

COMPARACIÓN DEL NIVEL DE INGRESO DEL JEFE DE HOGAR ENTRE AGLOMERADOS

ANÁLISIS DE DATOS MULTINIVEL

Autores:

María C. Leguizamón

Carlos I. Martínez

Adriana F. Panico

Facultad de Ciencias Económicas – U.N.T.

AÑO 2005

RESUMEN

El análisis multiniveles surgió en las áreas de la educación, la sociología y la demografía; siendo aplicable a diversas situaciones en las que unidades de nivel inferior, o “microunidades”, se imbrican en otras de nivel superior, o “macrounidades”. Este tipo de análisis también es conocido como Análisis de Datos en Panel en estudios económicos. Nos permite estudiar el efecto simultaneo de variables individuales y grupales, y sus respectivas interacciones, sobre una variable dependiente al nivel individual.

Los modelos multinivel son un caso particular de los Modelos Lineales generalizados, y se pueden aplicar a otros sistemas de análisis, como por ejemplo: al análisis de medidas repetidas, de datos categóricos, a los modelos de ecuaciones estructurales, a los procesos de series temporales, a la regresión no-lineal, etc.

El objetivo del presente trabajo es aplicar la metodología del análisis de datos en panel provenientes de la Encuesta Permanente de Hogares (EPH). Comparar el nivel de ingreso de los jefes de hogar entre aglomerados de acuerdo a las características de los mismos y de los hogares considerados en la muestra.

Los datos analizados surgen del aglomerado urbano del Gran Tucumán-Tafí Viejo y resto del NOA, relevados de la EPH continua del primer trimestre de 2005.

Se considera como niveles **MACRO** a los **Aglomerados** y **Hogares** y como nivel **MICRO** a los **Jefes de hogar**.

PALABRAS CLAVES: Análisis Multiniveles, Niveles MICRO y MACRO, EPH.

I. INTRODUCCIÓN

La EPH constituye un programa de alcance nacional orientado a obtener información que permite caracterizar a la población en términos demográficos, ocupacionales, habitacionales y de ingresos, desarrollando un sistema integrado de indicadores sociales.

Originalmente nos habíamos propuesto realizar una comparación del nivel de ingreso de los jefes de hogar entre los aglomerados urbanos San Miguel de Tucumán-Tafí Viejo con los aglomerados urbanos de las otras provincias que constituyen el NOA de acuerdo a las características de los mismo y de los hogares considerados en la muestra. Sin embargo en la EPH fuente no se discriminan los aglomerados urbanos de las provincias de Salta, Jujuy, Catamarca y Santiago del Estero en forma separadas, sino que se identifican todas juntas con el mismo código como “resto del NOA”. De acuerdo a los datos obtenidos de la fuente¹, decidimos considerar sólo dos aglomerados urbanos de importancia San Miguel de Tucumán-Tafí Viejo, con menos de 500.000 habitantes y resto del NOA con más de 500.000 habitantes.

La variable a medir es el Ingreso Total del Individuo (variable respuesta) y el objetivo es analizar si está influenciada por las variables a nivel **Macro**: estrategias del hogar² y tamaño del aglomerado y a nivel **Micro**: estado de actividad, condición de actividad, nivel de educación, edad, sexo, posee obra social y tiene un plan jefa-jefe de hogar.

II. METODOLOGÍA

La primera etapa del estudio requiere un análisis descriptivo de las variables, que nos dará una idea de cómo pueden estar relacionadas con la variable respuesta. Generalmente este análisis descriptivo se realiza a través de medidas y gráficos informativos.

La segunda etapa se refiere a la estimación de algún modelo multinivel que permitirá determinar las variables significativas o que expliquen el comportamiento del ingreso total del jefe de hogar según el aglomerado.

¹ EPH, Encuesta Permanente de Hogares Primer Trimestre de 2005

² Estrategias de obtención de recursos utilizadas por los hogares que provienen de fuentes no laborales.

III. VARIABLES

Las variables consideradas para las regresiones en los distintos niveles fueron:

- **Variable Respuesta (y)**
 - Ingreso Total Individual del Jefe de Hogar (**IngTot.ind**)
- **Variables Explicativas (X)**
 - Ingreso de la Ocupación Principal (**IngOc.Princ**)
 - Nivel de Educación de Jefe de Hogar (**nivel_ed**)
 - Estado de Actividad de Jefe de Hogar (**estado**)
 - Categoría Ocupacional (**cat-ocup**)
 - Edad de Jefe de Hogar (**edad**)
 - Sexo del Jefe de Hogar (**sexo**)
 - Posee un Plan Jefa-jefe de Hogar (**pj**)
 - Estrategias del Hogar (**v1, v2, ..., v19**)
 - Posee Obra Social (**O.Social**)

Donde:

IngTot.ind: Monto de los ingresos (continua)

IngOc.Princ.: Monto del ingreso de la ocupación principal (continua)

nivel_ed: Primaria incompleta = 1; Primaria completa = 2; Secundaria incompleta = 3; Secundaria completa = 4; Sup. Universitaria incompleta = 5; Sup. Universitaria completa = 6; Sin instrucción = 7

estado: Ocupado = 1, Desocupado = 2; Inactivo = 3; ns/nc = 0

cat-ocup: Patrón = 1; Cuenta propia = 2; Obrero/Empleado = 3; Trabajador familiar sin remuneración = 4; ns/nc = 0

edad: Edad de Jefe de Hogar (continua)

sexo: Varón = 1; Mujer = 2

pj: NO = 0; SI = 1

v1, v2, ..., v19: SI = 1; NO = 2

O.Social: TIENE = 1; NO TIENE = 0

IV. ANÁLISIS DESCRIPTIVO

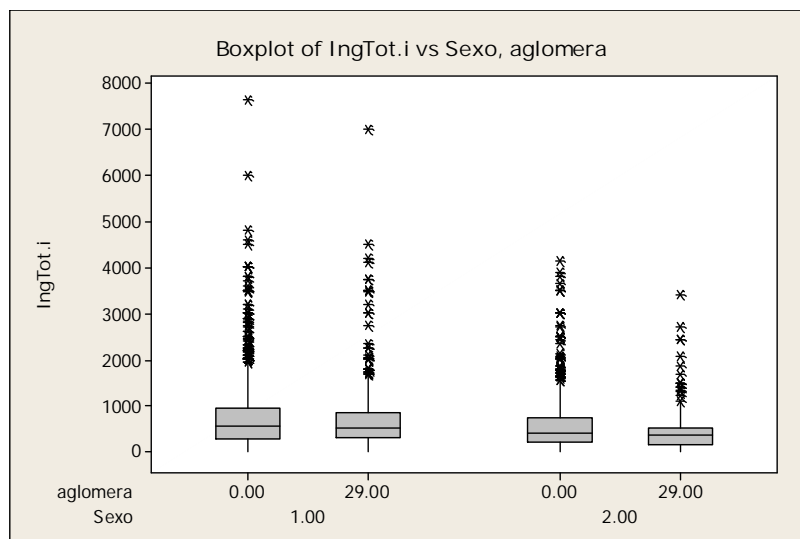
El primer paso en el análisis de datos es realizar un análisis descriptivo de la muestra. Este análisis nos permitirá controlar la presencia de posibles errores en la fase de introducción de los datos, es decir, detectaremos valores fuera de rango, o la presencia de valores perdidos. Este análisis inicial también nos proporcionará una idea de la *forma* que tienen los datos: su posible distribución de probabilidad con sus parámetros de centralización; media, mediana y moda; así como sus parámetros de dispersión; varianza, desviación típica, etc.

En los trabajos estadísticos en general, la pregunta de investigación es la base para el análisis. A partir de allí podemos recurrir a elementos y/o técnicas que nos permitirán determinar por dónde empezar nuestra investigación.

Como se planteó en la introducción, consideramos dos aglomerados urbanos de importancia dentro de la región NOA a nivel macro y variables de relevancia a nivel micro como: sexo, edad, estado de actividad, ingresos de la ocupación principal y otras, como posibles explicativas del ingreso total que percibe el jefe de hogar tipo según el aglomerado.

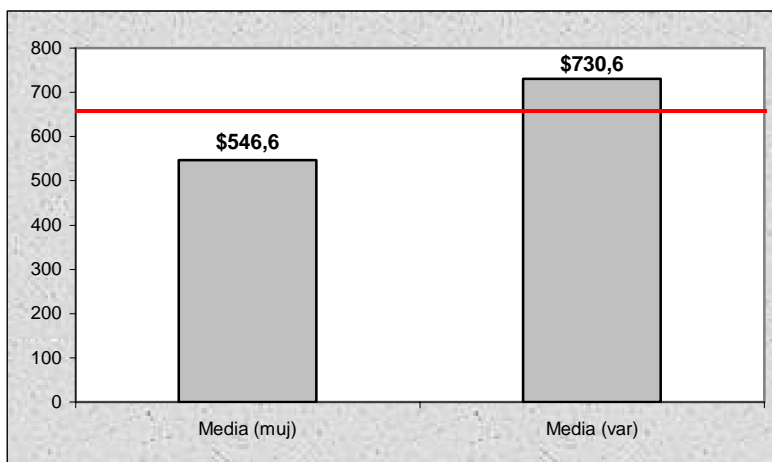
La figura 1 muestra un diagrama de caja del ingreso total individual por aglomerado y sexo. Se observa que el ingreso promedio de los hombres y de las mujeres de ambos aglomerados es similar, pero el ingreso medio de las mujeres es inferior a la de los hombres.

Figura 1: Ingreso Total Individual por Aglomerado y Sexo



El histograma de la figura 2 muestra que la media general del ingreso total individual es de 664,50 pesos mensuales para la región NOA. A partir de allí se observa que la media de los ingresos de los hombres es superior y la de las mujeres notablemente inferior a la media general, situación que se repite en ambos aglomerados.

Figura 2: Histograma de ingresos promedios



Las variables correspondientes al estado de actividad y a la categoría de ocupación del jefe de hogar podrían explicar de alguna manera el comportamiento de los ingresos individuales. Las figuras 3 y 4 muestran los histogramas de las distribución de dichas variables respectivamente.

Figura 3: Histograma del Estado de actividad

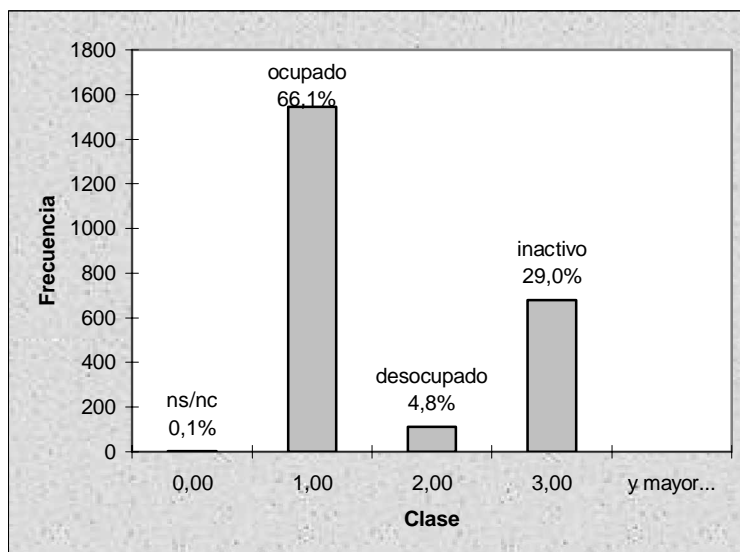
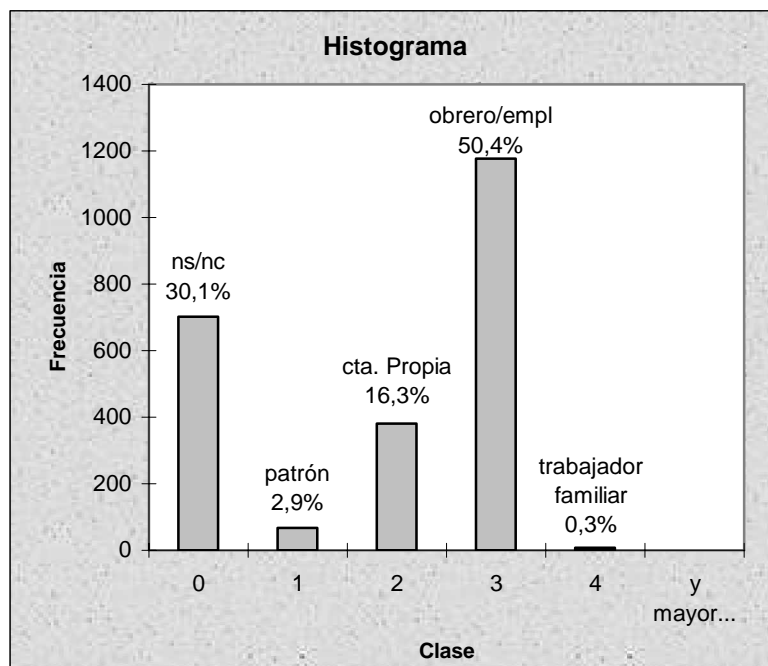


Figura 4: Histograma del Estado de Actividad por Aglomerado



Al comparar el estado de actividad del jefe de hogar por aglomerado observamos que el aglomerado Tucumán-Tafí Viejo tienen el 26,2% de ocupados con respecto al total.

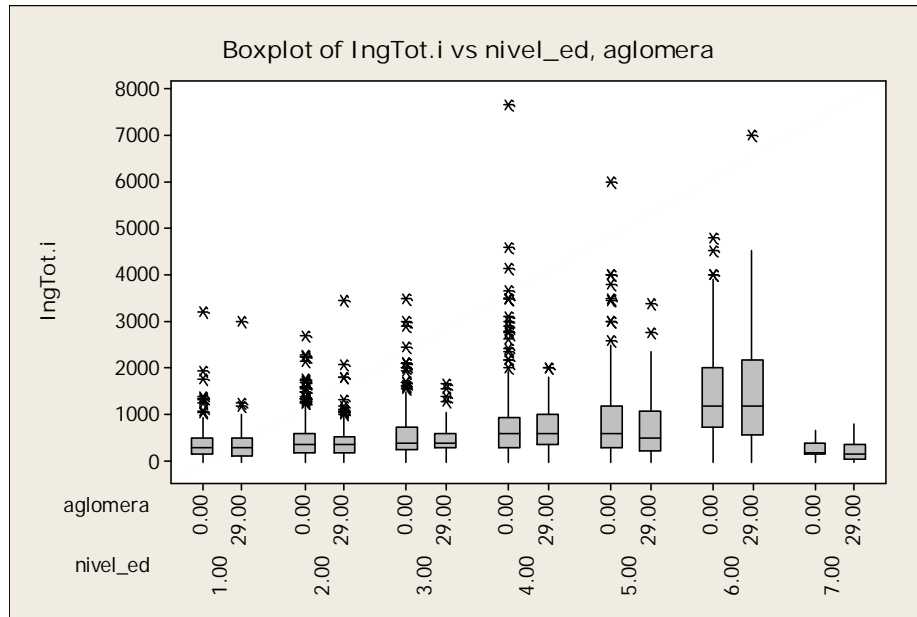
Figura 5: Histograma de la Categoría de ocupación



En la medición correspondiente al primer trimestre de 2005 se puede apreciar que el 50,4% de los jefes de hogar es obrero o empleado en relación de dependencia,

independientes (patrón y cuenta propia) sólo constituyen el 19% de la muestra y un alto porcentaje (30,1%) no responde a la pregunta.

Figura 6: Diagrama de cajas del Ingreso Total Individual por Aglomerado y Nivel de Educación



Podemos observar que los ingresos promedios según el nivel de educación no varían por aglomerado, es decir que existen similares características en Tucumán y el resto del NOA..

V. MODELOS ESTIMADOS

En la segunda etapa del trabajo procedemos a estimar modelos de regresión para determinar concretamente las variables que explican significativamente el porque del nivel de ingreso total individual de un jefe de hogar típico para la región NOA.

Definiendo X como el vector de variables explicativas en el modelo $X=(\text{sexo}, \text{edad}, \text{nivel_ed}, \text{estado}, \text{cat_ocup}, \text{IngOc.Prin})$ se realizaron las regresiones para estimar IngTot.ind utilizando:

Regresión lineal longitudinal (ingenua): en ella no se tiene en cuenta la estructura multinivel del problema planteado, agrupando las observaciones al nivel aglomerado. Estima por Mínimos Cuadrados Generalizados (efectos aleatorios)

Análisis multinivel de regresión lineal de efectos mixtos usando xtmixed: estos modelos permiten considerar efectos aleatorios y fijos tanto en el intercepto como en las pendientes. La estimación se realiza por optimización EM.

Se incluyeron al comienzo todas las variables propuestas como explicativas y se fueron eliminando de acuerdo al nivel de significación de los coeficientes y a la comparación de las verosimilitudes de los modelos con y sin la variable. Las regresiones se estimaron con el programa STATA v. 9.0.

VI. RESULTADOS

Regresión 1: Lineal longitudinal

```
xtreg ingtot_ind sexo edad nivel_ed estado cat_ocup ingoc_princ v9 v8 v10, re
```

```
Random-effects GLS regression      Number of obs      =      2337
Group variable (i): aglomerado_a   Number of groups   =         2
```

```
R-sq:  within = 0.6367              Obs per group:  min =         633
        between = 1.0000              avg =       1168.5
        overall = 0.6374              max =       1704
```

```
Random effects u_i ~ Gaussian      Wald chi2(9)       =      4089.89
corr(u_i, X) = 0 (assumed)         Prob > chi2        =         0.0000
```

ingtot_ind	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
sexo	-20.65951	19.26979	-1.07	0.284	-58.4276	17.10858
edad	8.095138	.6469411	12.51	0.000	6.827156	9.363119
nivel_ed	71.36662	5.946574	12.00	0.000	59.71155	83.02169
estado	207.1073	24.8926	8.32	0.000	158.3187	255.896
cat_ocup	41.79788	16.11625	2.59	0.009	10.21061	73.38515
ingoc_princ	1.054938	.0202361	52.13	0.000	1.015276	1.0946
v9	-320.9974	123.8622	-2.59	0.010	-563.7628	-78.23208
v8	-301.8005	55.59107	-5.43	0.000	-410.757	-192.8441
v10	182.2141	427.6092	0.43	0.670	-655.8846	1020.313
_cons	127.2771	903.4979	0.14	0.888	-1643.546	1898.1
sigma_u	0					
sigma_e	426.9938					
rho	0	(fraction of variance due to u_i)				

Todas las variables explicativas son significativas, excepto **sexo**. Este resultado contradice el análisis descriptivo donde observamos claramente (figura 2) que el ingreso medio de hombres y mujeres difieren por arriba y por debajo de la media general, respectivamente. Otra variable que resultó no significativa es **v10** (intereses, rentas de plazos fijos o inversiones cobradas- Estrategias del Hogar).

Probamos una segunda regresión eliminando las variables no significativas y no observamos ninguna mejora en el ajuste ni cambio en los parámetros estimados.

Regresión 2: Lineal de efectos mixtos con todas las variables

```
xtmixed ingtot_ind sexo edad nivel_ed estado ingoc_princ v8 v9 v10, ||
aglomerado_a:, covariance(identity)
```

Note: single-variable random-effects specification; covariance structure set to identity

Performing EM optimization:

Performing gradient-based optimization:

```
Mixed-effects REML regression          Number of obs      =      2337
Group variable: aglomerado_a           Number of groups   =         2

Obs per group: min =         633
                  avg =      1168.5
                  max =      1704
```

```
Log restricted-likelihood = -17441.306      Wald chi2(8)       =      4070.05
                                           Prob > chi2        =       0.0000
```

ingtot_ind	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
sexo	-19.43308	19.2867	-1.01	0.314	-57.23431 18.36815
edad	7.886358	.6429219	12.27	0.000	6.626254 9.146462
nivel_ed	71.31731	5.95689	11.97	0.000	59.64202 82.9926
estado	152.4214	13.17764	11.57	0.000	126.5937 178.2491
ingoc_princ	1.049724	.0201594	52.07	0.000	1.010213 1.089236
v8	-301.2771	55.66286	-5.41	0.000	-410.3743 -192.1799
v9	-315.0261	123.9839	-2.54	0.011	-558.0301 -72.02198
v10	170.4185	428.0777	0.40	0.691	-668.5983 1009.435
_cons	314.7233	901.5319	0.35	0.727	-1452.247 2081.693

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]
aglomerado_a: Identity			
sd(_cons)	8.622736	22.44807	.05244 1417.842
sd(Residual)	427.515	6.266684	415.4073 439.9756

```
LR test vs. linear regression: chibar2(01) =      0.06 Prob >= chibar2 = 0.4067
```

Considerando la regresión con efectos mixtos, nuevamente **sexo** y **v10** no son significativas, además no es posible incluir la variable **cat_ocup** porque el modelo no converge.

Regresión 3: Lineal de efectos mixtos con las variables estadísticamente significativas

```
xtmixed ingtot_ind edad nivel_ed estado ingoc_princ v8 v9, || aglomerado_a:,  
covariance(independent)
```

Note: single-variable random-effects specification; covariance structure set to identity

Performing EM optimization:

Computing standard errors:

Mixed-effects REML regression Number of obs = 2337
Group variable: aglomerado_a Number of groups = 2

Obs per group: min = 633
 avg = 1168.5
 max = 1704

Log restricted-likelihood = -17452.761 Wald chi2(6) = 4070.86
 Prob > chi2 = 0.0000

ingtot_ind	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
edad	7.884063	.642629	12.27	0.000	6.624533	9.143593
nivel_ed	70.75432	5.923429	11.94	0.000	59.14461	82.36402
estado	150.9418	13.07374	11.55	0.000	125.3178	176.5659
ingoc_princ	1.053101	.019885	52.96	0.000	1.014128	1.092075
v8	-301.4311	55.6524	-5.42	0.000	-410.5078	-192.3544
v9	-309.5421	123.8427	-2.50	0.012	-552.2694	-66.81491
_cons	621.8556	276.5066	2.25	0.025	79.91273	1163.799

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
aglomerado_a: Identity				
sd(_cons)	7.46933	24.21425	.0129968	4292.666
sd(Residual)	427.448	6.263032	415.3472	439.9012

LR test vs. linear regression: chibar2(01) = 0.03 Prob >= chibar2 = 0.4262

Eliminando las variables no significativas observamos una leve mejora en el ajuste y diferencias pequeñas en las estimaciones de los coeficientes de las variables explicativas, cuyo nivel de significación no se ve afectado.

VI. CONCLUSIONES

A partir del análisis descriptivo observamos que comparando el ingreso promedio del Jefe de Hogar perteneciente al aglomerado urbano San Miguel de Tucumán-Tafí Viejo con el ingreso promedio del Jefe de Hogar del resto del NOA, presentan diferencias significativas según el sexo o según el nivel de educación.

- § Los hombres tienen un ingreso promedio de un 33,6% más que el ingreso promedio de las mujeres.
- § Los ingresos promedio para los Jefes de Hogar con nivel de educación universitario completo es notablemente superior al resto.
- § Analizando el estado de actividad, el 66% de los jefes de hogar están ocupados, 4,8% desocupados y 29 % permanece inactivo.
- § Dentro de la categoría ocupados debemos distinguir que el 50,4% es asalariado, el 19% es independiente y el 30,1% no responde a la pregunta. ¿Podemos plantear que este porcentaje corresponde a trabajadores en “negro”?

Desde el punto de vista del análisis multinivel podemos determinar que:

- § El sexo del Jefe de Hogar deja de ser una variable estadísticamente significativa.
- § El nivel de educación se mantiene fuertemente significativo.
- § De las variables referidas a las estrategias de obtención de recursos utilizadas por los hogares que provienen de fuentes no laborales, resultaron significativas la que considera los beneficios y ganancias de negocios en el que no trabaja el Jefe de Hogar (**V9**) y la que incluye los alquileres de las propiedades del Jefe de Hogar (**V8**). La **V10** que considera intereses, rentas de plazos fijos e inversiones no resultó significativa.
- § El ingreso de la ocupación principal es otro factor determinante en el comportamiento del ingreso total individual.
- § El estado de actividad del Jefe de Hogar es determinante en la variable respuesta, aunque debemos hacer notar que el 66% de los jefes de hogar están ocupados.

Reflexión Final

En este trabajo nos propusimos analizar el ingreso total individual del jefe de hogar desde dos perspectivas diferentes: la del campo individual o estructural de cada trabajador (edad, sexo, nivel de educación, etc.) y la del campo contextual (estrategias de subsistencia del hogar, lugar de residencia). La importancia de centrarse en la problemática de los ingresos está relacionada con la creciente inequidad que presenta la distribución del ingreso causada, en parte, por el cambio del perfil productivo del país. Situación que hizo que en la década de los 90 el mercado laboral se vea profundamente alterado por la precarización de los puestos de trabajo, el ascenso del desempleo y la aparición de nuevas modalidades de contratación.

Se consideró que dos variables estructurales, como sexo y nivel de educación, son factores importantes en la determinación del ingreso total del jefe de hogar, esta hipótesis coincide con el análisis descriptivo pero, al ajustar un modelo de efectos mixtos no se puede sostener lo mismo con respecto a la variable sexo. Este resultado plantea una pregunta ***¿si una variable es estadísticamente no significativa podemos eliminarla del análisis cuando en la realidad se evidencia la contrario?*** Específicamente en el tema del rol de la mujer en el mercado laboral se vio que, históricamente, se desempeñó como sustituto del hombre, aceptando trabajos de menor calificación y menores sueldos puesto que la maternidad y obligaciones familiares (tradicionalmente asignadas a la mujer) la hicieron más vulnerable a la hora de discutir salarios.

BIBLIOGRAFÍA

- § **De Angelis, Carlos F.** “Utilización de la técnica del Análisis Factorial de la Varianza para la determinación de los Factores que influyen la variable Ingreso”. En torno de las metodologías. Abordajes cualitativo y cuantitativo.
- § **Twisk, Jos W. R.** “Applied Longitudinal Data Analysis for Epidemiology”. A Practical Guide. Cambridge University Press.