

**FLACSO / Facultad Latinoamericana de Ciencias Sociales**

**Maestría en Ciencias Políticas & Sociología**

**Materia: Metodología de la Investigación Social**

**Profesores: Agustín Salvia y Eduardo Donza**

**Trabajo Práctico N° 4**

**Ejercicio 1**

Con el fin de analizar la diversidad de ingresos de los asalariados del Gran Buenos Aires en el 2° trimestre del 2010, se propone considerar como factores explicativos el nivel educativo, la edad y la antigüedad laboral. Con este fin, se plantea un modelo de regresión lineal múltiple que presenta como variable dependiente el ingreso mensual y como variables independientes la cantidad de años de educación aprobados, la edad y la antigüedad laboral.

En la técnica de regresión se utilizó la modalidad de integración de variables independientes en pasos sucesivos por lo que se presentarán tres modelos: con una, dos y tres variables independientes, respectivamente.

Los resultados del procesamiento respectivo están a continuación. Desarrolle el siguiente análisis justificando las respuestas:

1. Con el fin de alcanzar la parsimonia, elegir entre los tres modelos presentados el que optimice el mayor porcentaje de explicación con la menor cantidad de variables posibles (considerar el  $r$  cuadrado, la significancia de la tabla de anova y la cantidad de variables). De ser necesario proponer un modelo que limite la cantidad de variables independientes planteadas originalmente.
2. Para el modelo elegido, analizar la correlación entre el conjunto de variables independientes y la dependiente por medio del  $r$  de Pearson, del coeficiente de determinación y del coeficiente de indeterminación. En función de esto, sacar conclusiones con respecto a la capacidad del modelo para explicar los ingresos (determinar qué proporción de la varianza de la variable dependiente es determinada conjuntamente por las varianzas de las variables independientes).

3. Construir la ecuación de predicción en función de los coeficientes observados. Además, ¿con qué confianza se puede decir que el modelo de regresión propuesto da cuenta de la distribución empírica de los datos?
4. Realizar la predicción del salario de una persona de 40 años de edad, 12 años de educación y 10 años de antigüedad laboral (utilizar solamente los valores necesarios según el modelo elegido).

**Variables introducidas/eliminadas<sup>a</sup>**

Modelo	Variables introducidas	Variables eliminadas	Método
1	Años de educación aprobados	.	Por pasos (criterio: Prob. de F para entrar $\leq$ ,050, Prob. de F para salir $\geq$ ,100).
2	Edad	.	Por pasos (criterio: Prob. de F para entrar $\leq$ ,050, Prob. de F para salir $\geq$ ,100).
3	Antigüedad laboral	.	Por pasos (criterio: Prob. de F para entrar $\leq$ ,050, Prob. de F para salir $\geq$ ,100).

a. Variable dependiente: Monto de ingreso de la ocupación principal percibido en ese mes

Fuente: elaboración propia en base a datos de EPH-INDEC.

**Resumen del modelo**

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,374 <sup>a</sup>	,140	,139	1612,050
2	,442 <sup>b</sup>	,196	,195	1558,871
3	,449 <sup>c</sup>	,201	,201	1553,640

a. Variables predictoras: (Constante), Años de educación aprobados

b. Variables predictoras: (Constante), Años de educación aprobados, Edad

c. Variables predictoras: (Constante), Años de educación aprobados, Edad, Antigüedad laboral

Fuente: elaboración propia en base a datos de EPH-INDEC.

ANOVA<sup>d</sup>

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	1,234E9	1	1,234E9	475,038	,000 <sup>a</sup>
	Residual	7,602E9	2925	2598704,816		
	Total	8,837E9	2926			
2	Regresión	1,730E9	2	8,651E8	355,997	,000 <sup>b</sup>
	Residual	7,106E9	2924	2430077,680		
	Total	8,837E9	2926			
3	Regresión	1,780E9	3	5,934E8	245,840	,000 <sup>c</sup>
	Residual	7,056E9	2923	2413796,916		
	Total	8,837E9	2926			

a. Variables predictoras: (Constante), Años de educación aprobados

b. Variables predictoras: (Constante), Años de educación aprobados, Edad

c. Variables predictoras: (Constante), Años de educación aprobados, Edad, Antigüedad laboral

d. Variable dependiente: Monto de ingreso de la ocupación principal percibido en ese mes

Fuente: elaboración propia en base a datos de EPH-INDEC.

Coeficientes<sup>a</sup>

Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	48,602	97,175		,500	,617
	Años de educación aprobados	177,613	8,149	,374	21,795	,000
2	(Constante)	-1239,049	130,224		-9,515	,000
	Años de educación aprobados	187,767	7,912	,395	23,731	,000
	Edad	31,097	2,177	,238	14,283	,000
3	(Constante)	-1186,642	130,296		-9,107	,000
	Años de educación aprobados	180,843	8,031	,381	22,518	,000
	Edad	32,420	2,189	,248	14,808	,000
	Antigüedad laboral	-5,077	1,115	-,077	-4,552	,000

a. Variable dependiente: Monto de ingreso de la ocupación principal percibido en ese mes

Fuente: elaboración propia en base a datos de EPH-INDEC.

## Ejercicio 2

En una investigación relativa a las estrategias asociativas de los trabajadores no asalariados (considerando a los cuentapropistas y a los patrones o empleadores) se sostiene que son los ocupados en el sector informal los que presentan un mayor grado de asociatividad. Además, se presume que el sexo de los trabajadores incide en esta relación.

Para verificarlo se obtuvieron datos provenientes de la Encuesta Permanente de Hogares del Gran Buenos Aires del 2º semestre de 2010 cuyos procesamientos se presentan a continuación. En función de ellos:

1. Determinar un modelo de análisis multivariado Lazarsfeld que involucre las variables analizadas. Determine el rol de cada una de ellas.
2. Analice los marginales y los parciales considerando el coeficiente de asociación y test de significancia de ji-cuadrado (chi-cuadrado). Sacar conclusiones con respecto al modelo.
3. Realizar una lectura de los porcentajes que se consideran pertinentes. Sacar conclusiones.
4. Analizar en conjunto los resultados expresados anteriormente en función de corroborar o refutar la hipótesis.

Relación originaria:

**Tabla de contingencia ¿En ese negocio/empresa/actividad ¿tiene socios o familiares asociados? \* Sector de Inserción**

% dentro de Sector de Inserción

		Sector de Inserción		Total
		Sector Formal	Sector Informal	
¿En ese negocio/empresa/actividad ¿tiene socios o familiares asociados?	Sí	38,5%	23,8%	26,2%
	No	61,5%	76,2%	73,8%
Total		100,0%	100,0%	100,0%

Fuente: elaboración propia en base a datos de EPH-INDEC.

**Pruebas de chi-cuadrado**

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	12,708 <sup>a</sup>	1	,000		
Corrección por continuidad <sup>b</sup>	11,955	1	,001		
Razón de verosimilitudes	11,920	1	,001		
Estadístico exacto de Fisher				,001	,000
Asociación lineal por lineal	12,692	1	,000		
N de casos válidos	817				

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 35,36.

b. Calculado sólo para una tabla de 2x2.

Fuente: elaboración propia en base a datos de EPH-INDEC.

**Medidas simétricas**

		Valor	Sig. aproximada
Nominal por nominal	Phi	,125	,000
	V de Cramer	,125	,000
N de casos válidos		817	

Fuente: elaboración propia en base a datos de EPH-INDEC.

Parciales:

**Tabla de contingencia ¿En ese negocio/empresa/actividad ¿tiene socios o familiares asociados? \* Sector de Inserción \* Sexo**

% dentro de Sector de Inserción

Sexo			Sector de Inserción		Total
			Sector Formal	Sector Informal	
Varón	¿En ese negocio/empresa/actividad ¿tiene socios o familiares asociados?	Sí	44,4%	21,9%	25,3%
		No	55,6%	78,1%	74,7%
	Total		100,0%	100,0%	100,0%
Mujer	¿En ese negocio/empresa/actividad ¿tiene socios o familiares asociados?	Sí	27,8%	27,6%	27,6%
		No	72,2%	72,4%	72,4%
	Total		100,0%	100,0%	100,0%

Fuente: elaboración propia en base a datos de EPH-INDEC.

**Pruebas de chi-cuadrado**

Sexo		Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Varón	Chi-cuadrado de Pearson	18,544 <sup>a</sup>	1	,000		
	Corrección por continuidad <sup>b</sup>	17,368	1	,000		
	Razón de verosimilitudes	16,835	1	,000		
	Estadístico exacto de Fisher				,000	,000
	Asociación lineal por lineal	18,509	1	,000		
	N de casos válidos	538				
Mujer	Chi-cuadrado de Pearson	,001 <sup>c</sup>	1	,974		
	Corrección por continuidad <sup>b</sup>	,000	1	1,000		
	Razón de verosimilitudes	,001	1	,974		
	Estadístico exacto de Fisher				1,000	,548
	Asociación lineal por lineal	,001	1	,974		
	N de casos válidos	279				

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 20,48.

b. Calculado sólo para una tabla de 2x2.

c. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 14,90.

Fuente: elaboración propia en base a datos de EPH-INDEC.

**Medidas simétricas**

Sexo			Valor	Sig. aproximada
Varón	Nominal por nominal	Phi	,186	,000
		V de Cramer	,186	,000
	N de casos válidos		538	
Mujer	Nominal por nominal	Phi	,002	,974
		V de Cramer	,002	,974
	N de casos válidos		279	

Fuente: elaboración propia en base a datos de EPH-INDEC.

Marginal 1:

**Tabla de contingencia Sector de Inserción \* Sexo**

% dentro de Sexo

		Sexo		Total
		Varón	Mujer	
Sector de Inserción	Sector Formal	15,1%	19,4%	16,5%
	Sector Informal	84,9%	80,6%	83,5%
Total		100,0%	100,0%	100,0%

Fuente: elaboración propia en base a datos de EPH-INDEC.

**Pruebas de chi-cuadrado**

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	2,462 <sup>a</sup>	1	,117		
Corrección por continuidad <sup>b</sup>	2,160	1	,142		
Razón de verosimilitudes	2,414	1	,120		
Estadístico exacto de Fisher				,136	,072
Asociación lineal por lineal	2,459	1	,117		
N de casos válidos	817				

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 46,10.

b. Calculado sólo para una tabla de 2x2.

Fuente: elaboración propia en base a datos de EPH-INDEC.

**Medidas simétricas**

		Valor	Sig. aproximada
Nominal por nominal	Phi	-,055	,117
	V de Cramer	,055	,117
N de casos válidos		817	

Fuente: elaboración propia en base a datos de EPH-INDEC.

## Marginal 2:

**Tabla de contingencia ¿En ese negocio/empresa/actividad ¿tiene socios o familiares asociados? \* Sexo**

% dentro de Sexo

		Sexo		Total
		Varón	Mujer	
¿En ese negocio/empresa/actividad ¿tiene socios o familiares asociados?	Sí	25,4%	27,6%	26,2%
	No	74,6%	72,4%	73,8%
Total		100,0%	100,0%	100,0%

Fuente: elaboración propia en base a datos de EPH-INDEC.

### Pruebas de chi-cuadrado

	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi-cuadrado de Pearson	,453 <sup>a</sup>	1	,501		
Corrección por continuidad <sup>b</sup>	,347	1	,556		
Razón de verosimilitudes	,450	1	,502		
Estadístico exacto de Fisher				,503	,277
Asociación lineal por lineal	,452	1	,501		
N de casos válidos	818				

a. 0 casillas (.0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 72,99.

b. Calculado sólo para una tabla de 2x2.

Fuente: elaboración propia en base a datos de EPH-INDEC.

### Medidas simétricas

		Valor	Sig. aproximada
Nominal por nominal	Phi	-,024	,501
	V de Cramer	,024	,501
N de casos válidos		818	

Fuente: elaboración propia en base a datos de EPH-INDEC.

## Ejercicio 3

Se presupone que en los departamentos de la provincia de Santa Fe el nivel del salario industrial depende de la capacidad de las empresas para generar valor agregado y que en esta relación puede influir el tamaño del departamento. De este hecho surge la siguiente hipótesis: “En el año 2003, los departamentos de la provincia de Santa Fe que poseen industrias con mayor valor agregado por ocupado también presentan un mayor promedio de retribuciones a los asalariados industriales. Incide en esta relación la cantidad de locales industriales del departamento”.

Se pone a prueba esta hipótesis con los resultados de la correlación parcial que están a continuación. Desarrolle el siguiente análisis justificando las respuestas:

1. Analizar la correlación entre las dos variables que componen la relación original por medio del r de Pearson, del coeficiente de determinación y del coeficiente de indeterminación.
2. Analizar la correlación parcial por medio del r de Pearson, del coeficiente de determinación y del coeficiente de indeterminación. Sacar conclusiones con respecto a la hipótesis propuesta. ¿Realmente incide en la relación original la cantidad de locales industriales?

Correlaciones			Valor agregado por ocupado	Remuneración media de asalariados	Cantidad de locales
-ninguno <sup>a</sup>	Valor agregado por ocupado	Correlación	1,000	,681	,043
		Significación (bilateral)	.	,001	,861
		gl	0	17	17
	Remuneración media de asalariados	Correlación	,681	1,000	,160
		Significación (bilateral)	,001	.	,513
		gl	17	0	17
	Cantidad de locales	Correlación	,043	,160	1,000
		Significación (bilateral)	,861	,513	.
		gl	17	17	0
Cantidad de locales	Valor agregado por ocupado	Correlación	1,000	,684	
		Significación (bilateral)	.	,002	
		gl	0	16	
	Remuneración media de asalariados	Correlación	,684	1,000	
		Significación (bilateral)	,002	.	
		gl	16	0	

a. Las casillas contienen correlaciones de orden cero (de Pearson).

Fuente: elaboración propia en base a datos de INDEC / IPEC - Censo Nacional Económico 2004/2005.