

FLACSO / Facultad Latinoamericana de Ciencias Sociales

Maestría en Ciencias Políticas & Sociología

Materia: Metodología de la Investigación Social

Profesores: Agustín Salvia y Eduardo Donza

Unidad II
Ejercitación N° 1

Ejercicio 1

A continuación se presentan indicadores univariados de la condición de actividad y de los ingresos laborales de la población del Gran Buenos Aires en el 2º trimestre de 2010.

En función de estos:

1. Identificar el nivel de medición de cada una de las variables.
2. Realizar una lectura de la distribución de frecuencias (sólo para la condición de actividad) y de las medidas de tendencia central y de dispersión. (Delimitar el análisis a las medidas pertinentes al nivel de medición de cada una de las variables)

Condición de actividad					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	ocupado	5812372	44,6	44,6	44,6
	desocupado	513188	3,9	3,9	48,6
	inactivo	4803427	36,9	36,9	85,4
	menor de 10 años	1899218	14,6	14,6	100,0
	Total	13028205	100,0	100,0	

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

Estadísticos

Condición de actividad

N	Válidos	13028205
	Perdidos	0
Media		2,21
Mediana		3,00
Moda		1
Desv. típ.		1,163
Varianza		1,352
Rango		3
Mínimo		1
Máximo		4
Percentiles	25	1,00
	50	3,00
	75	3,00

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

Estadísticos

'Ingresos Laborales incluye bonif no habituales'

N	Válidos	5763451
	Perdidos	0
Media		2367,6678
Mediana		2000,0000
Moda		2000,00
Desv. típ.		2018,74535
Varianza		4075332,799
Rango		29970,00
Mínimo		30,00
Máximo		30000,00
Percentiles	25	1100,0000
	50	2000,0000
	75	3000,0000

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

Ejercicio 2

En una investigación relativa a las estrategias de búsqueda de trabajo desarrolladas por los desocupados se plantea que el sexo incide en la tendencia de las personas a presentarse espontáneamente en un establecimiento para solicitar un empleo.

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

1. Definir una hipótesis que relacione las dos variables analizadas.

2. Realizar una lectura de los porcentajes que se consideran pertinentes. Sacar conclusiones.
3. Seleccione, fundamentando el criterio adoptado, el coeficiente de asociación más adecuado. Interprete el resultado.
4. Sacar conclusiones respecto al resultado del test de hipótesis de independencia estadística de ji cuadrado.
5. Analizar en conjunto los resultados expresados anteriormente en función de corroborar o refutar la hipótesis.

Tabla de contingencia Se presentó en establecimientos en la búsqueda de trabajo* Sexo

			Sexo		Total
			Varón	Mujer	
Se presentó en establecimientos en la búsqueda de trabajo	Sí	Recuento	44	32	76
		% dentro de Se presentó en establecimientos	57,9%	42,1%	100,0%
		% dentro de Sexo	29,7%	21,3%	25,5%
	No	Recuento	104	118	222
		% dentro de Se presentó en establecimientos	46,8%	53,2%	100,0%
		% dentro de Sexo	70,3%	78,7%	74,5%
Total	Recuento	148	150	298	
	% dentro de Se presentó en establecimientos	49,7%	50,3%	100,0%	
	% dentro de Sexo	100,0%	100,0%	100,0%	
	% del total	49,7%	50,3%	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	2,764 ^a	1	,096		
Corrección por continuidad ^b	2,340	1	,126		
Razón de verosimilitudes	2,773	1	,096		
Estadístico exacto de Fisher				,111	,063
Asociación lineal por lineal	2,755	1	,097		
N de casos válidos	298				

a. 0 casillas (,0%) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 37,74.

b. Calculado sólo para una tabla de 2x2.

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

Medidas simétricas

		Valor	Error típ. asint. ^a	T aproximada ^b	Sig. aproximada
Nominal por nominal	Phi	,096			,096
	V de Cramer	,096			,096
Ordinal por ordinal	Tau-c de Kendall	,084	,050	1,669	,095
	Gamma	,219	,128	1,669	,095
N de casos válidos		298			

a. Asumiendo la hipótesis alternativa.

b. Empleando el error típico asintótico basado en la hipótesis nula.

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

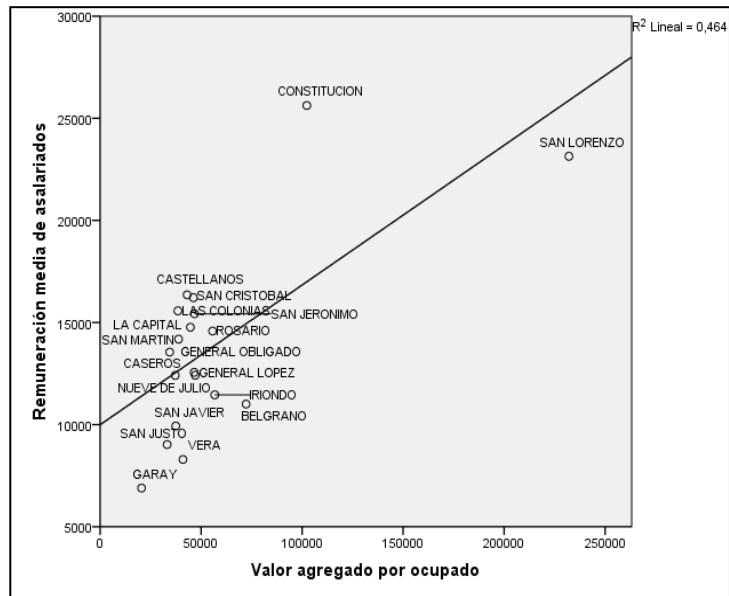
Ejercicio 3

Con el fin de analizar la diversidad del salario industrial entre los departamentos de la provincia de Santa Fe se propone considerar como un factor explicativo la capacidad de las empresas radicadas en cada uno de ellos para generar valor agregado. 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”.

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

1. Analizar, por medio del diagrama de dispersión y resultado del r de Pearson, si la distribución de puntos cumple con el supuesto de linealidad del modelo.
2. Analizar la correlación entre las dos variables por medio del r de Pearson, del coeficiente de determinación y del coeficiente de indeterminación. Sacar conclusiones con respecto a hipótesis propuesta.
3. Construir la ecuación de la recta de regresión. 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 promedio de salarios industriales de un supuesto departamento que posea un valor agregado industrial por ocupado de \$ 200.000.-

Gráfico 1
Departamentos de la provincia de Santa Fe según valor agregado industrial por ocupado y
media de remuneraciones en la industria. Año 2003
-Pesos de 2003-



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

Correlaciones

		Valor agregado por ocupado	Remuneración media de asalariados
Valor agregado por ocupado	Correlación de Pearson	1	,681**
	Sig. (bilateral)		,001
	N	19	19
Remuneración media de asalariados	Correlación de Pearson	,681**	1
	Sig. (bilateral)	,001	
	N	19	19

** La correlación es significativa al nivel 0,01 (bilateral).

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

Estadísticos descriptivos

	Media	Desviación típica	N
Remuneración media de asalariados	13864,1213	4603,35727	19
Valor agregado por ocupado	56548,1794	45834,99549	19

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

Resumen del modelo

Modelo	R	R cuadrado	R cuadrado corregida	Error típ. de la estimación
1	,681 ^a	,464	,433	3467,28035

a. Variables predictoras: (Constante), Valor agregado por ocupado

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

ANOVA^b

Modelo		Suma de cuadrados	gl	Media cuadrática	F	Sig.
1	Regresión	1,771E8	1	1,771E8	14,728	,001 ^a
	Residual	2,044E8	17	1,202E7		
	Total	3,814E8	18			

a. Variables predictoras: (Constante), Valor agregado por ocupado

b. Variable dependiente: Remuneración media de asalariados

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

Coefficientes^a

Modelo		Coeficientes no estandarizados		Coeficientes tipificados	t	Sig.
		B	Error típ.	Beta		
1	(Constante)	9994,688	1284,264		7,782	,000
	Valor agregado por ocupado	,068	,018	,681	3,838	,001

a. Variable dependiente: Remuneración media de asalariados

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

Ejercicio 4

Con el fin de analizar la diversidad de ingresos de los asalariados del Gran Buenos Aires en el 2º 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^a

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 ^a	,140	,139	1612,050
2	,442 ^b	,196	,195	1558,871
3	,449 ^c	,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^d

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

Coefficientes^a

Modelo		Coefficientes no estandarizados		Coefficientes 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.