

ESTRATIFICACIÓN DE HOGARES Y SEGMENTOS POR NIVELES DE INGRESO EN EL CENSO 2000 ¹

M.Sc. Johnny Madrigal Pana²

RESUMEN

El objetivo del estudio fue construir un indicador del ingreso para estratificar los hogares y segmentos del Censo 2000. Para hacerlo, se tomó como base la información de la Encuesta de Hogares de Propósitos Múltiples 2001 y se escalaron seis índices: educación, ocupación, pertenencias de la familia, hacinamiento, dependientes y ocupados. Éstos fueron las variables independientes en un modelo de regresión multivariable, cuya variable dependiente fue el ingreso per cápita del hogar. Con el objeto de incluir todos los hogares del país se ajustaron dos ecuaciones: una para hogares en los que el jefe se encontraba ocupado y otra en los que no. Los índices resultaron significativos en los modelos de regresión y, además, se incluyó la zona de residencia. El coeficiente de determinación (R^2) para los hogares con jefe ocupado fue de 61% y para los que no fue de 51%. Las ecuaciones fueron aplicadas a los hogares del Censo 2000 y con el uso del K-Medias (Cluster Analysis) se definieron tres niveles de ingreso. En la zona urbana el porcentaje de hogares en el nivel de ingreso bajo fue de 23%, en el medio de 67% y en el alto, 10%. En la parte rural los porcentajes fueron 54%, 44% y 2%, respectivamente. Con el uso del indicador para los hogares se estratificaron los segmentos censales usando nuevamente el K-Medias. En el área urbana el 30% de los segmentos fue clasificado en el nivel de ingreso bajo, el 53% en el medio y el 9% en el alto. Para la zona rural los porcentajes fueron de 88%, 11% y 1%, respectivamente. El coeficiente E_{θ} ² mostró que la estratificación de los segmentos aplicada al Marco Muestral de Viviendas 2000 (MMV-2000) explica el 80% de la variabilidad del ingreso per cápita en la zona urbana y el 63% en la rural.

JUSTIFICACIÓN

Uno de los factores que produce más desigualdad en las sociedades es el ingreso, ya que éste posibilita la adquisición de bienes y servicios valorados económicamente, incluyendo los que se requieren para satisfacer las necesidades elementales de la vida. Por esta razón, los investigadores sociales poseen un interés continuo por diferenciar y ordenar a las personas, familias o grupos de familias, según categorías o estratos de ingreso con diferentes propósitos.

Algunos investigadores, particularmente los sociólogos, orientan su trabajo a las causas de la diferenciación que produce el dinero y, por ello, existe una amplia literatura en la que se exponen las teorías sobre estratificación social y desigualdad. Otros, por el contrario, estudian más los efectos de la desigualdad que genera el ingreso y analizan su impacto en diversas áreas.

¹ Borrador para discusión – No citar.

² Instituto Nacional de Estadística y Censos (INEC). Consultor. e-mail jomapa@racsa.co.cr

Entre quienes investigan los efectos de la estratificación, principalmente antropólogos, psicólogos, politólogos, historiadores y también sociólogos, se encuentran los que están orientados a estudiar las relaciones de poder, la influencia de los individuos y la formación de estilos de vida, para analizar, principalmente, la estructura social de las clases, su composición y la relación que guarda con fenómenos culturales, económicos, políticos, demográficos y sociales. Otros, como los estadísticos, y particularmente los muestristas, lo hacen con la perspectiva de lograr una estratificación estadística, es decir, para disminuir la variancia de las estimaciones que se realizan a partir de las encuestas por muestreo. Adicionalmente, los investigadores que usan técnicas cuantitativas utilizan la estratificación estadística con fines analíticos, en los que la relación del ingreso con conocimientos, actitudes y prácticas son de particular importancia.

Esta breve clasificación de quienes estudian la estratificación que genera el ingreso se considera indispensable, debido a que el Censo ofrece la posibilidad de estudiarla desde diferentes puntos de vista. Para evitar confusión con respecto a los fines que persigue el presente trabajo es necesario aclarar que su realización obedece al deseo de lograr una estratificación estadística de los hogares y segmentos del Censo 2000 y que, por ello, los procedimientos y resultados obtenidos solamente tratan de cumplir ese objetivo.

La estratificación de los hogares fortalecerá el análisis de los datos censales, ya que los temas económicos, sociales y demográficos que indagó el censo podrían asociarse con el indicador del ingreso. Por su parte, la estratificación de los segmentos censales permitirá incluir en el Marco Muestral de Viviendas 2000 (MMV-2000) un elemento que fortalezca el diseño muestral de la Encuesta Nacional de Ingresos y Gastos que planea realizar el INEC en el 2003, así como el de otras encuestas por muestreo que le asignen al nivel de ingreso una importancia particular.

El documento que se presenta incluye ocho secciones. Una primera explica la metodología y la estrategia para construir, a partir de la Encuesta de Hogares de Propósitos Múltiples 2001 (EHPM-2001), el indicador del ingreso con base en un modelo de regresión múltiple. En la segunda, se presentan las variables de la Encuesta utilizadas en el estudio y, en una tercera, se muestra cómo se escalaron los índices (variables independientes). En la cuarta sección se estudia el ingreso registrado en la Encuesta y se examina su relación con los índices. En la siguiente se ajusta el modelo de regresión para construir la ecuación final del indicador del ingreso y se comentan los resultados. En la sexta sección se explica cómo se aplicó la ecuación del ingreso en los hogares del Censo 2000, cómo fueron estratificados y los resultados obtenidos. En la siguiente sección se hace lo mismo para los segmentos censales. El artículo finaliza con las principales conclusiones y algunas reflexiones sobre el trabajo realizado.

METODOLOGÍA

El objetivo del estudio es construir un indicador del ingreso de los hogares y segmentos del Censo 2000 para estratificarlos según niveles de ingreso. Para cumplir con este objetivo, primero hay que enfrentar la ausencia de información sobre ingreso en el Censo 2000³. Por esta razón, se

³ El censo no es un instrumento idóneo para recolectar este tipo de información con niveles de calidad aceptables. Generalmente, estos datos se obtienen de encuestas por muestreo, como la Encuesta Nacional de Ingresos y Gastos que planea realizar el INEC en el 2003 (ENIG-2003), en la que, por ejemplo, se confeccionan cuestionarios amplios, se capacita al personal rigurosamente en la temática y se establecen controles numerosos para evaluar la calidad de la

recurrió a la EHPM-2001, la que recoge datos sobre el ingreso de las personas (ingreso primario, secundario, los provenientes de transferencias regulares, rentas de la propiedad e inversiones) (INEC, 2002). Aunque es de esperar que el dato de ingreso de la Encuesta se encuentre subestimado, es decir, que se declaren ingresos menores a los reales, aspecto común en este tipo de encuestas⁴, es factible asumir, como se podrá corroborar posteriormente, que los montos registrados discriminan adecuadamente diferentes niveles de ingreso y que el sesgo potencial no afecta considerablemente la relación esperada del ingreso con otras variables de estudio, como lo son, por ejemplo, la educación y la ocupación. Otra de las bondades de la EHPM-2001 es que la boleta recolectó información similar a la incluida en los Censos 2000, tanto en el tema de población como en el de vivienda, por lo que su aprovechamiento fue medular para desarrollar el trabajo.

Por estas razones, la estrategia general para realizar el trabajo se hizo en tres pasos, los que se resumen a continuación:

1. Se usó el archivo de datos de la EHPM-2001 para construir una ecuación que estimara el ingreso per cápita de los hogares.
2. Con la ecuación se estimó el ingreso per cápita de los hogares del Censo 2000.
3. Se estratificaron los hogares y segmentos del MMV-2000 por niveles de ingreso.

A continuación se hacen observaciones específicas de carácter metodológico que es necesario tomar en cuenta durante la lectura del documento.

Por motivo de fechas se debió utilizar la base de datos de la Encuesta del año 2000, ya que fue ejecutada tres semanas después del trabajo de campo censal. Sin embargo, se optó por la del 2001 debido a que las ocupaciones fueron codificadas con el manual de OIT más reciente, tal y como se hizo con el Censo 2000 (INEC, 2000).

El uso de la muestra para el análisis de los datos merece otra observación. Concretamente, la muestra de 726 Unidades Primarias de Muestreo (UPM)⁵ de la EHPM-2001 fue dividida en dos mitades. En la primera se aplicaron los análisis requeridos y en la segunda se validaron los resultados. La división de la muestra se hizo ordenando geográficamente las UPM y luego se aplicó una selección sistemática de uno de cada dos.

Otro aspecto que es necesario aclarar es lo relacionado con el jefe del hogar, pues se sabe que no todos trabajan y, por lo tanto, no tienen información sobre ocupación. Como esta variable es básica para construir el indicador del ingreso, de pasar por alto este detalle se estaría dejando

información. Estas particularidades, que son opciones de una encuesta por muestreo bien diseñada, son difíciles de ejecutar en un censo por la envergadura del trabajo.

⁴ En Estados Unidos, una revisión de la literatura relacionada con el tema de la calidad de la medición del ingreso en las encuestas, encontró que el monto de la subestimación depende del tipo de ingreso. En la declaración de los salarios y en los pagos recibidos por la seguridad social la subestimación varía entre 5 y 8%. Sin embargo, en los ingresos provenientes de intereses o dividendos la subestimación alcanza el 50% o más (Moore, et.al 2000).

⁵ En la EHPM las UPM son segmentos censales en los que se seleccionó una submuestra de aproximadamente 15 viviendas.

fuera del análisis aproximadamente el 24% de los hogares del país. Para evitar esta situación, se modificó la definición de jefe de hogar de la siguiente manera:

1. Se seleccionó al jefe si se estaba ocupado (76.5% del total de hogares).
2. Cuando el hogar no tenía un jefe de hogar ocupado, se seleccionó como jefe al residente habitual de mayor edad que estaba ocupado (12.4%), excluyendo el servicio doméstico.
3. Se escogió al residente habitual de mayor edad cuando en el hogar no habían personas ocupadas (11.1%) (en el 90% de los casos el jefe coincidió con esta persona).

Esto originó dos ecuaciones para estimar el ingreso. La primera, para los hogares con “jefe” ocupado (grupos 1 y 2), y la segunda, para los hogares que no lo tenían (grupo 3). El cambio en la definición pretende fortalecer la estimación del ingreso con base en la información sobre ocupación y, al tener dos ecuaciones, posibilita el cálculo del indicador de ingreso en todos los hogares del Censo 2000.

En lo que respecta al análisis de los datos debe mencionarse que se utilizó la técnica de regresión múltiple (Kleinbaum, 1998). La variable dependiente del modelo fue el ingreso per cápita del hogar. Para construirla, se sumaron los ingresos del hogar, tal y como registran en la EHPM, y se dividió entre el total de residentes habituales. Este ingreso fue corregido utilizando los parámetros que se derivan de un estudio que realizó CEPAL para Costa Rica en 1991 (MEIC y DGEC, 1997). Concretamente, en los hogares urbanos del país el ingreso se incrementó 17.4% y en los rurales 35.8%. Debe tomarse en cuenta que este ajuste se basa en el análisis hecho por CEPAL con datos de 1988, por lo que se está suponiendo que estos factores de ajuste se han mantenido constantes en el tiempo.

Las variables independientes del modelo de regresión fueron seis índices que resumieron diferentes aspectos del hogar. Los índices son: educación y ocupación del jefe de hogar, y el de pertenencias (artefactos), hacinamiento, proporción de dependientes y número de personas ocupadas a nivel de hogar. Adicionalmente, la zona de residencia fue utilizada en el modelo como variable independiente.

Para construir los índices se utilizaron procedimientos simples, con la excepción del índice de ocupación, en el que se usó el “Cluster Analysis” para formar grupos de ocupaciones que discriminaran el ingreso. En este procedimiento se forman “cluster” o grupos de observaciones, usando como guía el criterio de minimizar la variancia entre grupos y maximizarla entre ellos. Para lograrlo, utiliza un procedimiento iterativo en el que cada observación j es primero examinada y luego asignada a un “cluster” k , para el que la distancia

$$\sum (x_{ij} - x_{ik})^2$$

sea un mínimo. El x_{ij} es la i -ésima variable para la j -ésima observación y x_{ik} es la media de la i -ésima variable en el k -ésimo “cluster” después de la inclusión de la j -ésima observación. El procedimiento continúa hasta que ninguna observación sea asignada a un nuevo “cluster” o hasta que un número máximo de interacciones prefijado haya sido alcanzado (Golder, 1973).

Para realizar la estratificación de hogares y segmentos se usó una modificación del “Cluster Analysis”, al que se le denomina K-Medias. Éste posee la característica de manejar bases de datos más grandes que el Cluster, de 200 datos o más, y facilita obtener los límites de los estratos contruidos (SPSS, 1999).

Con relación a los cuadros generados a partir del archivo de datos de la EHPM-2001 debe mencionarse que se anotó con la letra “n” o con la leyenda “Casos en la muestra”, el número de observaciones de la muestra sin ponderar (muestra natural), para tener presente el número real de casos utilizado.

EXAMEN DE LAS VARIABLES INDIVIDUALES DEL ESTUDIO

A continuación se presentan las variables individuales utilizadas en el estudio y sus características, las que posteriormente -en la siguiente sección-, serán transformadas a índices para que funcionen como variables independientes en los modelos de regresión.

La selección de las variables individuales incluye las que están asociadas, teóricamente, con el ingreso familiar. Es decir, se realizó una selección inicial de variables con base en la experiencia. A esto se le llama validez lógica o facial (Babbie, 1989). Éstas fueron divididas inicialmente en cuatro grupos, como se detalla a continuación:

<i>GRUPO</i>	<i>VARIABLE</i>
Relacionadas con el jefe del hogar	1. Nivel de educación 2. Ocupación principal 3. Categoría ocupacional 4. Sector Institucional
Relacionadas con las pertenencias del hogar	5. TV Color 6. Refrigeradora 7. Lavadora 8. Teléfono 9. Ducha para agua caliente 10. Microondas 11. Vehículo (no de trabajo) 12. Microcomputadora 13. Tanque para agua caliente
Relacionados con aspectos económicos	14. Número de residentes ocupados 15. Número de dormitorios
Relacionadas con aspectos demográficos	16. Número de residentes habituales 17. Número de residentes menores de 15 18. Número de residentes de 65 años o más 19. Zona de residencia

Los resultados para el jefe del hogar se muestran en el Cuadro 1. Como puede observarse, el porcentaje que no posee educación es relativamente bajo, de 5%. Los que poseen algún nivel de primaria son la mitad y en secundaria se encuentra el 26%. En el nivel de educación universitaria se concentra el 17%. En términos relativos, los jefes tienen niveles educativos más elevados en la zona urbana que en la rural.

La ocupación principal indica que, a nivel nacional, los jefes se concentran más en el trabajo no calificado (21.3%)⁶. Le sigue en orden de importancia los que están en las ocupaciones calificadas de la producción artesanal, la construcción, la mecánica y artes mecánicas, las artes gráficas y la industria manufacturera (12.8%) y los de venta en locales y prestación de servicios directos a personas (12.6%). Los que se encuentran en ocupaciones de nivel técnico y profesional medio también presentan un porcentaje importante (11.6%). Otros grupos muestran magnitudes menores al 10%.

Las diferencias según zona de residencia son evidentes, ya que los jefes en ocupaciones generadoras de mayores ingresos son más frecuentes en la zona urbana que en la rural.

Cuadro 1
Distribución porcentual de los jefes de hogar según algunas características de interés por zona de residencia

Características	Urbano	Rural	Total
Casos en la muestra (n)	(2.201)	(2.949)	(5.150)
Total	100	100	100
Nivel de educación			
Ninguna	3,8	9,4	5,5
Primaria incompleta	12,6	26,2	18,1
Primaria completa	26,9	38,8	31,7
Secundaria incompleta	16,9	11,6	14,7
Secundaria completa	15,8	6,5	12,0
Universitaria incompleta	8,3	3,1	6,2
Universitaria completa	15,7	3,8	10,9
Ns-Nr	0,9	0,7	0,8
Ocupación principal 1/			
Nivel Direc en Adm Publ y Emp Priv	3,7	1,4	2,8
Nivel Prof Cient Intelec	11,4	2,9	8,0
Nivel Tec y Prof Medio	14,3	7,5	11,6
Apoyo administrativo	5,8	2,3	4,4
Venta locales y servicios	14,8	9,4	12,6
Act Agrop Agric y Pesq	2,2	14,7	7,3
Prod Artesa Const Meca	13,9	11,0	12,8
Montaje y oper de instal	10,2	9,3	9,8
No calificadas	14,1	32,1	21,3
Ns-Nr	9,5	9,5	9,5
Categoría ocupacional 1/ 2/			
Empleado del Estado	7,7	4,9	6,6
Empleado Inst Autónoma	9,1	2,8	6,6
Empleado Empresa Privada	40,4	45,9	42,6
Cuenta Propia	20,0	23,9	21,6
Patrono-Socio Activo	10,4	9,3	10,0
Servidor doméstico	2,7	3,1	2,8
Ns-Nr	9,7	10,1	9,8

1/Incluye jefes de hogar ocupados

1/Variable reconstruida con base en el sector institucional

Fuente: INEC- Archivo de datos de Encuesta de Hogares y de Propósitos Múltiples del 2001 (50% de la muestra)

La categoría ocupacional muestra que los jefes del hogar se desempeñan más como empleados en la empresa privada (42.6%). También, un porcentaje importante es cuenta propia (21.6%). El

⁶ Éste aglomera vendedores ambulantes y afines, trabajadores de servicios prestados en la calle y otras ocupaciones desempeñadas por personal no calificado.

10% tiene la categoría de patrono o socio activo, y el Estado y las Instituciones Autónomas aglomeran el 7% cada una. La categoría ocupacional varía menos por zona de residencia que la educación y la ocupación.

En el Cuadro 2 se presenta el resto de las variables individuales de estudio. Las relacionadas con las pertenencias del hogar muestran un amplio rango de variación. El TV Color y la refrigeradora son los artefactos más comunes en los hogares (87% ambos) y el vehículo no de trabajo (26.8%), la microcomputadora (16.8%) y el tanque para agua caliente (3.9%) los menos frecuentes. Sin excepción, todos los porcentajes son más altos en los hogares urbanos que en los rurales, principalmente el teléfono, la ducha para agua caliente y el microondas.

Cuadro 2
Características de los hogares por zona de residencia

Variables	Urbano	Rural	Total
Casos en la muestra (n)	(2.201)	(2.949)	(5.150)
Pertenencias del hogar en porcentajes^{1/}			
TV Color	92,9	78,2	87,0
Refrigeradora	92,0	79,0	86,8
Lavadora	89,3	74,7	83,4
Teléfono	76,2	38,7	61,1
Ducha para agua caliente	55,5	22,9	42,4
Microondas	50,7	20,0	38,4
Vehículo (no de trabajo)	32,5	18,3	26,8
Microcomputadora	23,5	6,8	16,8
Tanque para agua caliente	5,8	1,2	3,9
Aspectos económicos del hogar			
Número promedio de ocupados	1.62	1.53	1.58
Número promedio de dormitorios	2.66	2.47	2.59
Aspectos demográficos del hogar			
Número promedio de residentes habituales	3.94	4.17	4.03
Promedio de menores de 15 años	1.09	1.44	1.23
Promedio de personas con 65 años o más	0.26	0.22	0.24

^{1/} Solamente se encontraron dos valores faltantes en cada respuesta

Fuente: INEC-Archivo de datos de Encuesta de Hogares de Propósitos Múltiples del 2001 (50% de la muestra)

En los aspectos económicos se observa que, el número de ocupados es 1.58 y el de dormitorios es 2.59. Las dos variables son levemente mayores en la zona urbana que en la rural.

Las variables demográficas muestran que el número de residentes habituales es 4.03 por hogar, que el de personas menores de 15 años es 1.23 y las de 65 años o más, 0.24. La zona de residencia indica que, mientras el promedio de jóvenes por hogar es menor en la zona urbana que en la rural (1.09 y 1.44, respectivamente), el de las personas mayores es más alto en la primera que en la segunda (0.26 y 0.22, respectivamente). Esto es el efecto de un patrón de fecundidad que favorece menos hijos en la zona urbana del país y, por lo tanto, de un proceso de envejecimiento más evidente en esa zona.

ESCALAMIENTO DE VARIABLES INDEPENDIENTES DEL MODELO DE REGRESIÓN

A continuación se explica cómo, a partir de las variables presentadas en la sección anterior, fueron escalados los índices que serán utilizados como variables independientes en el modelo de regresión. Se procede desde los menos complejos hasta el que requirió más elaboración (índice de ocupación).

Para el índice de educación del jefe del hogar se consideraron dos opciones: los años de educación y el nivel educativo. Al respecto, se comprobó que la correlación con el ingreso per cápita era prácticamente igual (0.57 y 0.56, para los años de educación y el nivel educativo), por lo que se decidió adoptar el procedimiento más simple, es decir, el nivel educativo. De esta manera, el índice se escaló de la siguiente manera:

<i>CATEGORÍA</i>	<i>ESCALAMIENTO</i>
Sin estudios	0
Primaria incompleta	1
Primaria completa	2
Secundaria incompleta	3
Secundaria completa	4
Universitaria incompleta	5
Universidad completa	6

Teóricamente, se está postulando que a mayor nivel educativo del jefe del hogar, mayor será el ingreso, motivo por el que el escalamiento se incrementa conforme lo hace la educación.

Las pertenencias del hogar fueron resumidas en un índice que indica el número de artefactos en el hogar. Para ello, se recodificó la tenencia como uno y la ausencia de éste como cero. De esta manera, el índice se construyó sumando la tenencia de las pertenencias. El razonamiento subyacente sugiere que un número mayor de artefactos en el hogar es producto de un ingreso per cápita mayor. Por ello, cuando la tenencia es nula (no posee artefactos) es de esperar un ingreso menor que el obtenido cuando es alta (digamos nueve).

Las variables número residentes habituales y número de dormitorios fueron resumidas en un índice de hacinamiento, es decir, se calculó la razón de personas por dormitorio en la vivienda⁷. La fórmula puede expresarse de la siguiente manera:

$$\text{Índice de hacinamiento} = \frac{\text{Número de residentes habituales}}{\text{Número de dormitorios}}$$

Se supone que el hacinamiento está correlacionado inversamente con el ingreso per cápita del hogar. Es decir, si el número de personas por dormitorio es alto (por ejemplo 3), entonces se supone que el hogar no posee los ingresos suficientes para que sus residentes descansen más confortablemente. Por el contrario, si el número es bajo (por ejemplo 1), puede ser porque existe más comodidad, producto de un ingreso mayor. El índice de hacinamiento se concibió como un indicador que sintetiza características sociales y económicas de los hogares.

⁷ Los dormitorios hacen referencia a los cuartos o habitaciones de la vivienda que son dedicados exclusivamente para dormir.

Con los aspectos demográficos de los residentes de la vivienda se construyó la proporción de dependientes⁸. Para calcularlo se procedió de la siguiente manera:

$$\text{Proporción de dependientes} = \frac{\text{Residentes menores de 15 años} + \text{Residentes de 65 años o más}}{\text{Total de residentes habituales}}$$

Básicamente, se está asumiendo que los residentes menores a 15 años y los de 65 años o más, generalmente no participan de las actividades económicas. Esto significa que proporciones altas de dependientes reflejan una carga económica importante hacia las personas que trabajan en el hogar, por lo que es de esperar ingresos per cápita menores. O a la inversa, proporciones bajas de dependientes sugieren una carga económica menor para los que trabajan en el hogar, por lo que es de esperar ingresos per cápita mayores. En otras palabras, es de esperar una asociación inversa entre la proporción de dependientes y el ingreso per cápita del hogar.

El número de ocupados fue el penúltimo índice construido. Para calcularlo solamente se contó a los residentes habituales del hogar que cumplían con la definición de ocupado. Es de esperar que a mayor número de ocupados mayor sea el ingreso per cápita o viceversa.

La construcción del índice de ocupación fue más compleja. Para empezar, se buscó cómo relacionar el ingreso con la ocupación, la categoría ocupacional y el sector institucional, debido a que es lógico suponer que el ingreso varía dependiendo de estas variables.

Para proceder, se tomaron dos decisiones iniciales. La primera fue trabajar con el ingreso total mensual del jefe del hogar y, la segunda, construir una variable nueva, denominada categoría ocupacional del jefe del hogar, que tomara en cuenta el sector institucional.

De esta manera, se tabuló la información que se presenta en el Cuadro 3. En éste se puede observar el ingreso total mensual promedio del jefe del hogar por ocupación principal y por categoría ocupacional. Analizando en primera instancia la columna de total para la ocupación principal (última columna), se obtiene que el ingreso mensual varía de 73029 colones en las ocupaciones no calificadas hasta 463915 colones en los niveles directivos de la administración pública y la empresa privada. La categoría ocupacional (última línea del Cuadro 3), muestra menos variación, aunque es evidente que los empleados de instituciones autónomas perciben los salarios mayores y los servidores domésticos los menores.

Cuadro 3
Ingreso Total Mensual promedio del Jefe del Hogar por ocupación principal según categoría ocupacional

OCUPACIÓN PRINCIPAL	CATEGORÍA OCUPACIONAL						Total
	Empleado del Estado	Empleado Institución Autónoma	Empleado Empresa Privada	Cuenta Propia	Patrón-Socio Activo	Servidor doméstico	
	TOTAL (n=7.997)						
Nivel Direc en Adm Publ y Emp Priv	373.905	472.970	515.149	*310.068	*456.700		463.915
Nivel Prof Cient Intelec	272.730	509.199	376.331	436.483	517.524		391.401

⁸ La otra forma de proceder era calculando la razón de dependencia económica por hogar, la que incluiría en el denominador de la fórmula anterior, la población de 15 a 64 años. Sin embargo, se evitó este procedimiento debido a que el cálculo se indefiniría cuando en el hogar residen únicamente personas de 65 años o más, hecho que cada vez es más frecuente en el país.

Nivel profesional	Cta propia	2+4	436.483	5.891	44	11,1	9	9
Nivel directivo	Patrono	1+5	456.700	7.420	76	18,9	9	
Directivo+Apoyo	Autónoma+Patrono	1-4+2-5	488.927	3.566	31	11,3	10	10
Nivel profesional	Autónoma	2+2	509.199	12.984	102	8,0	10	
Nivel Directivo	Privada	1+3	515.149	8.611	64	8,4	10	
Nivel profesional	Patrono	2+5	517.524	6.486	52	14,0	10	

1/ Ocupación principal Nivel Direc en Adm Publ y Emp Priv (1), Nivel Prof Cient Intelec (2), Nivel Técnico y Prof Medio (3), Apoyo Administrativo (4), Venta locales y servicios (5), Act Agrop Agric y Pesq (6), Prod Artesa Const Meca (7), Montaje y oper de instal (8), No calificadas (9).

Categoría ocupacional Empleado del Estado (1), Empleado de Inst Autónoma (2), Empleado Empresa Privada (3), Cuenta propia (4), Patrono o socio activo (5), Servidor Doméstico (6).

Fuente: INEC-Archivo de datos de Encuesta de Hogares de Propósitos Múltiples del 2001.

Se utilizó el 100% de la muestra.

En éste se incluye, para cada ingreso promedio la ocupación principal y la categoría ocupacional respectiva. Este cuadro es el resultado de ordenar la información del Cuadro 3 y permite saber fácilmente qué ocupaciones y categorías ocupacionales presentan menores o mayores ingresos. También muestra el número de casos (ponderado y sin ponderar) con el que se estimó el ingreso total mensual promedio del jefe del hogar y aspectos relacionados con la variabilidad en cada categoría.

La información de este cuadro es básica para construir el índice de ocupación. El punto a resolver aquí es cómo clasificar las ocupaciones y categorías de manera óptima, es decir, procurando puntos de corte por ingreso que garanticen un número de grupos menor, de manera que la variancia del ingreso dentro de los grupos sea la más pequeña posible y entre los grupos sea la mayor. Este dilema se resolvió usando un “Cluster Analysis”. Es decir, introduciendo la información del Cuadro 4 en el procedimiento “Cluster” y especificándole el número de categorías deseado.

Para explorar diferentes posibilidades se construyeron 7 clasificaciones diferentes, contemplando la formación de 8 hasta 14 grupos. Para decidir cuál era la mejor clasificación se correlacionó cada agrupación con el ingreso per cápita del hogar. El resultado indicó que al utilizar entre 8 y 14 grupos la correlación no varió considerablemente (0.54 a 0.56), por lo que se decidió dejar 10 grupos. Los resultados que también se sintetizan en el Cuadro 4, permiten determinar las ocupaciones y sus códigos respectivos cuando se forman 10 grupos de ocupaciones. El escalamiento dado a los grupos fue de 1 hasta 10, como se muestra en la última columna del cuadro en mención y se puede afirmar que es un procedimiento “arbitrario” que diferencia los niveles de ingreso según la ocupación.

EXAMEN DE LAS RELACIONES BIVARIADAS

Hasta este momento, la única variable que no se ha presentado es la del ingreso per cápita del hogar registrada en la EHPM-2001. Por esta razón, en esta sección se obtienen sus características. Además, como es usual, previo a la estimación del modelo de regresión se analizan las relaciones bivariadas de la variable dependiente (ingreso per cápita) con las independientes (índices), ya que la dirección y la intensidad de la asociación permite corroborar los planteamientos teóricos argumentados en la sección anterior. Además, también es útil estudiar la relación entre los índices, debido a que las altas correlaciones ocasionan problemas de multicolinealidad.

En el Cuadro 5 se presenta la información del ingreso per cápita del hogar, el que se estimó en 77115 colones a nivel nacional. Para la zona urbana fue de 93346 colones y para la rural de

54864. Esto significa que el ingreso per cápita de los hogares urbanos es 1.7 veces mayor que los rurales⁹.

Por otra parte, un aspecto de la variable dependiente que generalmente se estudia es la forma de su distribución, debido a que la regresión requiere el cumplimiento del supuesto de normalidad de esta variable. Para ello se procedió de la manera usual, es decir, estimando el estadístico de Kolmogorov-Smirnov y graficando la variable.

Los resultados muestran que a nivel nacional, y en las zonas urbana y rural, la distribución del ingreso per cápita no cumple con el supuesto de normalidad (Cuadro 5 y Gráfico 1). Por este motivo, se procedió a transformar la variable con el logaritmo natural. Los resultados de esta transformación modifican el panorama, ya que el estadístico de K-S indica que no existe evidencia para asumir que la distribución resultante sea diferente a la normal. El Gráfico 2 permite tener una idea más clara, pues la forma de la distribución es evidente. Este resultado hizo que, en adelante, se siguiera trabajando con la transformación.

Las variables independientes, es decir, los índices construidos, se muestran en el Cuadro 6. Un resultado de interés son las variaciones que se observan por zona de residencia. El índice promedio de la educación es 3.19 en el área urbana y de 2.04 en la rural, y el de artefactos es 5.19 y 3.40, respectivamente. Con excepción del hacinamiento y de la proporción de dependientes, todos los promedios son mayores en la zona urbana que en la rural.

Cuadro 5
Estadísticos relacionados con el Ingreso per cápita del hogar

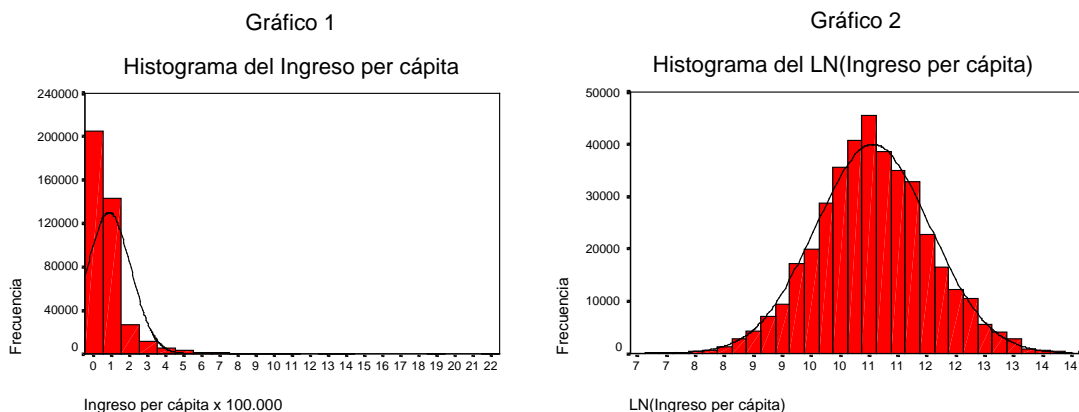
Estadísticos	Urbano	Rural	Total
Ingreso per cápita 1/			
Media (en colones)	93346	54864	77115
Desviación estándar	113469	72833	100216
Kolmogorov-Smirnov Z 2/	3,57	3,56	4,80
Significancia asintótica (2 colas)	0,00	0,00	0,00
LN(Ingreso per cápita)			
Media	11,01	10,5	10,8
Desviación estándar	0,91	0,94	1,01
Kolmogorov-Smirnov Z 2/	0,77	0,81	0,74
Significancia asintótica (2 colas)	0,59	0,52	0,63

1/ El ingreso per cápita reportó un 18.1% de valores faltantes.

2/ Para calcular los estadísticos relacionados con K-S se dividió entre 1000 el factor de expansión para evitar que el tamaño de muestra afectara la prueba de significancia.

Fuente: INEC- Archivo de datos de Encuesta de Hogares de Propósitos Múltiples del 2001 (50% de la muestra)

⁹ Los resultados de la Encuesta Nacional de Ingresos y Gastos de 1988 (ENIG-88) muestran un resultado similar, ya que el ingreso per cápita del hogar resultó 1.8 mayor en la zona urbana que en la rural (MEIC y DGEC, 1989). Aunque la ENIG-88 no es una encuesta reciente y los montos de los ingresos absolutos han variado sustancialmente desde entonces, es difícil pensar que la relación del ingreso por zona de residencia haya sufrido cambios radicales desde 1988 hasta el presente. Por este motivo, se puede asumir que el ajuste introducido al ingreso registrado en la Encuesta de Hogares para la zona urbana y rural es aceptable y que también lo son las diferencias observadas por zona de residencia.



Cuadro 6
Estadísticos de resumen para los índices construidos

Índices	Promedio			Mínimo	Máximo	Error estándar de la media
	Total	Urbano	Rural			
Educación	2.72	3.19	2.04	0	6	0.0024
Ocupación	4.06	4.60	3.26	1	10	0.0033
Artefactos	4.47	5.19	3.40	0	9	0.0032
Hacinamiento	1.69	1.59	1.85	0.2	13	0.0015
Dependientes	0.35	0.34	0.38	0	1	0.0004
Número de ocupados	1.58	1.62	1.53	0	8	0.0015

Fuente: INEC-Archivo de datos de Encuesta de Hogares de Propósitos Múltiples del 2001 (50% de la muestra)

Las correlaciones de Pearson muestran que, a nivel nacional, los índices y el logaritmo natural del ingreso per cápita están correlacionados (Cuadro 7). La educación, la ocupación y los artefactos lo hacen de manera positiva, en magnitudes mayores a 0.5. El número de ocupados muestra una correlación modesta (0.24), aunque también es positiva. El hacinamiento y la proporción de dependientes correlacionaron negativamente con el ingreso transformado, como era de esperar, y lo hicieron en magnitudes de aproximadamente 0.35.

Cuadro 7
Coeficientes de Correlación de Pearson para los índices y el LN(Ingreso per cápita del hogar)

Índices	Educ	Ocupa	Artef	Hacin	Depen	Ocup
Educación	1,00					
Ocupación	0,67	1,00				
Artefactos	0,57	0,57	1,00			
Hacinamiento	-0,18	-0,21	-0,27	1,00		
Dependientes	-0,17	-0,10	-0,13	0,18	1,00	
Número de ocupados	0,05	0,03	0,15	0,21	-0,32	1,00
LN(Ingreso per cápita)	0,56	0,55	0,57	-0,36	-0,34	0,24
Casos en la muestra	(4.228)	(3.816)	(4.250)	(4.249)	(4.250)	(4.250)
URBANO						
Educación	1,00					
Ocupación	0,67	1,00				
Artefactos	0,53	0,53	1,00			
Hacinamiento	-0,18	-0,22	-0,25	1,00		
Dependientes	-0,15	-0,08	-0,11	0,15	1,00	
Número de ocupados	0,03	-0,01	0,14	0,24	-0,31	1,00
LN(Ingreso per cápita)	0,57	0,56	0,54	-0,35	-0,32	0,24
Casos en la muestra	(1.723)	(1.548)	(1.731)	(1.730)	(1.731)	(1.731)

	RURAL					
Educación	1,00					
Ocupación	0,53	1,00				
Artefactos	0,46	0,48	1,00			
Hacinamiento	-0,11	-0,14	-0,23	1,00		
Dependientes	-0,14	-0,07	-0,12	0,21	1,00	
Número de ocupados	0,04	0,05	0,15	0,21	-0,32	1,00
LN(Ingreso per cápita)	0,40	0,42	0,45	-0,33	-0,35	0,24
Casos en la muestra	(2.505)	(2.268)	(2.519)	(2.519)	(2.519)	(2.519)

Fuente: INEC-Archivo de datos Encuesta de Hogares del 2001 (50% de la muestra).

Por zona geográfica muestran que la correlación con los índices de educación, ocupación y artefactos es mayor en la zona urbana que en la rural. El resto de los índices muestra correlaciones prácticamente iguales en ambas zonas.

El otro aspecto que puede observarse es la correlación entre los índices. La máxima correlación encontrada es entre la educación y la ocupación (0.67), con signo positivo. El índice de artefactos también muestra una correlación positiva importante con la ocupación y la educación (0.57 en ambos casos). El resto presenta correlaciones modestas o bajas entre ellos. Este patrón se cumple también en la zona urbana y en la rural.

Los resultados encontrados muestran que las correlaciones entre los índices y el logaritmo natural del ingreso per cápita se dan en la dirección esperada y en magnitudes aceptables. Además, no se observaron correlaciones extremadamente altas entre los índices, por lo que momentáneamente, no se considera la exclusión de alguno de ellos del análisis.

ECUACIÓN DEL INGRESO

Para estimar la ecuación del ingreso se utilizará, como ya se indicó, un modelo de regresión múltiple, en el que la variable dependiente será el ingreso per cápita. Debe recordarse que se ajustarán dos modelos; uno para hogares en los que su “jefe” posee información sobre ocupación y otro para los hogares en los que no habían personas ocupadas (11.1% de los hogares). Además, se incluyó en el modelo la zona de residencia como una variable “dummy” (1=urbano, 0=rural). Estas decisiones se tomaron para establecer un procedimiento que diferencie por zona de residencia y sea aplicable a todos los hogares del Censo 2000.

Los resultados se muestran en el Cuadro 8. En primer lugar debe mencionarse que ambos modelos resultaron significativos, es decir, que los índices contribuyeron a explicar significativamente la variación del logaritmo natural del ingreso per cápita. En segundo lugar, se obtuvo que cada uno de los índices construidos resultó significativo, o sea, que es razonable incluirlos a todos en las ecuaciones. Esto incluye la zona de residencia, variable cuya pertenencia al modelo era de esperar.

Para estimar el logaritmo natural del ingreso per cápita en los hogares con jefe ocupado se utiliza la siguiente ecuación, obtenida del Cuadro 8 (primera columna):

$$LN(YPC) = 10.15 + Edu * 0.12 + Ocu * 0.09 + Art * 0.08 - Hac * 0.18 - Dep * 0.86 + Nocu * 0.15 + Zon * 0.01 \quad (R^2 = 0.61)$$

En los hogares donde no habían personas ocupadas se usa la siguiente:

$$LN(YPC) = 9.50 + Edu * 0.26 + Art * 0.14 - Hac * 0.38 + Dep * 0.32 + Zon * 0.06 \quad (R^2 = 0.51)$$

Los coeficientes de determinación (R^2) mostraron que en los hogares con jefe ocupado, los índices lograron explicar el 61% de la variabilidad del logaritmo natural del ingreso per cápita. Para los hogares con jefe no ocupado éste coeficiente fue menor, de 51%. Las pruebas de autocorrelación, multicolinealidad y heterocedasticidad mostraron que los modelos no presentan estos problemas.

Para validar los modelos ajustados se utilizó el archivo de la segunda submuestra. Con base en las ecuaciones construidas se estimó el logaritmo natural del ingreso per cápita del hogar y se correlacionó con la respectiva transformación del reportado en la Encuesta. La correlación de Pearson a nivel nacional fue de 0.75; para la zona urbana fue de 0.73 y para la rural de 0.70. Otra prueba que se hizo en el archivo de la segunda submuestra fue estimar nuevamente los modelos. Los resultados corroboraron que los índices, la magnitud de sus coeficientes, los signos y la significancia, no diferían de los obtenidos con la primera submuestra. Los coeficientes de determinación también mostraron resultados muy similares.

Cuadro 8
Resultados del Modelo de Regresión Multivariable utilizando LN(Ingreso per cápita) del hogar como variable dependiente

Indíces	Coeficiente Beta	Error Estándar	Coeficiente Beta Estandarizado	t	Sig	IC (95%) para Beta	
						Inferior	Superior
HOGARES CON JEFE OCUPADO (n=3.643)							
Constante	10.1494	0.004	--	2657	0	10.142	10.157
Educación	0.1203	0.001	0.227	145.5	0	0.119	0.122
Ocupación	0.0902	0.001	0.225	146.6	0	0.089	0.091
Artefactos	0.0837	0.001	0.210	139.4	0	0.082	0.085
Hacinamiento	-0.1803	0.001	-0.214	-170.6	0	-0.182	-0.178
Dependientes	-0.8550	0.004	-0.231	-190.3	0	-0.864	-0.846
Número de ocupados	0.1493	0.001	0.158	134.0	0	0.147	0.151
Zona de residencia 1/	0.0118	0.002	0.007	5.5	0	0.008	0.016
HOGARES CON JEFE NO OCUPADO (n=473)							
Constante	9.4993	0.014	--	689.206	0	9.472	9.526
Educación	0.2563	0.003	0.393	92.893	0	0.251	0.262
Artefactos	0.1361	0.002	0.287	65.411	0	0.132	0.140
Hacinamiento	-0.3834	0.005	-0.259	-72.664	0	-0.394	-0.373
Dependientes	0.3180	0.010	0.111	31.091	0	0.298	0.338
Zona de residencia 1/	0.0616	0.008	0.029	7.684	0	0.046	0.077

1/ 1=Zona urbana, 0=Zona rural

Fuente: INEC-Archivo de datos Encuesta de Hogares de Propósitos Múltiples del 2001 (50% de la muestra)

ESTRATIFICACIÓN DE LOS HOGARES

Las dos ecuaciones obtenidas en la sección anterior son básicas para estimar el ingreso per cápita de cada hogar en los Censos 2000. Para hacerlo, se siguió el siguiente procedimiento:

1. En los hogares del Censo 2000 se dividen los que tienen un jefe ocupado y los que no lo tienen, de la misma manera que se hizo en la EHPM.

2. Para cada hogar, se construyen los índices correspondientes (educación, artefactos, ocupación, etc.), de la misma forma que se hizo en la EHPM-2001.
3. Se sustituyen los valores de los índices en cada hogar del Censo 2000 en la ecuación correspondiente para estimar el logaritmo del ingreso per cápita.
4. Se aplica la función exponencial al resultado obtenido anteriormente.

El resultado de la estimación del ingreso per cápita en el Censo 2000 se comparó con el registrado en la EHPM-2001, calculando los deciles (Cuadro 9). Al respecto, se observa que en los primeros deciles el Censo estima montos de ingreso per cápita levemente mayores a los registrados en la Encuesta hasta el decil tres. A partir del cuarto, los registrados en la Encuesta son mayores y la diferencia se incrementa conforme lo hacen los deciles.

Cuadro 9
Comparación por deciles del ingreso per cápita de los hogares en la EHPM-2001 y el Censo 2000

Deciles	Ingreso per cápita		Diferencia
	Registrado en EHPM-2001	Estimado Censo 2000	
Total de casos^{1/}	969950	959144	
1	14675	18514	-3839
2	22884	25959	-3075
3	30500	32150	-1650
4	39133	38452	681
5	48626	45515	3111
6	61433	54255	7178
7	78266	66260	12006
8	105660	85555	20105
9	170122	125410	44712

1/Para la EHPM-2001 se muestra el número de casos expandido. Para el Censo 2000 es el real.

Fuente: INEC- Archivo de datos Encuesta de Hogares de Propósitos Múltiples 2001 y archivo del Censo 2000.

El por qué de estas diferencias puede tener diversas explicaciones. Sin embargo, la que parece más probable radica en que la EHPM-2001 y el Censo 2000 muestran diferencias en algunas variables clave. Por ejemplo, la proporción de hogares con jefe ocupado y no ocupado difiere levemente entre el Censo y la Encuesta, tal y como se muestra a continuación:

TIPO DE HOGAR	CENSO	EHPM	INGRESO PER CÁPITA MEDIO REGISTRADO EN EHPM 2001
Hogares con jefe ocupado	84.0	88.9	83543
Hogares con jefe no ocupado	16.0	11.1	51341

El origen de esta diferencia puede deberse a un sesgo en la Encuesta, en el Censo o en ambas fuentes de información. Debido a la naturaleza de la EHPM, principalmente en el aspecto de que los entrevistadores reciben una capacitación mayor que en el Censo, puede suponerse que la Encuesta posee los parámetros correctos. De ser así, la menor proporción de hogares con jefe no ocupado en el Censo produciría una subestimación del ingreso per cápita, ya que el ingreso en los hogares cuyo jefe no se encuentra ocupado es significativamente menor al que reportan los hogares con jefe ocupado. Este sesgo puede explicar una subestimación del ingreso en el Censo, pero no el patrón de las diferencias de los deciles que se deduce del Cuadro 9. Para ello, puede

analizarse la distribución porcentual de la categoría ocupacional de los hogares con jefe ocupado en la siguiente tabla:

CATEGORÍA OCUPACIONAL	CENSO	EHPM
Empleado del Estado	7.5	7.2
Empleado Inst Autónoma	7.6	7.3
Empleado Empresa Privada	54.5	47.4
Cuenta Propia	23.9	24.1
Patrono-Socio Activo	5.7	10.9
Servidor doméstico	0.7	3.1
Total	100	100

Puede observarse que el Censo reporta una magnitud de patronos y servidores domésticos menor que la Encuesta. Pareciera que ambos grupos (patronos y servidores domésticos) se “trasladaron” a los empleados de la empresa privada, el que es mayor en el Censo que en la Encuesta. Si la información de la Encuesta es correcta, este sesgo afectaría los extremos de los montos de los deciles que muestra el Censo en el Cuadro 9, ya que los servidores domésticos están ubicados en los ingresos más bajos y los patronos en los más altos. Este hecho estaría produciendo una subestimación del ingreso en los deciles más altos del Censo, debido a la falta de patronos, y una sobreestimación del ingreso en los más bajos, debido a la falta de servidores domésticos en el Censo, lo que explicaría el patrón de las diferencias observadas en el Cuadro 9.

Conviene señalar que, debido a que el objetivo del trabajo es realizar una estratificación con fines estadísticos, las diferencias observadas no afectarán considerablemente los resultados y, por esta razón, se tomó la decisión de evitar la introducción de factores de ponderación o funciones de ajuste que corrigieran los sesgos observados¹⁰.

Para clasificar los hogares por nivel de ingreso se usó el procedimiento K-Medias (SPSS-1999) y se le solicitó que construyera seis agrupaciones para analizar diferentes posibilidades de estratificación. El procedimiento no diferenció por zona de residencia. En otras palabras, aún sabiendo que el ingreso per cápita en la zona rural es más bajo que en la urbana, se prefirió clasificar los hogares sin tomar en cuenta la zona de residencia. Esto hace que los montos de ingreso que designarán lo que es, por ejemplo, bajo, medio o alto nivel de ingreso, son iguales para ambas zonas. Este procedimiento permitirá comparar, posteriormente, los resultados del nivel de ingreso por zona de residencia en términos absolutos. La otra opción es hacer una clasificación independiente en cada zona. Sin embargo, este procedimiento no permitiría comparar los niveles de ingreso por urbano y rural en términos absolutos, ya que los montos de ingreso per cápita que establecerían lo que es, por ejemplo, alto, medio o bajo nivel de ingreso diferirían entre zonas, y, por ello, las comparaciones tendrían que hacerse en términos relativos.

Los resultados del K-Medias para seis agrupaciones se presentan en el Cuadro 10. Se prefirió optar por un número de seis, aunque posteriormente será reducido, para evaluar la efectividad del procedimiento en términos de las diferencias del ingreso per cápita que se producen entre los estratos. Como puede observarse, en el primer estrato se agrupa el 35% de los hogares del país y este porcentaje es menor en la zona urbana (22.8%) que en la rural (53.7%). Por el contrario, en

¹⁰ Si se quisiera usar la ecuación para realizar una estimación precisa del ingreso, se debe tomar en cuenta la limitación mencionada.

el estrato más alto (seis) quedan pocos hogares (1.8% a nivel nacional), y son más en la zona urbana (2.7%) que en la rural (0.4%).

Por otra parte, y como era de esperar, el ingreso per cápita medio de los hogares se incrementa a medida que lo hace el estrato y pasa de 23005 en el de ingresos más bajos hasta 266440 en el de los más altos. La misma relación se encuentra por zona de residencia. Los coeficientes de variación muestran el efecto de la estratificación de los hogares. Si no se estratifica, el coeficiente de variación para el ingreso per cápita del hogar es de 82%. Si se estratifica por zona de residencia, sin tomar en cuenta los seis niveles de ingreso, los coeficientes disminuyen levemente (74.4 y 78.0% en la zona urbana y rural, respectivamente). Sin embargo, cuando se considera la zona de residencia y los seis estratos construidos, los coeficientes de variación disminuyen radicalmente a un promedio de 34% en el de ingresos más bajos y a cifras inferiores a 16% en el resto de los estratos.

Cuadro 10
Resultados del Cluster Analysis para el ingreso per cápita de los hogares del Censo 2000
construyendo seis estratos por zona de residencia

ESTRATOS	URBANO	RURAL	TOTAL	URBANO	RURAL	TOTAL
	NÚMERO DE HOGARES			DISTRIBUCIÓN PORCENTUAL		
Total	580470	378674	959144	100	100	100
1	132461	203456	335917	22.8	53.7	35.0
2	187661	120654	308315	32.3	31.9	32.1
3	130884	36595	167479	22.5	9.7	17.5
4	73673	11768	85441	12.7	3.1	8.9
5	40055	4788	44843	6.9	1.3	4.7
6	15736	1413	17149	2.7	0.4	1.8
	INGRESO PER CÁPITA MEDIO			COEFICIENTE DE VARIACIÓN		
Total	73748	40769	60728	74.4	78.0	82.1
1	24388	22105	23005	30.7	36.2	34.3
2	47887	45805	47072	16.2	16.4	16.4
3	78141	75959	77664	13.4	13.3	13.4
4	121309	119782	121099	12.0	11.9	12.0
5	180443	178588	180245	11.5	11.3	11.5
6	266848	261887	266440	14.5	13.7	14.5

Fuente: INEC-Archivo de datos del Censo 2000

Si el número de estratos se disminuye a tres se dan los resultados del Cuadro 11. Lo que se hizo fue reagrupar los seis estratos del cuadro anterior: al bajo le correspondió el primer estrato y el alto agrupó los estratos cinco y seis. El medio abarcó el resto. El nombre de los estratos como bajo, medio y alto, es el resultado de comparar los ingresos per cápita por estrato, por lo que también pueden denominarse como nivel de ingreso bajo, medio y alto.

Cuadro 11
Resultados de la reagrupación del ingreso per cápita de los hogares del Censo 2000
en tres niveles de ingreso por zona de residencia

ESTRATOS	URBANO	RURAL	TOTAL	URBANO	RURAL	TOTAL
	NÚMERO DE HOGARES			DISTRIBUCIÓN PORCENTUAL		
Total	580470	378674	959144	100	100	100
Bajo	132461	203456	335917	22.8	53.7	35.0
Medio	392218	169017	561235	67.6	44.6	58.5

	55791	6201	61992	9.6	1.6	6.5
	INGRESO PER CÁPITA MEDIO			COEFICIENTE DE VARIACIÓN		
Total	73748	40769	60728	74.4	78.0	82.1
Bajo	24388	22105	23005	30.7	36.2	34.3
Medio	71774	57484	67471	40.7	39.6	41.8
Alto	204814	197569	204089	23.1	21.6	23.0

Fuente: INEC-Archivo de datos del Censo 2000

En el estrato bajo de la zona urbana se concentra el 23% de los hogares y en el mismo, para la zona rural, se agrupa el 54%. Por el contrario, en el estrato alto, para la zona urbana el porcentaje es de 10% y en la rural es apenas de 2%. Los ingresos per cápita promedio varían de acuerdo a lo esperado, ya que se incrementan a medida que lo hace el estrato. Los coeficientes de variación todavía reportan una ganancia importante, ya que el coeficiente en el estrato bajo es 2.4 veces menor si se compara con el total (82.1%), en el estrato medio esta relación es de 2.0 y en el alto de 3.6.

Para evaluar la bondad de la estratificación propuesta, en términos del análisis de los datos censales, se trabajó con una muestra del 10% del Censo 2000 que proporcionó el INEC y se obtuvieron tabulaciones donde se espera que el nivel de ingreso discrimine adecuadamente. Se escogieron variables relacionadas con la vivienda, el seguro social, la inmigración internacional, la fecundidad y el trabajo.

Los resultados muestran, para las variables relacionadas con la vivienda, que la calificación del estado del piso como bueno se incrementa a medida que aumenta el nivel de ingreso, tanto en la zona urbana como en la rural (Gráfico 3). El porcentaje de tenencia de teléfono también se incrementa de acuerdo al nivel del ingreso y esto se cumple en la zona urbana y en la rural. Como puede observarse en el Gráfico 4, mientras que en el estrato de altos ingresos de la zona urbana el porcentaje de hogares con teléfono es de 86%, en el bajo es de 56%. En la zona rural estos porcentajes son 57% y 22%, respectivamente.

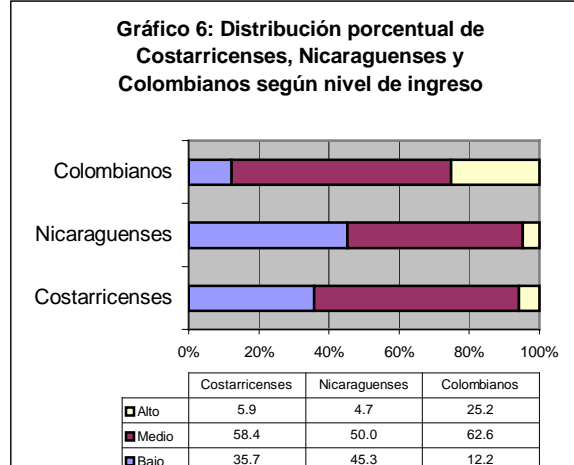
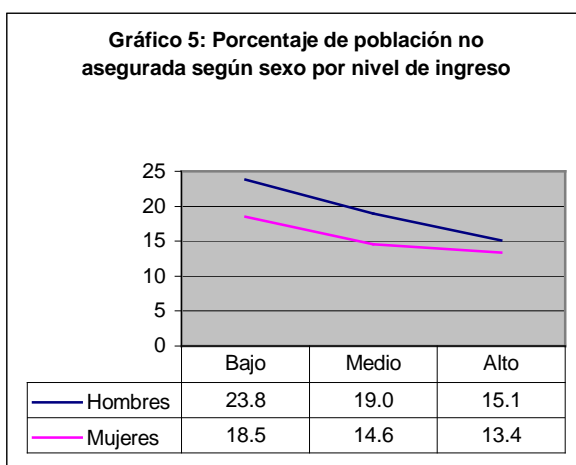
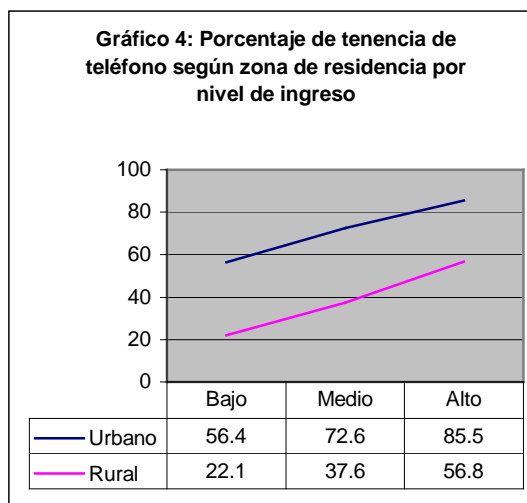
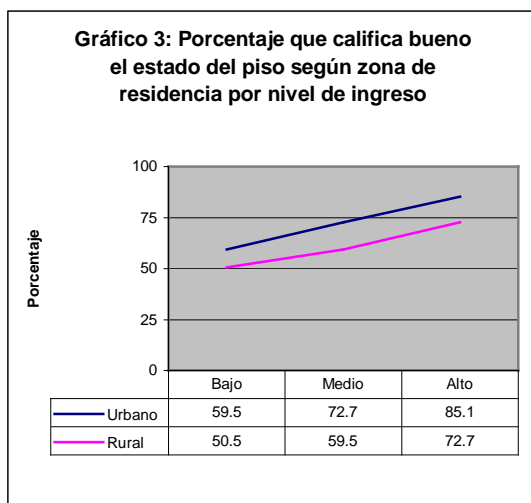
La variable relacionada con el seguro social consideró las personas que no tienen seguro. Este resultado señala que el porcentaje de población no asegurada es mayor en el nivel de ingreso bajo que en el alto, tanto en la zona urbana como rural (Gráfico 5).

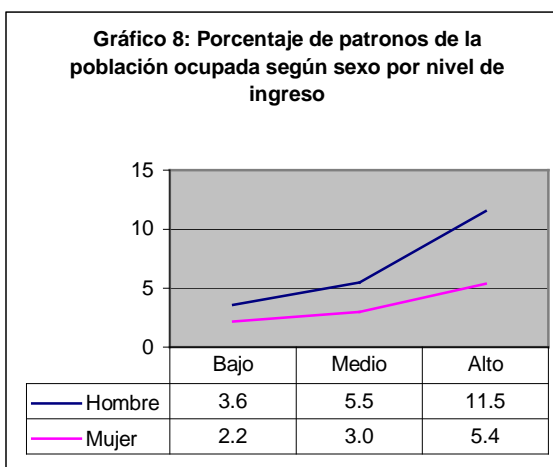
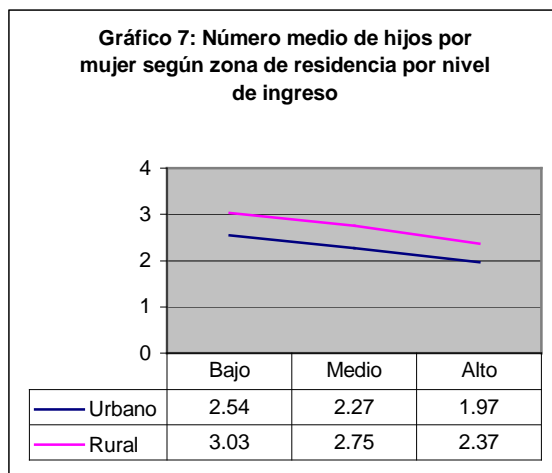
En el Gráfico 6 se presenta la distribución porcentual de costarricenses de acuerdo al nivel de ingreso y se compara con los inmigrantes de Nicaragua y de Colombia. Los resultados muestran que los costarricenses se clasifican principalmente en el nivel de ingreso medio (58.4%), el bajo concentra 36% y el alto 6%. Si se compara con los nicaragüenses, resulta que el porcentaje de estos inmigrantes en el estrato bajo (45.3%) es superior que el de los costarricenses (35.7%), en el estrato medio es menor y en el alto también. Esto significa que los inmigrantes nicaragüenses en el país se encuentran proporcionalmente más concentrados que los costarricenses en bajos niveles de ingreso. Por el contrario, solamente el 12% de los colombianos fue clasificado en el nivel de ingreso bajo y en el alto la proporción (25.2%) es, aproximadamente, 4 veces mayor que la de costarricenses (5.9%) y nicaragüenses (4.7%). Este resultado pone en evidencia las diferencias que existen en la inmigración internacional al país, ya que los nicaragüenses se llegan a

concentrar más que los nacionales en el nivel de ingreso bajo y los colombianos se aglomeran más en el nivel de altos ingresos.

En lo que se refiere a la fecundidad también se observan los patrones esperados, ya que el número medio de hijos vivos es mayor en el nivel de ingreso bajo que en el alto y, además, es más alta en la zona rural que en la urbana (Gráfico 7).

Para la variable relacionada con el trabajo se escogió la proporción de patronos en la población ocupada. Los resultados muestran que el porcentaje de patronos se incrementa a medida que lo hace el nivel de ingreso (Gráfico 8). Esta relación se cumple en ambos sexos, pero es más evidente en los hombres, ya que pasa de 4% en el nivel de ingreso bajo a 11% en el alto, y en las mujeres los mismos porcentajes son de 2% y 5%, respectivamente.





Los resultados encontrados contribuyen a juzgar la efectividad de la estratificación de los hogares por nivel de ingreso y, por ello, el énfasis de la interpretación se hizo en las relaciones encontradas. Por la naturaleza del estudio se evitaron las interpretaciones sustantivas que podrían argumentar causas o efectos de las relaciones encontradas, sin embargo, existe conciencia de la importancia por profundizar en estos temas a futuro.

ESTRATIFICACIÓN DE LOS SEGMENTOS (MMV-2000)

Para obtener el MMV-2000 el personal del INEC acumuló para cada segmento censal, entre otra información relevante, el número de viviendas individuales que reportó el Censo 2000. Cada segmento trae incluida la identificación de provincia, cantón y distrito. Además, posee un código que denota la zona de residencia (urbano-rural) y la región de planificación en la que está ubicado.

Para calcular el ingreso per cápita en cada segmento censal se sumaron los ingresos totales de los hogares y se dividió entre el número de residentes habituales. Al MMV-2000 también se le incluyó información referente a los índices construidos.

Utilizando los segmentos censales como unidades de análisis se calcularon los coeficientes de correlación de Pearson entre los índices y el ingreso per cápita, controlando por zona de residencia, para corroborar la asociación. Los resultados indican que, en general, la correlación de los índices con el ingreso es alta (Cuadro 12). A nivel nacional, el índice de educación, ocupación y artefactos muestran correlaciones muy altas y positivas con el ingreso (de 0.90 o más). La correlación con el hacinamiento y la proporción de dependientes es menos intensa y negativa (-0.635 y -0.555, respectivamente). Para el número de ocupados es moderada y positiva (0.412). La comparación por zona de residencia indica que la dirección y la intensidad de las correlaciones se mantiene. No obstante, las asociaciones son mayores en la zona urbana que en la rural. Estos resultados indican que la intensidad de las correlaciones se fortalece al utilizar el segmento censal como unidad de análisis. Además, se corrobora que la correlación mantiene la dirección esperada.

Cuadro 12
Coefficientes de Correlación de Pearson usando los segmentos como
unidad de análisis

Indíces	Ingreso per cápita		
	Total	Urbano	Rural
(N)	(17257)	(9046)	(8211)
Educación	0.945	0.944	0.887
Ocupación	0.946	0.953	0.846
Artefactos	0.895	0.914	0.850
Hacinamiento	-0.635	-0.748	-0.496
Dependientes	-0.555	-0.521	-0.445
Número de ocupados	0.412	0.320	0.355

Fuente: INEC-Archivo de datos del Marco Muestral de Viviendas Censo 2000

Para evaluar la intensidad de la asociación del ingreso per cápita a nivel de segmento se aprovechó un trabajo anterior realizado por el INEC, en el que se “empataron” los segmentos de la EHPM-2000 con los respectivos del Censo 2000. Aunque la cartografía de la Encuesta no coincide exactamente con la del Censo, se incluyó todo el segmento del Censo, ya sea que coincidiera total o parcialmente. Esta comparación desfavorece el cálculo de la correlación, debido a que la Encuesta incluye en promedio 12 viviendas por segmento y los segmentos censales tienen en promedio 60. Aún así, el coeficiente de correlación de Pearson entre el ingreso per cápita por segmento registrado de la EHPM-2000 y el Censo 2000 fue de 89% para el total del país; en el área urbana fue de 87% y en la rural de 66%.

Para estratificar los segmentos del MMV-2000 se procedió igual que con los hogares. Es decir, se usó el procedimiento K-Medias (SPSS, 1999) y se clasificaron los segmentos en seis estratos para analizar diferentes posibilidades de estratificación. En el Cuadro 13 se muestran los resultados, los que posteriormente fueron reagrupados en tres estratos: a los dos primeros estratos se les designó como nivel de ingreso bajo, al tercero y cuarto como medio y al quinto y sexto como alto.

Los resultados de la reagrupación se muestran en el Cuadro 14, en el que se obtiene que el MMV-2000 está compuesto por 17257 segmentos censales; 9046 en la zona urbana y 8211 en la rural. A nivel nacional, el 58% de los segmentos censales fue clasificado como bajo; en el área urbana fue el 30% y en la rural de 88%. La alta concentración de segmentos en la parte rural era de esperar, ya que los ingresos urbanos son casi dos veces mayores a los rurales, según se había comentado anteriormente.

Los segmentos de estrato medio son el 33% a nivel nacional y la proporción es mayor en la zona urbana que en la rural (52.9 y 11.4%, respectivamente). En el nivel alto está el 9% de los segmentos del país y el porcentaje es de 16.7% en la zona urbana y prácticamente nulo en la rural.

Cuadro 13
Resultados del Cluster Analysis para el ingreso per cápita de los segmentos censales
del MMV-2000 construyendo seis estratos por zona de residencia

Estratos	Número de segmentos	Porcentaje	Ingreso per cápita promedio	Desviación estándar	CV(%)
TOTAL					
1	5102	29.6	27871	5795	20.8

2	4874	28.2	44361	5130	11.6
3	3555	20.6	63665	6085	9.6
4	2163	12.5	86775	7842	9.0
5	1103	6.4	119691	11461	9.6
6	460	2.7	170172	19852	11.7
Total	17257	100.0	56947	32565	57.2

URBANO

1	612	6.8	30550	4390	14.4
2	2137	23.6	45763	5012	11.0
3	2767	30.6	64162	6117	9.5
4	2019	22.3	86977	7846	9.0
5	1064	11.8	119598	11427	9.6
6	447	4.9	170333	19905	11.7
Total	9046	100.0	74400	33965	45.7

RURAL

1	4490	54.7	27506	5867	21.3
2	2737	33.3	43266	4952	11.4
3	788	9.6	61920	5640	9.1
4	144	1.8	83943	7247	8.6
5	39	0.5	122213	12245	10.0
6	13	0.2	164623	17734	10.8
Total	8211	100.0	37718	15891	42.1

Fuente: INEC-Archivo de datos del Marco Muestral de Viviendas del Censo 2000

Cuadro 14
Resultados de la reagrupación del ingreso per cápita de los segmentos censales del MMV-2000 en tres niveles de ingreso por zona de residencia

Estratos	Número de segmentos	Porcentaje	Ingreso per cápita promedio	Desviación estándar	CV(%)
TOTAL					
Bajo	9976	57.8	35927	9898	27.6
Medio	5718	33.1	72407	13111	18.1
Alto	1563	9.1	134548	27168	20.2
Total	17257	100.0	56947	32565	57.2
URBANO					
Bajo	2749	30.4	42376	7992	18.9
Medio	4786	52.9	73787	13212	17.9
Alto	1511	16.7	134607	27304	20.3
Total	9046	100.0	74400	33965	45.7
RURAL					
Bajo	7227	88.0	33474	9440	28.2
Medio	932	11.4	65323	9919	15.2
Alto	52	0.6	132816	23012	17.3
Total	8211	100	37718	15891	42.1

Fuente: INEC-Archivo de datos del Marco Muestral de Viviendas del Censo 2000

Estos resultados sugieren que si bien es cierto en la zona urbana del país se puede trabajar provechosamente con tres niveles de ingreso, en la parte rural son solo necesarios dos (uno bajo y otro medio). También sugieren que la estratificación siempre debería tomar en cuenta la distinción urbano-rural, debido a las diferencias tan marcadas que se encontraron.

Por otra parte, puede observarse que el ingreso per cápita promedio se incrementa a medida que lo hace el nivel del ingreso, tanto en la zona urbana como en la rural. Para medir la efectividad de la estratificación aplicada al marco se procedió a calcular el coeficiente *Etha*. Éste toma en cuenta que el nivel de medición de la variable de control es nominal y que la dependiente es continua. Tiene, además, la bondad de que el cuadrado de *Etha* puede interpretarse como el coeficiente de determinación (R^2) de la regresión múltiple (Freeman, 1971). Los resultados se presentan a continuación:

ZONA	ETHA	ETHA²
Urbano	0.895	0.801
Rural	0.795	0.632

De esta manera se encontró que en la zona urbana la estratificación explica aproximadamente un 80% de la variabilidad del ingreso per cápita y en la zona rural el 63%.

Otro punto de interés se observa en los coeficientes de variación del ingreso per cápita para los estratos. En la zona urbana, los coeficientes para cada nivel de ingreso son aproximadamente del 20% en los tres niveles de ingreso. En la zona rural el rango de variación es mayor (mínimo de 15.2% y un máximo de 28.2%). El punto de interés es que obviando la estratificación por ingreso el coeficiente sería de 46% en la zona urbana y 42% en la rural. De igual manera, si no se toma en cuenta la zona de residencia, el coeficiente para cada nivel de ingreso sería de aproximadamente 33%. Finalmente, si no se toma en consideración la zona de residencia ni el nivel de ingreso, el coeficiente de variación sería de 57%. Estos detalles ponen en evidencia la importancia de controlar el nivel de ingreso en el MMV-2000 para realizar estudios en los que la variable ingreso sea de interés.

En el MMV-2000 el promedio nacional de viviendas individuales por segmento es de 60 y en el estrato bajo es menor (56 viviendas) que en el medio y alto (65 y 64 viviendas, respectivamente) (Cuadro 15). En la zona urbana el promedio por segmento es de 67 y, como es lo usual, es mayor al de la zona rural (52). Por niveles de ingreso, el promedio de viviendas varía más en la zona rural que en la urbana. Esto es más fácil observarlo en los coeficientes de variación, ya que mientras en el sector urbano el coeficiente es de 24% en la rural es de 39%.

Cuadro 15
Resultados de la estratificación del MMV-2000 con respecto al número de viviendas por segmento controlando por zona de residencia

Estratos	Urbano	Rural	Total
Promedio de viviendas por segmento			
Bajo	69	51	56
Medio	67	59	65
Alto	65	57	64
Total	67	52	60
Coefficiente de variación (%)			
Bajo	24.6	39.2	37.5
Medio	22.4	28.8	24.6
Alto	23.1	40.4	23.4

Total	23.9	38.5	31.7
Fuente: INEC-Archivo de datos del Marco Muestral de Viviendas del Censo 2000			

La información del número de viviendas por segmento y su coeficiente de variación, será útil para justificar acciones de revisión del tamaño de los segmentos censales en el MMV-2000. Los resultados muestran que, ante una posible revisión del marco, la zona rural tendría prioridad. Sin embargo, eso dependerá en gran medida de los recursos disponibles. Además, esta información debería complementarse con otros datos del marco para realizar un trabajo mejor fundamentado. Por ejemplo, detectando los segmentos censales cuyo número de viviendas sea extremo. Este procedimiento, que ya se está aplicando en el INEC, está discriminando los segmentos con menos de 40 viviendas o más de 100 en la zona urbana del país y los que tienen menos de 30 o más de 80 en la zona rural, para proceder a fusionarlos, en caso de que sean relativamente pequeños, o a dividirlos, si el número de viviendas es relativamente grande.

CONCLUSIONES

En el trabajo realizado se estimó el ingreso per cápita de los hogares y segmentos del Censo 2000 y éste dato se consideró un indicador del ingreso que fue útil para estratificarlos por niveles de ingreso. Los modelos de regresión obtenidos son considerados aceptables para los propósitos del estudio, tomando en cuenta el cumplimiento de los supuestos, la significancia de los modelos y los coeficientes de determinación. En el tanto los resultados obtenidos se usen con fines de estratificación estadística, se considera que los sesgos observados en la estimación del ingreso no afectan la estratificación propuesta los hogares y segmentos del Censo 2000.

Este trabajo puede verse como el inicio de diversas investigaciones que en el futuro es posible realizar. El indicador del ingreso para los hogares constituye ahora una variable que podría utilizarse para analizar la relación del ingreso con diferentes aspectos contemplados en el Censo 2000. Con respecto al MMV-2000 puede decirse que ahora dispondrá de una variable de estratificación adicional que fortalecerá los diseños muestrales de las encuestas que requieran de esta información. Otras estratificaciones disponibles serán, al menos, la zona de residencia, las regiones de planificación y combinaciones o arreglos por zonas como lo son, por ejemplo, el Gran Área Metropolitana y Resto del País.

El trabajo realizado incluyó la zona rural del país, a diferencia del realizado con la información del Censo de 1984, el que contempló únicamente el área urbana (Madrigal, 1986). Sin embargo, es fácil deducir que la información disponible no es suficiente para estratificar más provechosamente el área rural y que se requiere de otras variables no incluidas en el Censo para mejorar. La decisión de cuáles son estas variables invita a la discusión y es de esperar que la realización de la ENIG-2003 contribuya a este propósito.

AGRADECIMIENTOS

A la Directora de los Censos 2000, Lic. María Elena Gonzalez y al personal a su cargo, particularmente de la Lic. Elizabeth Solano, ya que sin su apoyo e intervención hubiese sido imposible realizar el trabajo propuesto en tan poco tiempo. A la Lic. Giselle Arguello y el Lic. Mario Umaña, destacados en el área de Muestreo del INEC, por su colaboración en el trabajo de diversas maneras. Las sugerencias de la Lic. Floribel Mendez, Coordinadora de la Unidad de Estadísticas Económicas del INEC, fueron un apoyo importante. Al Bach. Iván Villegas por su labor de asistencia. Al Máster Miguel Gómez, miembro del Consejo Directivo del INEC, por las valiosas críticas y sugerencias al trabajo realizado. Al Dr. Edgar Gutiérrez, Director del Observatorio del Desarrollo de la Universidad de Costa Rica, por el interés demostrado en el trabajo y por el tiempo brindado.

BIBLIOGRAFÍA

1. Babbie, Earl. **The Practice of Social Research**. 5 Edition. Wadsworth Publishing Company. USA. 1989. p393
2. Freeman, Linton. **Elementos de Estadística Aplicada**. EUROAMERICA S.A. Madrid, España. 1971. p 160-173.
3. Golder P.A and Yeomans K.A. The Use of Cluster Analysis for Stratification. **Applied Statistics**, Volume 22, Issue 2 (1973), 213-219.
4. Instituto Nacional de Estadística y Censos. **Encuesta de Hogares de Propósitos Múltiples, Julio del 2001**. San José, Costa Rica. Junio del 2002.
5. Instituto Nacional de Estadística y Censos. **Clasificación de Ocupaciones de Costa Rica-2000**. San José, Costa Rica. 2000. Documento basado en la Clasificación Internacional Uniforme de Ocupaciones CIU-88 de la Oficina Internacional del Trabajo.
6. Kish, Leslie. **Survey Sampling**. Wiley Classics Library Edition Published 1995. Copyright by John Wiley & Sons, Inc. United States of America. 1965.
7. Kerbo, Harold. **Estratificación Social y Desigualdad. El conflicto de clases en perspectiva histórica y comparada**. Mc Graw Hill. España, 1998. p20.
8. Kleinbaum, David et. al. **Applied Regression Analysis and Other Multivariable Methods**. Third edition. Duxbury Press. United States of America. 1998.
9. Madrigal, Johnny. **Metodología y Construcción de un Indicador del Ingreso Familiar: Aplicación a los Censos Nacionales de Vivienda y Población 1984**. Asociación Demográfica Costarricense. San José, Costa Rica, 1986.
10. Ministerio de Economía, Industria y Comercio y Dirección General de Estadística y Censos. **Metodología: Encuesta Nacional de Ingresos y Gastos de los Hogares**. San José, Costa Rica. Vol. 2. Sin fecha.
11. Ministerio de Economía, Industria y Comercio, Dirección General de Estadística y Censos y Caja Costarricense de Seguro Social. **Encuesta de Hogares de Propósitos Múltiples. Modulo de Empleo. Julio 1996**. San José, Costa Rica. Febrero, 1997. p48-49
12. Ministerio de Economía, Industria y Comercio, Dirección General de Estadística y Censos y Caja Costarricense de Seguro Social. **Encuesta Nacional de Ingresos y Gastos 1988**. San José, Costa Rica. 1989. Cuadro 12, p27
13. Moore J, Stinson, L and Welniak, J. Income Measurement Error in Surveys: a review. *Journal of Official Statistics*. Vol. 6, No. 4, 2000. pp. 332.
14. SPSS Base 10.0 **Applications Guide**. Copyright 1999 by SPSS Inc. United States of America. p 296.