobre el tema en específico tratados mediante regresión logística, de modo que se pretende incentivar con base a los resultados anteriormente señalados. Por otra parte, considero que sería importante realizar o evaluar este tipo de temas con otras variables adicionales que puedan ser analizadas como posibles factores para que la mujer universitarias se incorpore o no a la PEA, como lo son el grado de endeudamiento del individuo, el aspecto cultural del Jefe de Hogar, grado de satisfacción con respecto a su condición de actividad, etc, esto porque la EHPM no es específica para estudiar el fenómeno del la incorporación al PEA de la mujer universitaria. Las técnicas utilizadas fueron muy consistentes en los resultados, demostrando que es confiable el uso de las mismas.

BIBLIOGRAFÍA

Abraira, Victor, 1996. **Métodos multivariantes en bioestadística.** Editorial Centro de Estudios Ramón Areces. Madrid 1996.

Albán López Mary, 1991. **Situación de la Mujer Costarricense y su integración al desarrollo económico y social del país**. Universidad de Costa Rica, San José, Costa Rica.

Arroyo Roxana, 1999. **El estudio protege equitativamente y eficazmente los derechos de los grupos étnicos de género.** Estado de la Nación, San José, Costa Rica.

Asamblea Legislativa de Costa Rica, 2005. **Período Legislativo 2006-2010**. Departamento de Relaciones Públicas, Prensa y Protocolo. Costa Rica. http://www.asamblea.go.cr/diputado/dip_sj2006.htm

Barquero B, Jorge A, 2004. **Población y Salud en Mesoamérica.** Revista. Volumen1, número 2, artículo 5. Centro Centroamericano de Población. Editorial Universidad de Costa Rica. San José, Costa Rica.

Brennes Isabel, 2003. **Los géneros de la Ecuación Superior Universitaria en Costa Rica**. OPES. San José, Costa Rica.

CEPAL, 2002 (Comisión Económica para América Latina y el Caribe). **América Latina y el Caribe: Indicadores seleccionados con una perspectiva de género.** Boletín Demográfico No70. CELADE, División de Población y Unidad Mujer y Desarrollo. Santiago, Chile.

Chavarría C Freddy, Molina C Isabel y Zamora H Juan Carlos, 1997. **Diferenciales del desempleo, subempleo y del sexo del jefe de hogar utilizando la Metodología de Regresión Logística.** Universidad de Costa Rica, San José, Costa Rica.

Chinchilla Carlos, Lápiz Carlos y Segura Olman, 2006. **Agenda Universitaria hacia la equidad entre mujeres y hombres**. Instituto de Estudios de la Mujer. San José. Costa Rica.

FLACSO, 2000 (Facultad Latinoamericana de Ciencias Sociales). **Mujeres Latinoamericanas en cifras.** FLACSO - Facultad Latinoamericana de Ciencias Sociales, San José, Costa Rica.

FLACSO, 2004 (Facultad Latinoamericana de Ciencias Sociales). **Centroamérica en cifras.** FLACSO - Facultad Latinoamericana de Ciencias Sociales, San José, Costa Rica.

Gálvez Thelma, 2001. **Aspectos Económicos de la Equidad de Género.** Series CEPAL. Unidad Mujer y Desarrollo. No 35. Santiago, Chile.

Hernández Rodríguez Oscar, 1998. **Análisis Estadístico Multivariado**. Editorial Universidad de Costa Rica. San José, Costa Rica.

Hosmer David y Lemeshow Stanley, 1989. **Applied Logistic Regression.** John Wiley & Sons. Nueva York, Estados Unidos.

INEC, 1973 (Instituto Nacional de Estadística y Censos). **Censo poblacional 1973.** Instituto Nacional de Estadística y Censos, San José, Costa Rica.

INEC, 1984 (Instituto Nacional de Estadística y Censos). **Censo poblacional 1984.** Instituto Nacional de Estadística y Censos, San José, Costa Rica.

INEC, 2000 (Instituto Nacional de Estadística y Censos). **Censo poblacional 2000.** Instituto Nacional de Estadística y Censos, San José, Costa Rica.

INEC, 2004 (Instituto Nacional de Estadística y Censos). **Encuesta de Hogares de Propósitos Múltiples julio 2004. Principales Resultados.** Instituto Nacional de Estadística y Censos, San José, Costa Rica.

Mazzei Nogueir Claudia, 2006. **El trabajo feminino y las desigualdades en el mundo produtivo.** III Conferencia Internacional La obra de Carlos Marx y los desafíos del Siglo XXI.
http://www.nodo50.org/cubasigloXXI/congreso06/conf3_mazei.pdf

Molinero Luis M, 2003. **¿Qué es el método de estimación de máxima verosimilitud y cómo se interpreta?**.
http://www.seh-lelha.org/maxverosim.htm

Montgomery, E. A. Peck and G. G. Vining , 2002. **Introduction to linear regression analysis by D. C.** Editor Elsevier Science Publishers B. V. Hannover, Germany

Mujica Petit Javier, 2006. **Desafíos y estrategias frente a las políticas migratorias de la Unión Europea** . Programa de Derechos Humanos. Centro de Asesoría Laboral del Perú (CEDAL). Boletín #14. Perú.

Proyecto Estado de la Nación, 2002. **Aporte para el análisis de las brechas de equidad entre los géneros. Insumos para su medición**. San José, Costa Rica.

Quiros M.. Teresa, 1984. **La Mujer en Costa Rica y su participación política económica en el desarrollo del país**. FLACSO, avance No.51. San José, Costa Rica.

Ramírez Moreira Olman, 1995. **Factores Asociados con la desnutrición crónica en indígenas guatemaltecos 1995.** Universidad de Costa Rica, San José Costa Rica.

Ramírez Moreira Olman, 2000. **Factores determinantes de los arreglos de convivencia de la población adulta mayor.** Tesis de Maestría. Universidad de Costa Rica, San José, Costa Rica.

Stata Press, University Drive, 1995 **Getting Started with STATA for windows**. Collage Station, Texas, Estado Unidos.

Stata Corporation. Stata Statistical Software, 1997. **Realease 5.0. College Station, TX**: Stata Corporation. Chicago, Estados Unidos.

UCR, 1960-1961, 1970 (Universidad de Costa Rica). **Estadística Universitaria.** Editorial Universidad de Costa Rica. San José, Costa Rica.

http://163.178.80.12/catalogos/doctextcomp/opes/2003/informe_g%C3%A9nero_%20Resumen.pdf

ANEXOS

A. DEFINICIONES

a. Mercado laboral

Para entender el Mercado de Trabajo se debe estudiar los comportamientos de los oferentes y de los demandantes de trabajo frente a un conjunto de incentivos pecuniarios y no pecuniarios. Es decir, los empleadores y los empleados precisan sus acciones y reacciones frente a una serie de decisiones, que afectan su interés como agentes económicos. Estas decisiones se refieren al salario, precio, beneficio y condiciones relacionadas al empleo. Mientras los empleadores tratan de maximizar el beneficio, los empleados se empeñan en maximizar la remuneración que les permite asegurar la mayor satisfacción posible.

b. Población Económicamente Activa (PEA)

Personas de 12 años o más de edad que trabajaron al menos una hora en la semana de referencia o que, sin hacerlo, buscaron trabajo en las últimas cinco semanas. [EHPM, 2004]

c. Condición de actividad

Es la participación de cada persona de 12 años o más en las actividades económicas que se desarrollan en el país, en cualquiera de sus sectores (industria, comercio, agropecuario, servicios, etc.) Para determinar la condición de cada entrevistado se indaga su situación respecto a la tenencia o no de un trabajo o empelo, si buscó trabajo o estuvo inactivo en el período de referencia. [Baquero, 2004: 28]

d. Las variables “dummy”

A veces necesitamos incorporar al modelo de regresión logística variables independientes que no son numéricas sino categóricas. Ello consiste en generar $n - 1$

variables dicotómicas con valores cero y uno, siendo n el número de categorías de la variable original.

e. Variables del análisis para ajustar el modelo

Variable Dependiente

- **Condición económicamente activa (conact):** esta variable dicotómica (0 = Económicamente inactiva, 1 = Económicamente Activa) se utilizó para analizar el tema sobre todas las mujeres universitarias incorporadas al PEA.

Variables Independientes

Personales

- **Relación de parentesco con el jefe de familia (felpar):** es una variable dicotómica en donde 1 = Jefe de Hogar y 0 = No Jefe de Hogar, y representa la relación de parentesco que tienen las mujeres universitarias en estudio con el Jefe de Hogar.

- **Edad (edad):** es una variable continua que señala la edad en años cumplidos de las mujeres universitarias en estudio.

- **Estado civil (estciv):** es una variable dicotómica en donde muestra si 1 = Con cónyuge y 0 = sin cónyuge, de las mujeres universitarias en estudio.

Académicos

- **Cantidad de años de escolaridad (anoses):** es una variable continua que representa la cantidad de años que han realizado las mujeres universitarias en estudio en su último nivel educativo.

- **Título (titul):** es una variable dicotómica en donde corresponde 1 = Con Título y 0 = sin Titulo, de las mujeres universitarias en estudio.

- **Carrera en Arte (carte):** es una variable dicotómica en donde corresponde 1 = presencia y 0 = ausencia, de la carrera de las mujeres universitarias en estudio

- **Carrera en Medio Ambiente (cmedamb):** es una variable dicotómica en donde corresponde 1 = presencia y 0 = ausencia, de la carrera en las mujeres universitarias en estudio

- **Carrera en Ciencias Económicas (ceconom):** es una variable dicotómica en donde corresponde 1 = presencia y 0 = ausencia, de la carrera en las mujeres universitarias en estudio

- **Carrera Física, Matemática y estadística (cfimaes):** es una variable dicotómica en donde corresponde 1 = presencia y 0 = ausencia, de la carrera en las mujeres universitarias en estudio

- **Carrera en Ciencias de la Salud (csalud):** es una variable dicotómica en donde corresponde 1 = presencia y 0 = ausencia, de la carrera en las mujeres universitarias en estudio

- **Carrera en Ciencias Sociales (csocial):** es una variable dicotómica en donde corresponde 1 = presencia y 0 = ausencia, de la carrera en las mujeres universitarias en estudio

- **Carrera en Derecho (cderec):** es una variable dicotómica en donde corresponde 1 = presencia y 0 = ausencia, de la carrera en las mujeres universitarias en estudio

- **Carrera en Docencia (cdocen):** es una variable dicotómica en donde corresponde 1 = presencia y 0 = ausencia, de la carrera en las mujeres universitarias en estudio

- **Carrera en Ingenierías (cingen):** es una variable dicotómica en donde corresponde 1 = presencia y 0 = ausencia, de la carrera en las mujeres universitarias en estudio

Demográficos

- **Zona de residencia(zona):** esta es una variable dicotómica en donde 1 = Urbano y 0 = Rural, y representa la zona residencial de las mujeres universitarias en estudio.

- **Cantidad de aposentos en la casa de habitación (aponse):** esta es una variable continua que indica la cantidad de aposentos dentro de la casa de habitación de las mujeres universitarias en estudio.

Del hogar

- **Cantidad de activos en el hogar (acthog):** es una variable continua que indica la cantidad de miembros activos en cada hogar

- **Cantidad de hijos (hij):** es una variable continua que muestra la cantidad de hijos en el hogar

- **Cantidad de hijos en condición económicamente activa (hijact):** es una variable continua que indica la cantidad de hijos en condición activa en el hogar.

- **LN ingreso mensual total (lningt):** es una variable continua que muestra el logaritmo natural del ingreso total de hogar en forma mensual.

- **Cantidad de miembros (miemb):** es una variable continua que representa la cantidad de miembros en el hogar

Del cónyuge

- **Edad (edcony):** es una variable continua en la que muestra la edad en años cumplidos del cónyuge de las mujeres universitarias en estudio.

- **Nivel de instrucción (nivicony):** es una variable dicotómica que representa el nivel educativo del cónyuge de las mujeres universitarias en estudio. En donde 1 significa que es universitario y 2 no es universitario

- **Cantidad de años de escolaridad (anescony):** es una variable continua que muestra los años educativos que ha tenido el cónyuge de las mujeres universitarias en estudio.

- **Condición económicamente activa (actcony):** esta variable dicotómica (0 = No Activa, 1 = Activa) muestra la condición activa del cónyuge de las mujeres universitarias en estudio.

- **Título universitario (titulcony):** es una variable dicotómica en donde 1 = Con título universitario y 0 = sin título universitario, del cónyuge de las mujeres universitarias en estudio.

B. CUADROS

Cuadro A 4.2

COSTA RICA: Total de población con grado universitario Por sexo. 1973, 1984 Y 2000

	TOTAL	HOMBRES	MUJERES
Censo 1973	56,905	51.7%	48.3%
Censo 1984	141,483	52.2%	47.8%
Censo 2000	400,415	48.4%	51.6%

Fuente: Instituto Nacional de Estadísticas y Censos (INEC), Censos 1973, 1984 y 2000.

Cuadro A 4.3

MUNDIAL: Tasa Neta de la población económicamente activa, de 15 o más años, Según región. 2001

Región	Tasa (%)	Índice (1990 = 100)	Tasa masculina
Países en desarrollo	55.7	101	67
Países menos adelantados	64.2	99	74
Estados Arabes	32.7	117	41
Asia Oriental y el Pacífico	68.8	99	82
América Latina y el Caribe	42.2	109	52
Asia meridional	43.6	106	52
África subsahariana	62.2	99	73
Europa Central y Oriental y la CEI	57.5	99	81
OCDE	51.3	106	71
Países de la OCDE de ingresos altos	52.0	106	73
Desarrollo humano alto	50.7	106	70
Desarrollo humano medio	56.7	100	69
Desarrollo humano bajo	56.7	102	66
Ingresos altos	51.9	106	73
Ingresos medios	59.1	100	73
Ingresos bajos	51.9	103	62
Tota mundial	55.2	102	68

Fuente: Organización Internacional del Trabajo, cálculos basados en datos relativos a la población económicamente activa y la población, 2002.

Cuadro A 4.4

COSTA RICA: Distribución relativa de la población, Según nivel de instrucción, rama de actividad y tasa de participación, Por año y sexo. 1973, 1984 y 2000.

CARACTERISTICA	AÑO Y SEXO					
	1973		1984		2000	
	HOMBRES	MUJERES	HOMBRES	MUJERES	HOMBRES	MUJERES
Población	938.535	933.245	1.208.216	1.208.593	1.902.614	1.907.565
	50,1	49,9	50,0	50,0	49,9	50,1
Nivel de instrucción (población de 15 años y más)						
Sin instrucción	11,9	11,6	8,1	7,8	5,8	5,2
Primaria incompleta	41,1	42,0	26,6	27,7	18,6	18,9
Primaria completa	23,6	23,2	29,2	27,5	30,5	29,0
Secundaria incompleta	13,1	13,5	16,3	16,7	20,6	21,0
Secundaria completa	4,6	4,5	10,1	11,6	9,4	10,1
Superior	5,7	5,2	9,7	8,7	15,1	15,8
Tasa neta de participación en la fuerza de trabajo	78,4	18,6	74,5	20,7	69,2	27,0
Población ocupada por rama de actividad	434.154	108.176	579.940	166.920	922.768	378.776
Agricultura, caza, silvicultura y pesa	46,7	4,1	40,5	4,4	25,6	4,4
Industrias manufactureras	11,6	16,2	12,6	18,2	17,0	16,7
Comercio al por mayory menos	11,2	16,4	10,2	15,7	17,0	23,6
Servicios comunales, sociales y personales	12,2	69,2	15,5	49,9	17,1	46,3
Construcción	8,6	0,2	6,7	0,2	8,7	0,5
Otras	9,7	3,9	14,5	11,6	14,6	8,5

Fuente: Idem cuadro A 4.2.

Cuadro A 4.5

COSTA RICA: Distribución relativa de la población femenina entre 25 y 49 años Según nivel de educación. 1973, 1984 Y 2000

	1973	1984	2000
Total	27,489	67,680	206,541
NA / Sin instrucción	27.8%	24.7%	15.9%
Primaria	57.6%	50.1%	49.9%
Secundaria	11.6%	19.6%	23.3%
Univers / ParaUniver	2.9%	5.6%	10.8%

Fuente: Idem cuadro A 4.2.

Cuadro A 4.6

COSTA RICA: Distribución relativa de mujeres económicamente activas entre 25 y 29 años, según nivel de educación, por cantidad de hijos. 1984 Y 2000.

Nivel de educación	No tiene	1-2	3-4	5-6	Más de 7	Total activas
Total de activas	**51,891**	**118,364**	**78,724**	**19,007**	**6,688**	**274,674**
Ninguno	0	0.5%	0.6%	0.4%	0.4%	2.0%
Primaria	3.6%	11.5%	11.1%	4.1%	1.7%	32.0%
Secundaria	5.1%	14.4%	9.0%	1.7%	0.3%	30.6%
Univers / ParaUniver	9.9%	16.7%	8.0%	0.7%	0.1%	35.4%
Total	**18.9%**	**43.1%**	**28.7%**	**6.9%**	**2.4%**	**100.0%**

Fuente: Instituto Nacional de Estadísticas y Censos (INEC), Censo 2000.

Cuadro A 4.7

Codificación de la variable nueva "Años de escolaridad"

Nivel de Instrucción	Años de escolaridad
Ignorado	0
Ninguno	0
Ensenanza Especial	1
Preparatoria	0
Primaria Ignorada	1
Primaria: 1°	1
Primaria: 2°	2
Primaria: 3°	3
Primaria: 4°	4
Primaria: 5°	5
Primaria: 6°	6
Secundaria Ignorada	7
Secundaria: 7°	7
Secundaria: 8°	8
Secundaria: 9°	9
Secundaria: 10°	10
Secundaria: 11°	11

Nivel de Instrucción	Años de escolaridad
Secundaria Técnica: 9°	7
Secundaria Técnica: 8°	8
Secundaria Técnica: 9°	9
Secundaria Técnica: 10°	10
Secundaria Técnica: 11°	11
Secundaria Técnica: 12°	12
Parauniversitaria: 1er ano	12
Parauniversitaria: 2do ano	13
Parauniversitaria: 3er ano	14
Universitaria Ignorada	12
Universitaria: 1er ano	12
Universitaria: 2do ano	13
Universitaria: 3er ano	14
Universitaria: 4to ano	15
Universitaria: 5to ano	16
Universitaria: 6to ano	17
Universitaria: 7mo ano	18
Universitaria: 8vo ano	19

Cuadro A 4.8
Estadísticas descriptivas de las variables independientes cuantitativas

	Variables independientes cuantitativas								
Medición	**edad**	**anoes**	**aponse**	**acthog**	**hij**	**hijact**	**miemb**	**edcony**	**anescony**
Media	34,80	14,35	3,12	1,99	1,95	0,63	4,14	1,11	0,31
Mediana	32,00	14,00	3,00	2,00	2,00	0,00	4,00	0,00	0,00
Moda	22,00	14,00	3,00	2,00	2,00	0,00	4,00	0,00	0,00
Desviación estándar	13,21	1,78	0,94	1,06	1,41	0,92	3,37	6,73	1,97
Varianza de la muestra	174,62	3,18	0,88	1,13	1,98	0,84	11,36	45,36	3,88
Rango	82,00	8,00	8,00	8,00	10,00	5,00	98,00	54,00	17,00
Mínimo	17,00	12,00	1,00	1,00	0,00	0,00	1,00	1,00	1,00
Máximo	99,00	20,00	8,00	8,00	10,00	5,00	12,00	54,00	17,00

Cuadro A 4.9

Matriz de correlación con el total de variables

	conact	zona	aponse	relpar	edad	anoes	titul	edcony	nivicony	anescony	titcony	actcony	acthog	hij	hijact	estciv	lningt	miemb	carte	cmedamb	ceconom
conact	1,00																				
zona	0,05	1,00																			
aponse	-0,06	-0,05	1,00																		
relpar	0,09	0,10	-0,17	1,00																	
edad	-0,06	0,12	-0,05	0,33	1,00																
anoes	0,26	0,09	-0,09	0,10	0,25	1,00															
titul	0,21	0,12	-0,17	0,18	0,34	0,61	1,00														
edcony	0,03	0,01	-0,05	0,30	0,09	0,04	0,06	1,00													
nivicony	0,02	0,04	-0,05	0,22	0,05	0,06	0,05	0,66	1,00												
anescony	0,04	0,03	-0,05	0,30	0,07	0,05	0,04	0,93	0,84	1,00											
titcony	0,02	0,04	-0,05	0,22	0,05	0,06	0,05	0,66	1,00	0,84	1,00										
actcony	0,05	0,03	-0,04	0,32	0,07	0,03	0,04	0,91	0,62	0,90	0,62	1,00									
acthog	0,38	-0,05	0,28	-0,26	-0,24	0,00	-0,06	0,04	0,01	0,04	0,01	0,06	1,00								
hij	-0,02	-0,13	0,33	-0,17	-0,16	-0,10	-0,17	-0,04	-0,07	-0,06	-0,07	-0,06	0,42	1,00							
hijact	0,19	-0,03	0,29	-0,11	-0,13	-0,02	-0,06	-0,02	-0,04	-0,03	-0,04	-0,02	0,69	0,48	1,00						
estciv	0,00	-0,01	-0,15	-0,28	0,28	0,15	0,17	0,15	0,12	0,15	0,12	0,17	-0,04	-0,16	-0,34	1,00					
lningt	-0,02	-0,02	0,00	0,00	0,01	-0,03	0,08	-0,05	-0,05	-0,06	-0,05	-0,05	-0,03	0,01	-0,03	0,04	1,00				
miemb	-0,06	-0,08	0,20	-0,17	-0,11	-0,02	-0,06	-0,02	-0,03	-0,03	-0,03	-0,03	0,19	0,36	0,20	-0,05	-0,07	1,00			
carte	-0,05	-0,06	-0,02	-0,06	0,02	0,01	0,00	-0,02	-0,01	-0,02	-0,01	-0,02	-0,02	-0,06	-0,04	0,07	0,00	-0,02	1,00		
cmedamb	0,00	-0,01	-0,02	-0,02	-0,04	-0,01	-0,02	-0,02	-0,01	-0,02	-0,01	-0,02	0,02	0,02	-0,01	0,02	-0,01	0,00	-0,01	1,00	
ceconom	-0,03	0,04	0,03	0,03	0,00	-0,01	-0,02	-0,04	-0,01	-0,02	-0,01	-0,02	-0,02	0,02	0,00	-0,04	0,03	0,00	-0,07	-0,08	1,00
cfimaes	0,01	-0,06	-0,04	-0,01	0,02	0,08	0,03	0,04	-0,01	0,02	-0,01	-0,02	-0,02	-0,01	-0,01	0,02	-0,02	0,00	-0,01	-0,01	-0,07
csalud	0,00	-0,02	0,04	0,02	-0,03	-0,03	-0,04	0,00	0,02	0,02	0,02	0,02	0,03	0,07	0,05	-0,04	0,04	0,01	-0,04	-0,04	-0,22
csocial	-0,04	0,01	0,02	0,02	0,03	0,03	0,03	0,04	0,05	0,04	0,05	0,02	-0,07	-0,05	-0,03	-0,01	-0,06	-0,03	-0,04	-0,04	-0,23
cderec	-0,01	-0,05	-0,02	-0,04	-0,03	-0,03	0,00	0,02	-0,02	0,00	-0,02	0,02	0,01	-0,02	-0,01	0,00	-0,07	-0,02	-0,02	-0,02	-0,09
cdocen	0,07	0,03	-0,03	-0,03	0,06	0,05	0,05	0,02	0,01	0,00	0,01	0,01	0,03	0,00	0,00	0,07	0,03	0,05	-0,07	-0,08	-0,43
cingen	0,01	0,04	-0,08	0,03	-0,02	0,02	0,06	-0,02	-0,03	-0,03	-0,03	-0,04	-0,03	-0,03	-0,02	0,00	-0,04	-0,03	-0,03	-0,03	-0,18

Cuadro A 4.10

Modelo completo: Regresión logística

Logit estimados — Número de obs = 1,039

Prob > chi2 = 0.0000

Log verosimilitud = -474.36437 — Pseudo R2 = 0.4067

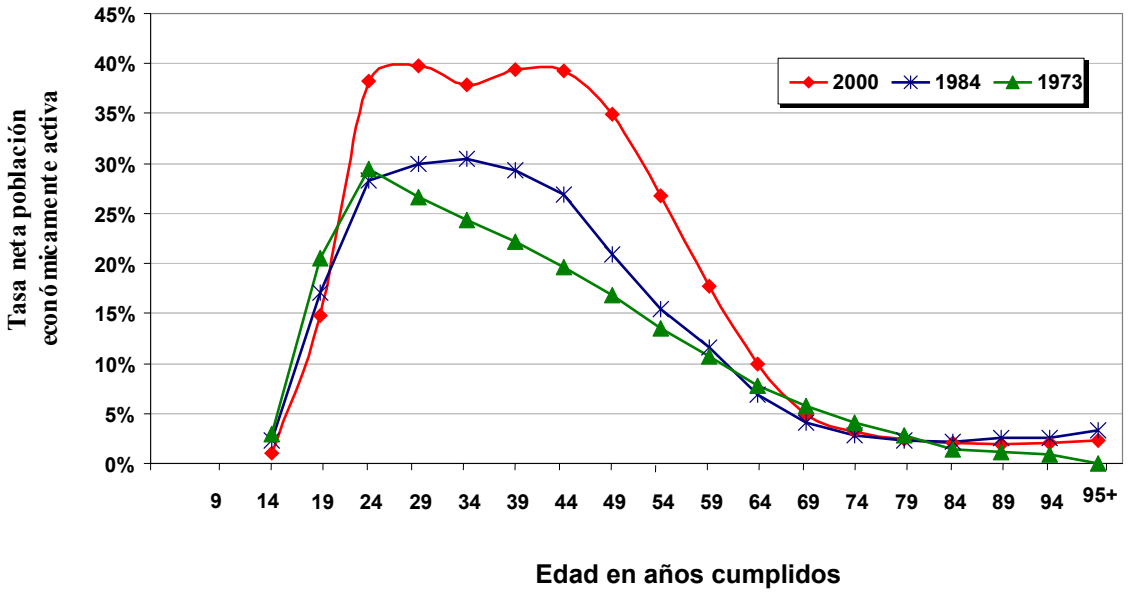
conact	**Coef.**	**Std. Err.**	**z**	**P>z**	**[95% Conf.**	**Interval]**
zona	.1199888	.169863	0.71	0.480	-.2129365	.452914
aponse	-.0971605	.0994013	-0.98	0.328	-.2919835	.0976625
relpar	1.182715	.2792669	4.24	0.000	.635362	1.730068
edad	-.0209769	.0075139	-2.79	0.005	-.0357039	-.0062499
anoes	.3569459	.0688734	5.18	0.000	.2219566	.4919352
titul	.3764593	.2333386	1.61	0.107	-.080876	.8337945
acthog	1.938307	.1578216	12.28	0.000	1.628982	2.247632
hij	.2294302	.1110195	2.07	0.039	.011836	.4470244
hijact	-.3667448	.1549524	-2.37	0.018	-.670446	-.0630436
estciv	.0329692	.2225322	0.15	0.882	-.4031858	.4691243
lningt	-.0140338	.0183785	-0.76	0.445	-.050055	.0219874
miemb	-.6247595	.1054977	-5.92	0.000	-.8315311	-.4179879
carte	-.84635	.799407	-1.06	0.290	-2.413159	.7204589
cmedamb	.0370945	.7736421	0.05	0.962	-1.479216	1.553405
ceconom	.2171756	.3806418	0.57	0.568	-.5288686	.9632198
cfimaes	.1744405	.8298622	0.21	0.834	-1.452059	1.80094
csalud	.1676934	.4267842	0.39	0.694	-.6687883	1.004175
csocial	.1274039	.4294114	0.30	0.767	-.7142269	.9690347
cderec	-.2247276	.6355843	-0.35	0.724	-1.47045	1.020995
cdocen	.537598	.3818023	1.41	0.159	-.2107208	1.285917
cingen	.1917072	.4602926	0.42	0.677	-.7104497	1.093864
_cons	-5.290011	1.032505	-5.12	0.000	-7.313683	-3.266338

C. GRÁFICOS

Gráfico A 4.1

COSTA RICA: Tasa neta de población femenina económicamente activa

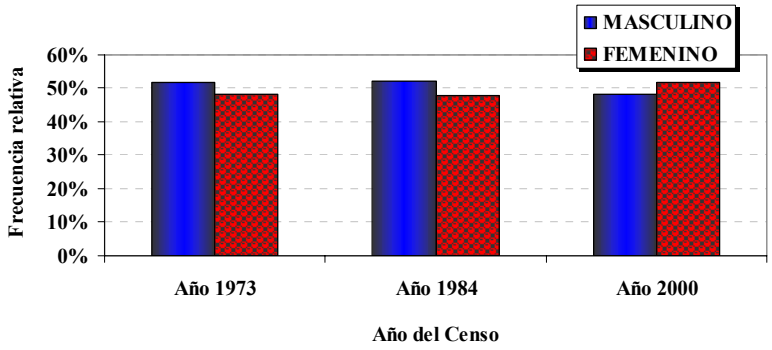
Según la edad en años cumplidos. 1973, 1984 y 2000



Fuente: Instituto Nacional de Estadísticas y Censos (INEC), Censos 1973, 1984 y 2000.

Gráfico A 4.2

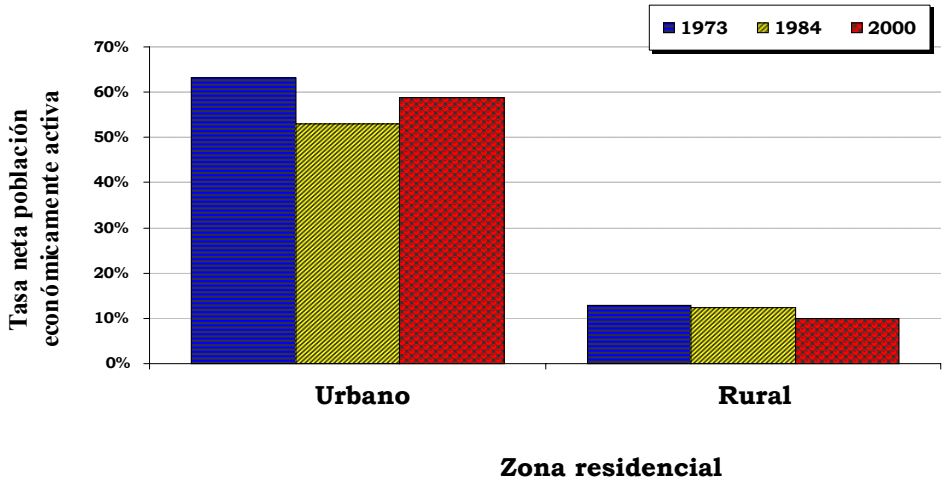
COSTA RICA: Distribución relativa de la población, según nivel universitario. 1973, 1984 y 2000



Fuente: Idem gráfico A 4.1.

Gráfico A 4.4

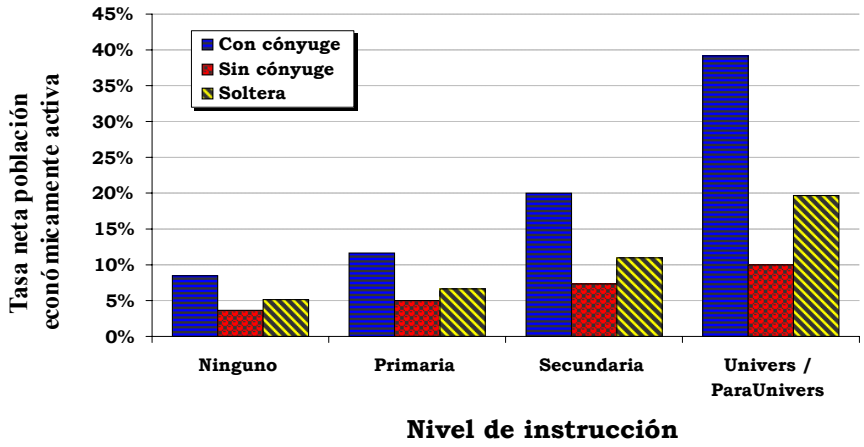
COSTA RICA: Tasa neta de población femenina universitaria económicamente activa entre 25 y 59 años de edad, según la zona de residencia. 1973, 1984 y 2000.



Fuente: Idem gráfico A 4.1.

Gráfico A 4.5

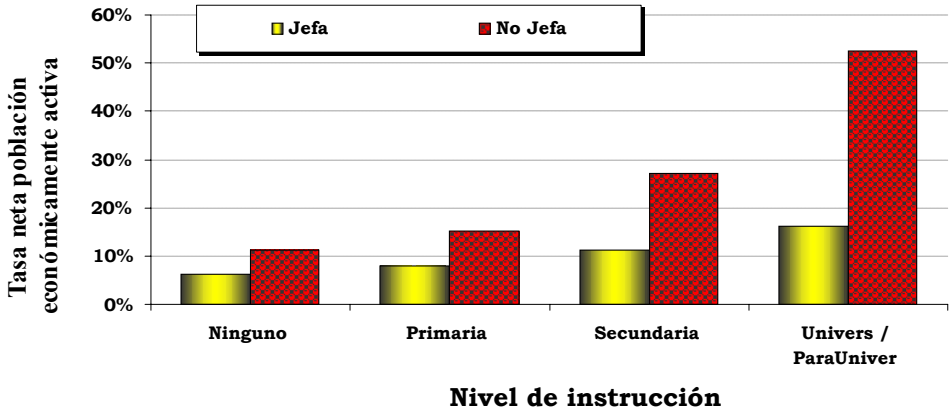
COSTA RICA: Tasa neta de población femenina económicamente activa, entre los 25 y 59 años de edad según estado civil y nivel de instrucción. 2000



Fuente: Instituto Nacional de Estadística y Censos (INEC), Censo 2000.

Gráfico A 4.6

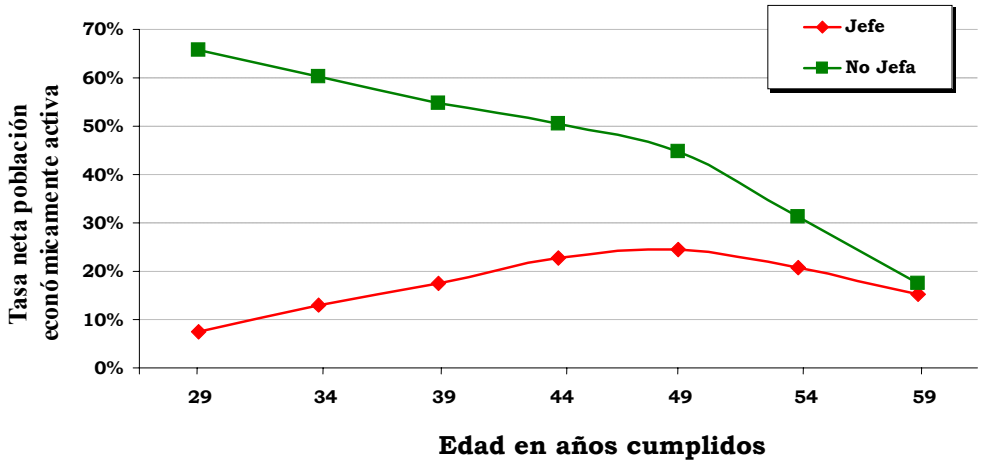
COSTA RICA: Tasa neta de población femenina activa entre los 25 y 59 años de edad, según jefatura y nivel de educación. 2000.



Fuente: Idem gráfico A 4.5.

Gráfico A 4.7

COSTA RICA: Tasa neta de población femenina universitaria económicamente activa, entre los 25 y 59 años de edad según Jefatura. 2000



Fuente: Idem gráfico A 4.5.