



METODOLOGÍA Y TÉCNICAS DE INVESTIGACIÓN EN CIENCIAS SOCIALES

Titular: Agustín Salvia

Clase 9

**RELACIONES BIVARIDAS. CORRELACIÓN Y
REGRESIÓN**

Eduardo Donza



CORRELACIÓN Y REGRESIÓN

Datos de variables años de estudio e ingresos

**Nivel de medición
numérico**

Años de estudio (años)	Ingresos (\$)
5	1.700
6	2.000
7	2.300
8	2.600
9	2.900
10	3.200
11	3.500
12	3.800
13	4.100
14	4.400
16	5.000
17	5.300

Diagrama de dispersión años de estudio e ingresos

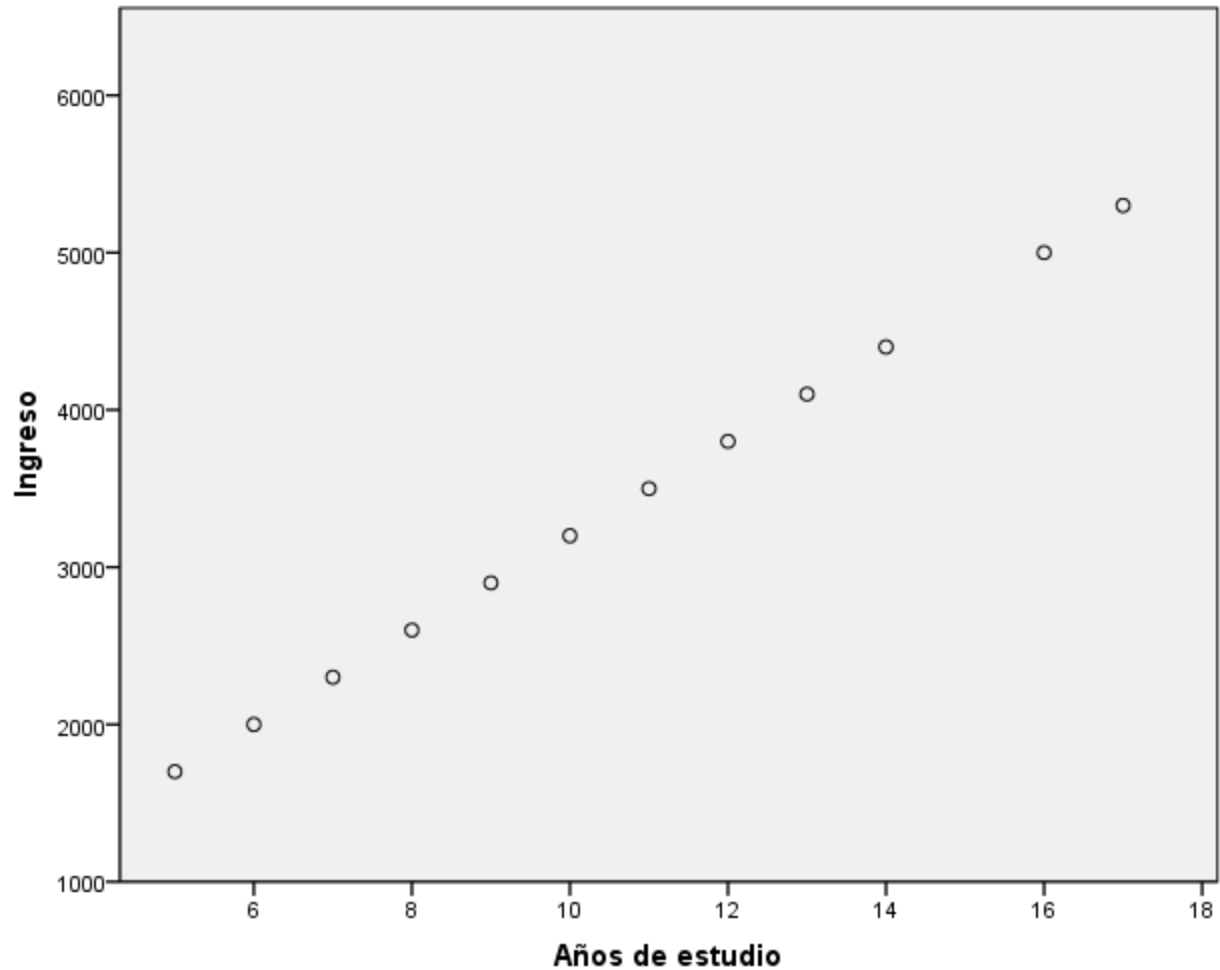
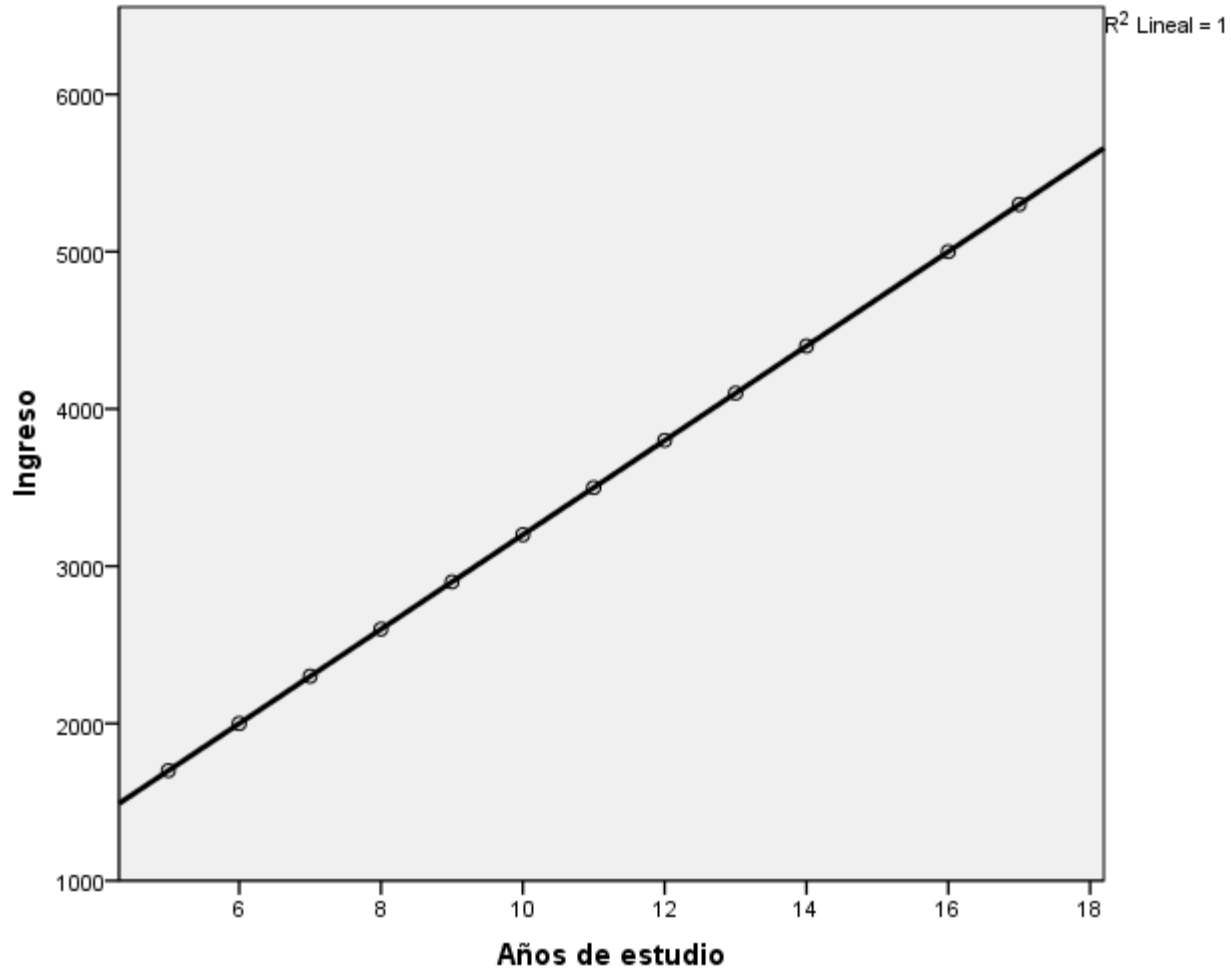
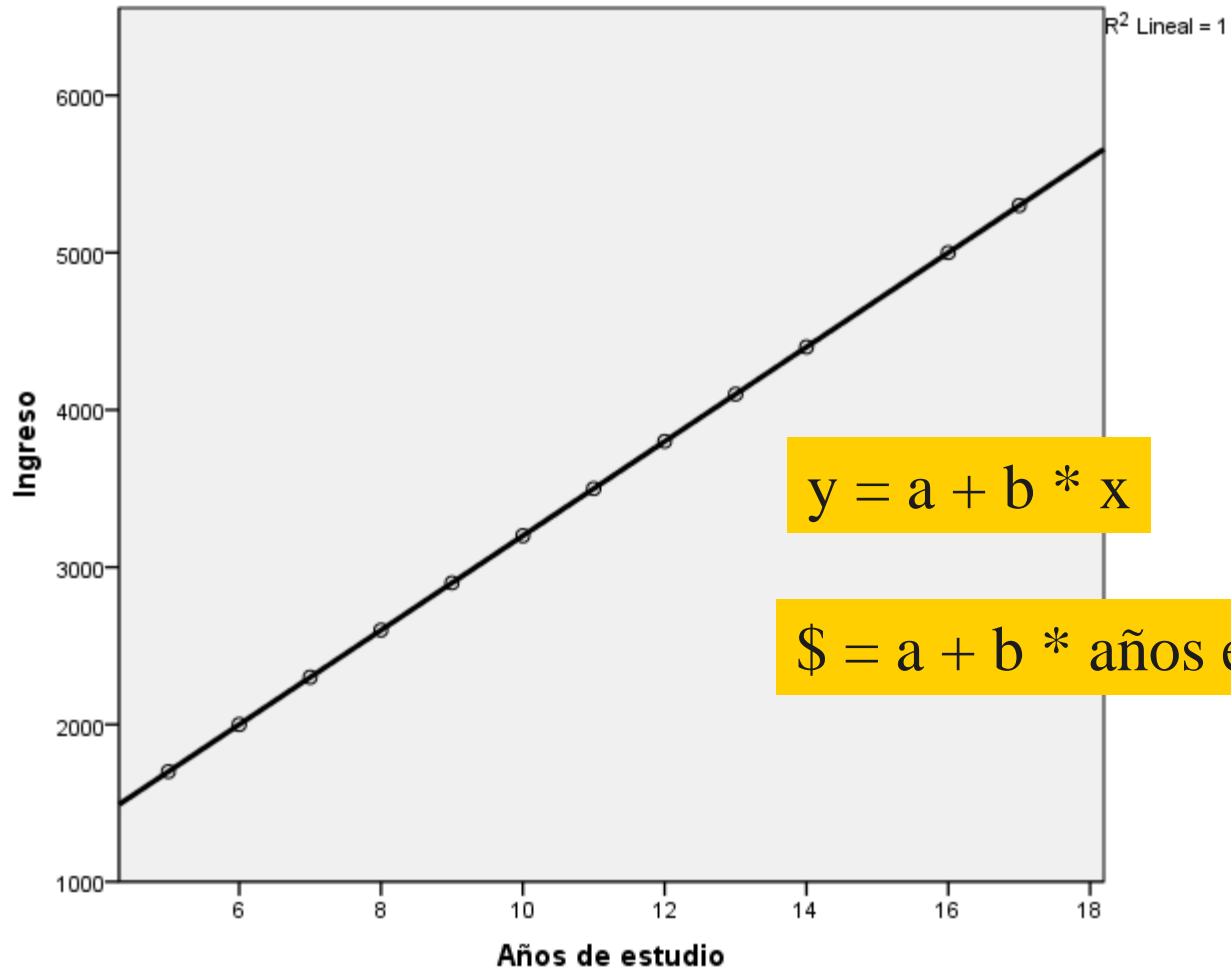


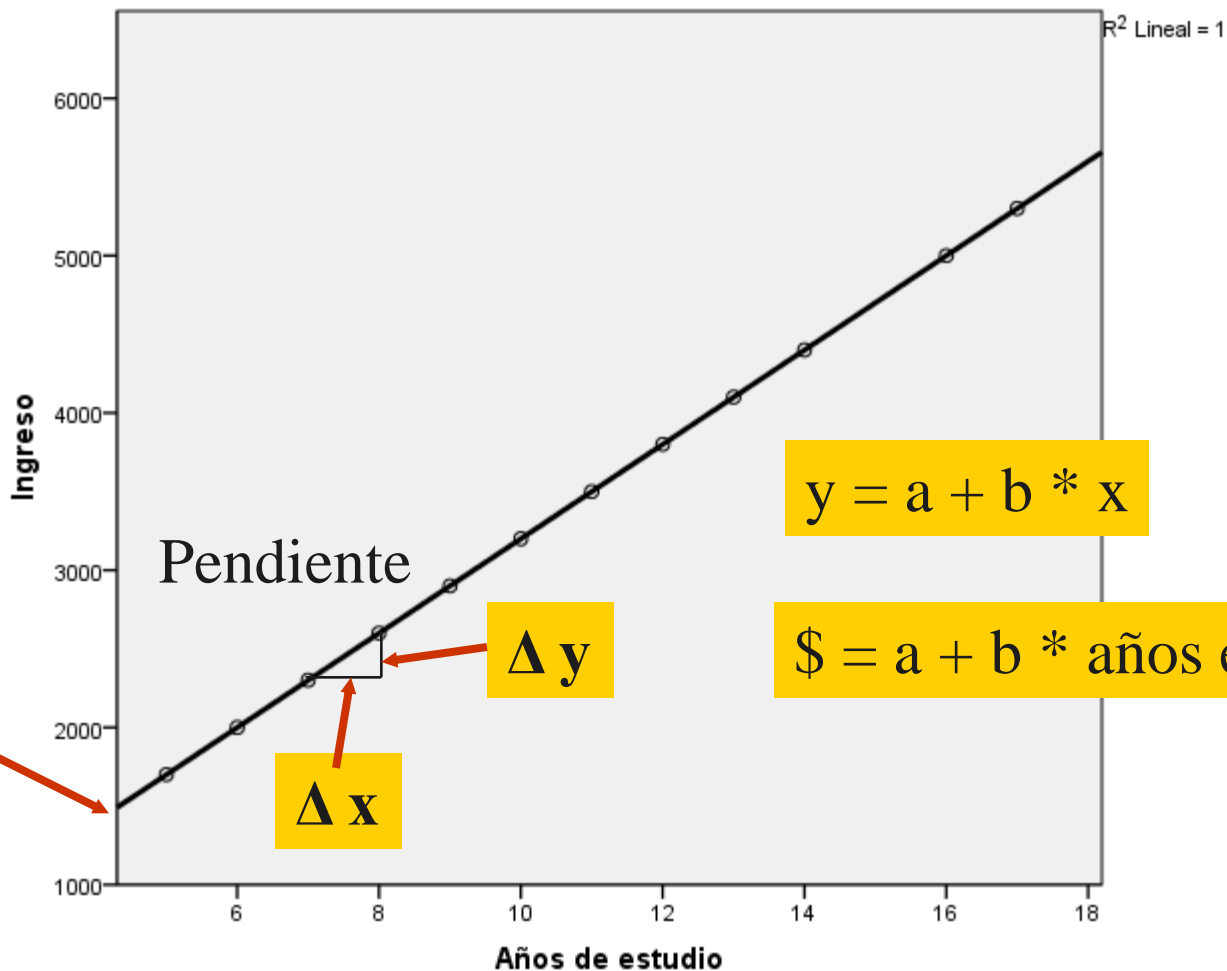
Diagrama de dispersión años de estudio e ingresos



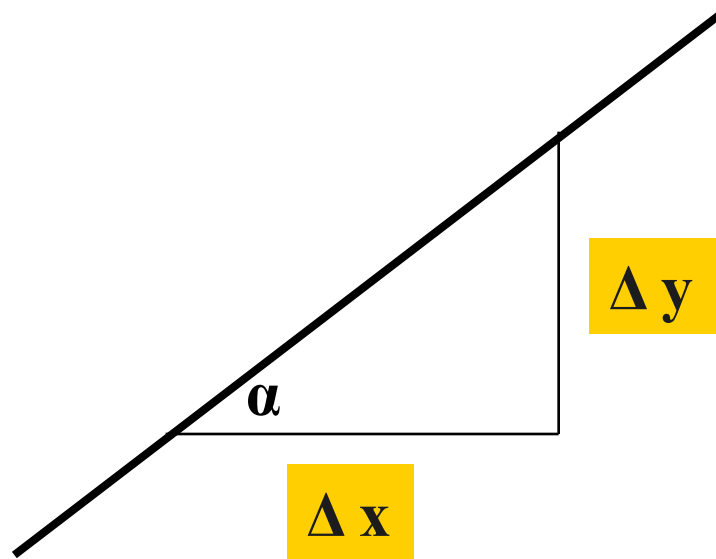
Recta de regresión



Particularidades de recta de regresión

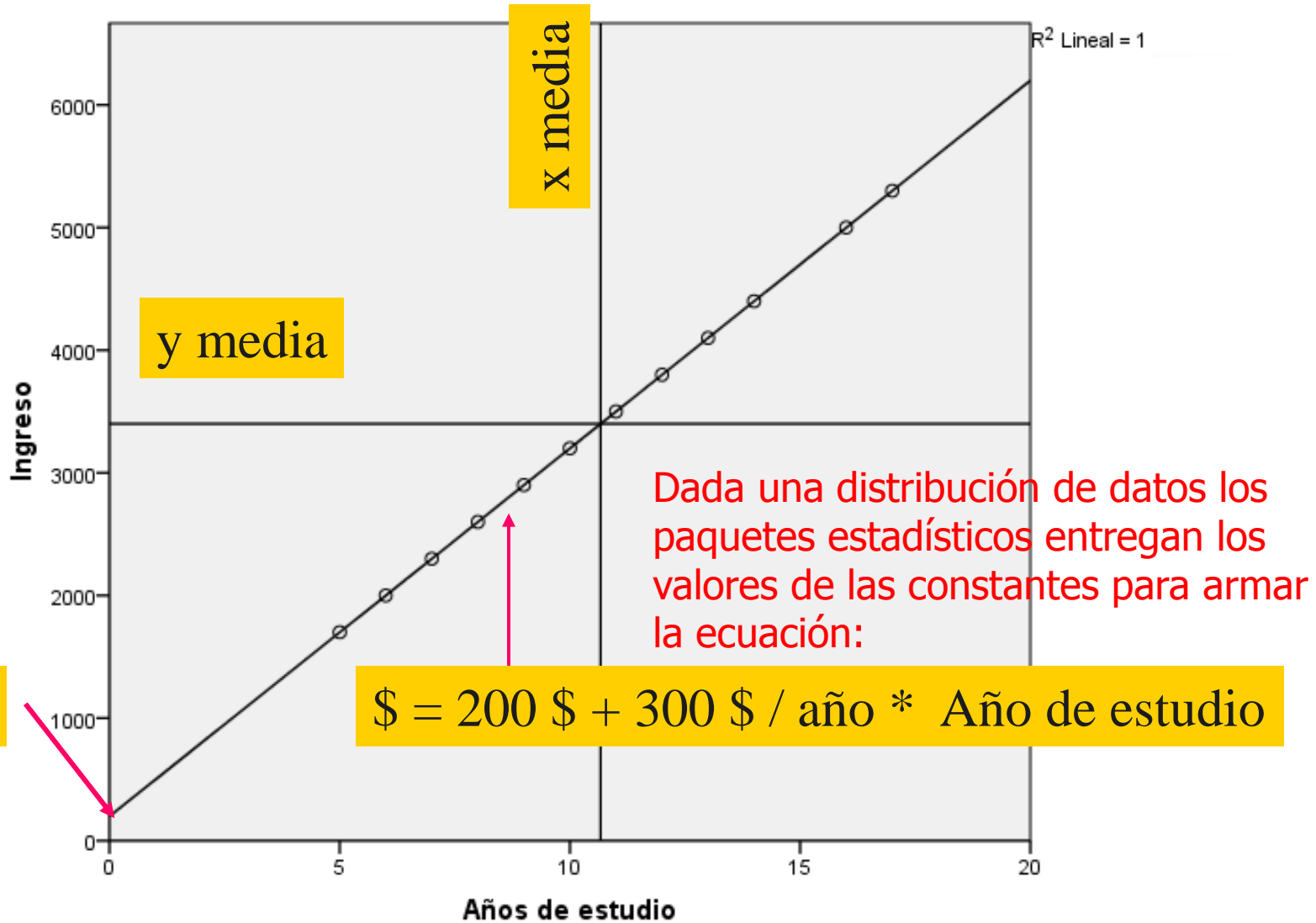


Pendiente de recta de regresión



$$b = \operatorname{tg} \alpha = \frac{\Delta y}{\Delta x}$$

Recta de regresión

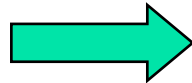




Predicción por medio de la ecuación

$$\$ = 200 \$ + 300 \$ / \text{año} * \text{Años de estudio}$$

Si años estudio = 15



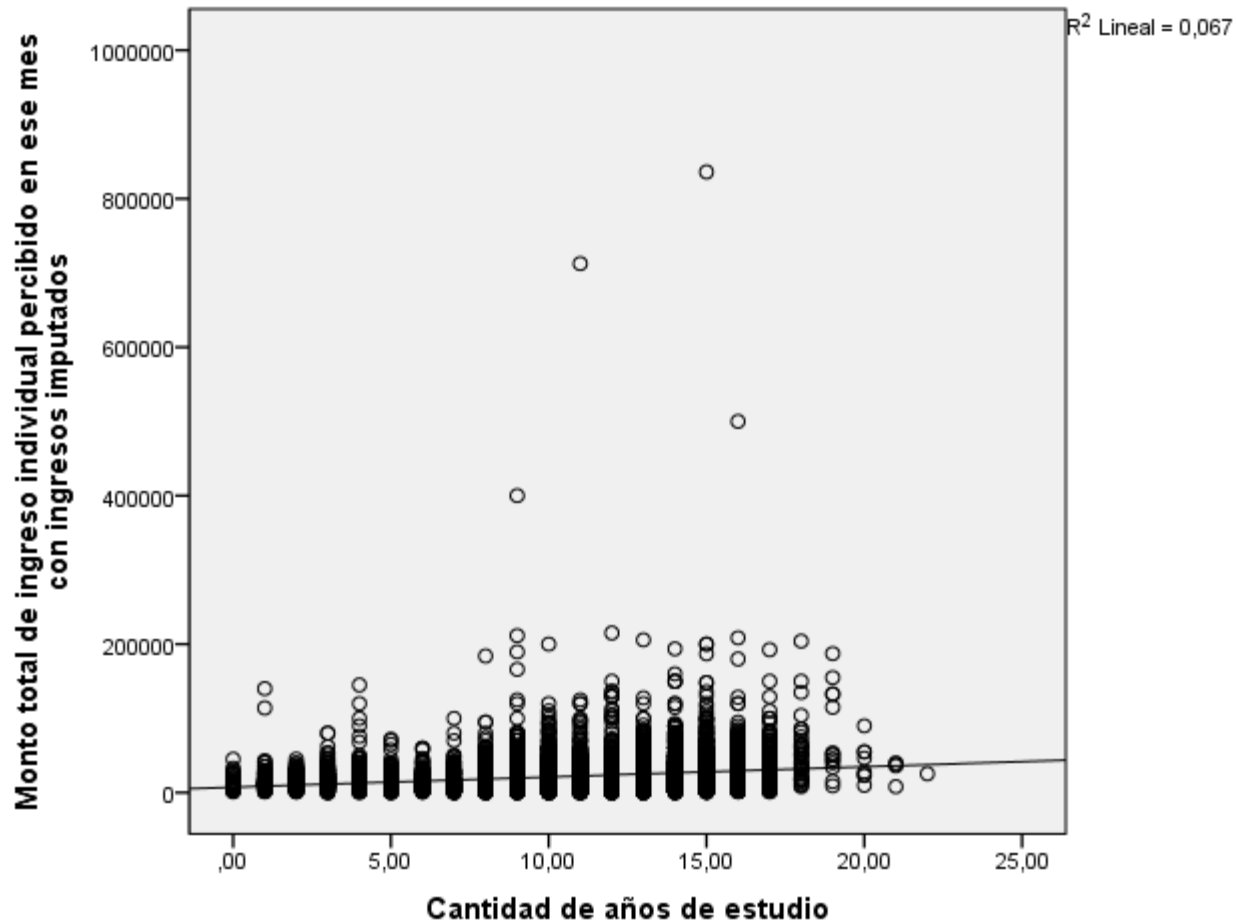
$$\$ = 200 \$ + 300 \$ / \text{años} * 15 \text{ años}$$

$$\$ = 200 \$ + 4500 \$$$

$$\$ = 4700 \$$$

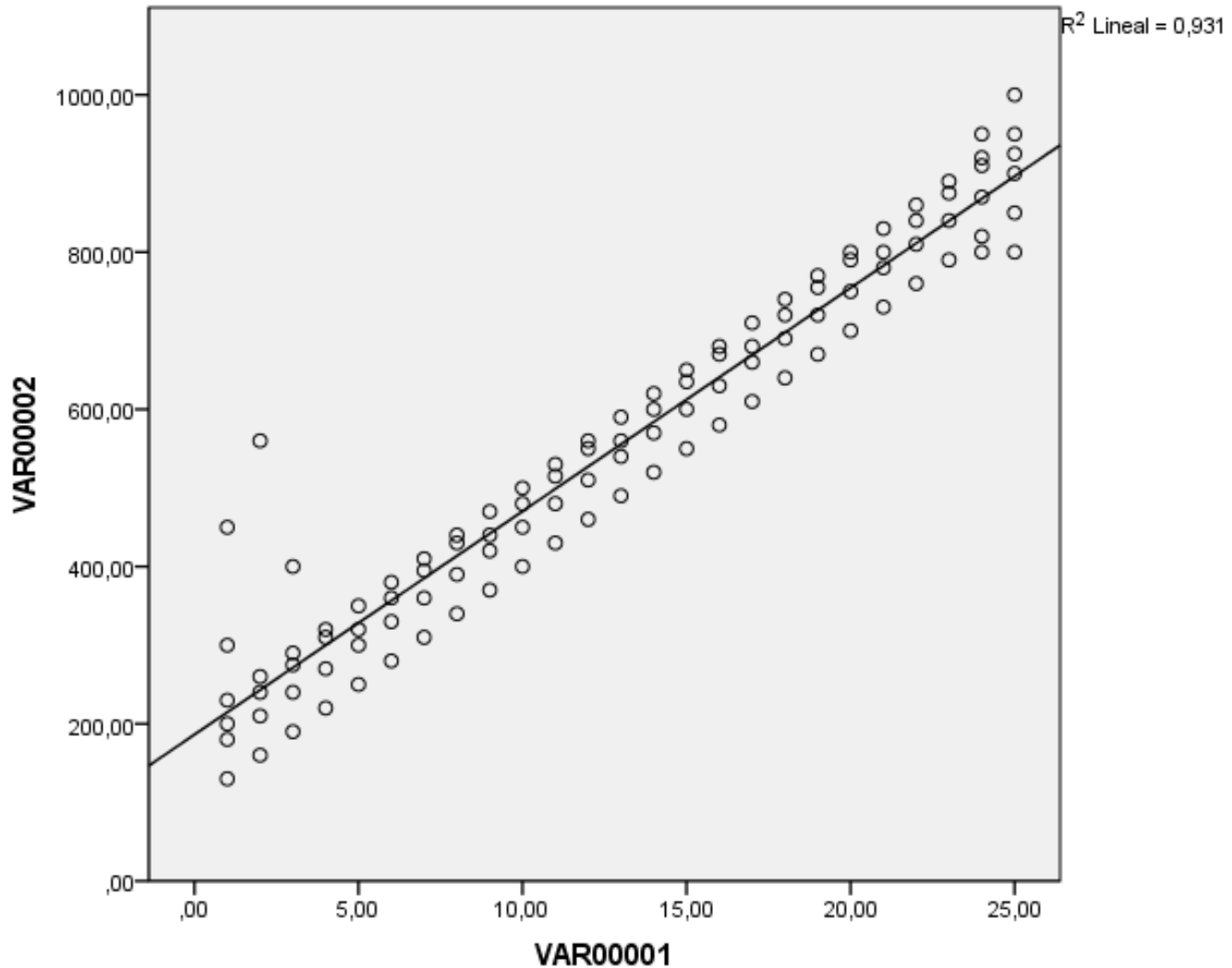
Dispersión de casos reales

Total de ingresos personales por cantidad de años de estudio. Encuesta Permanente de Hogares Total urbano EPH. 4° trim. 2019



Casos ponderados por Ponderador sin elevar

Recta de regresión / Técnica de mínimos cuadrados





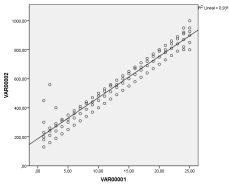
Correlación y regresión

Permiten:

- Medir la fuerza y el sentido de la relación por medio de un coeficiente denominado r de Pearson.
- Construir un modelo matemático que da cuenta de la distribución de la nube de puntos. Realizar predicciones de valores no conocidos de una de las variables.
- Determinar el nivel de confianza con que se puede asegurar que existe relación entre las variables en el universo observando los datos de la muestra.

Ejemplo de conclusiones al poner a prueba una hipótesis de variables numéricas

Datos



$$r = + 0,40$$

$$r^2 = 0,16$$

$$\text{Significancia} = 3\%$$

- Observando el diagrama de dispersión se puede aplicar una regresión lineal.
- La fuerza de la relación es moderada y el sentido es el propuesto por la hipótesis.
- El 16% de la variación de una variable esta determinado por la variación de la otra variable.
- La prueba de independencia estadística nos indica que se puede decir que hay relación entre las variables en el universo con un 97% de confianza.



METODOLOGÍA Y TÉCNICAS DE INVESTIGACIÓN EN CIENCIAS SOCIALES

Titular: Agustín Salvia

Clase 10

NOCIONES DE ANÁLISIS MULTIVARIADO

Eduardo Donza



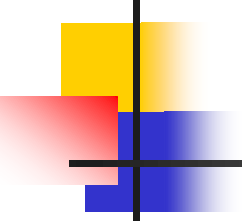
Usos del análisis multivariado

- **Aumentar el porcentaje de determinación de un evento (tienen más de una variable independiente)**
- **Identificar posibles interpretaciones espurias. Lograr explicación (tienen una variable de control)**
- **Clasificar unidades de análisis (clasificar unidades según uno o más de sus atributos)**

Principales técnicas de análisis multivariado según los usos

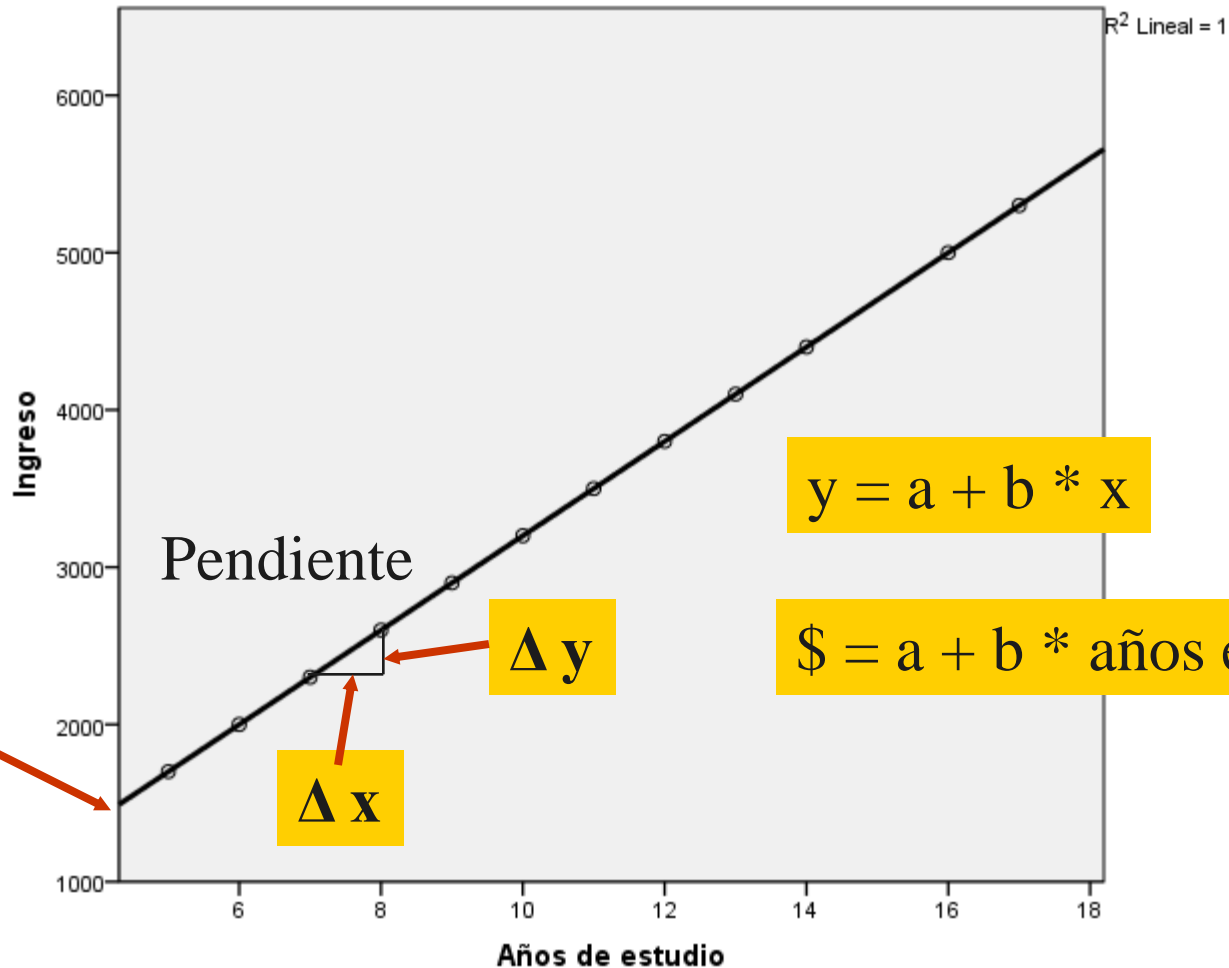
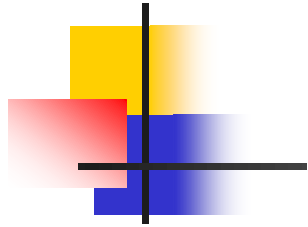


- **Aumentar el porcentaje de determinación de un evento**
 - Regresión múltiple
 - Regresión logística
 - Etc.
- **Identificar posibles interpretaciones espurias**
 - Modelo Lazarfeld
 - Correlación parcial
 - Hiloglineal
 - Etc.
- **Clasificar unidades de análisis o relaciones entre variables**
 - Análisis de cluster
 - Etc.



**Aumento del porcentaje de determinación:
Regresión lineal múltiple**

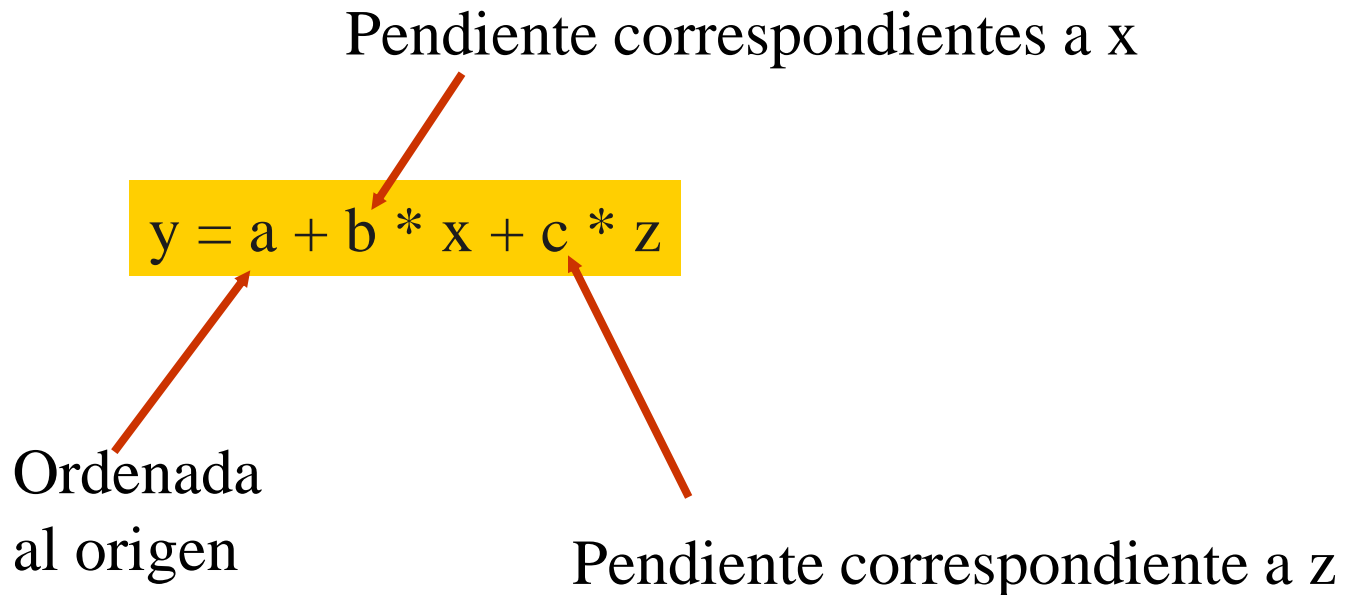
Particularidades de recta de regresión lineal simple



Ordenada
al origen

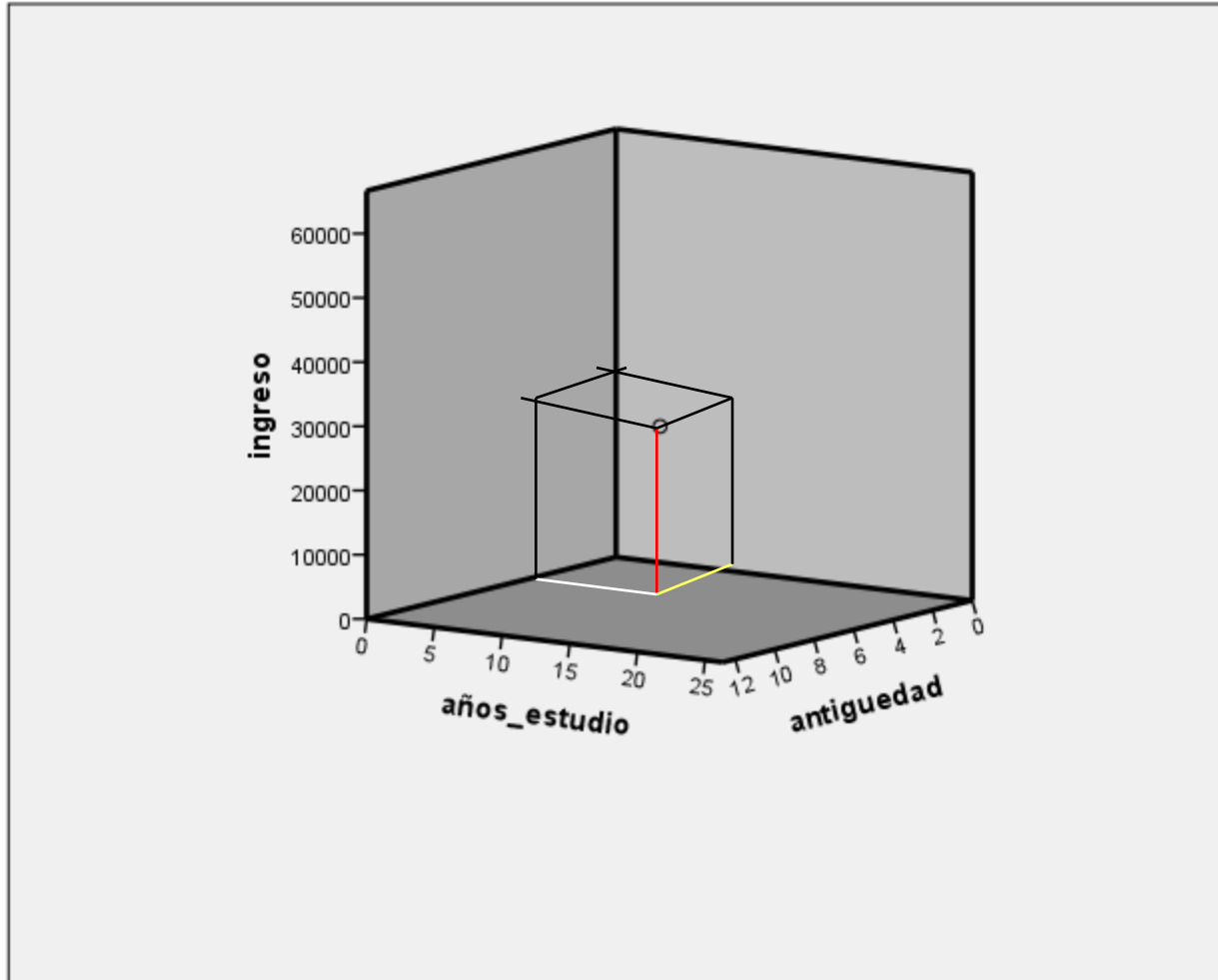
Regresión lineal múltiple

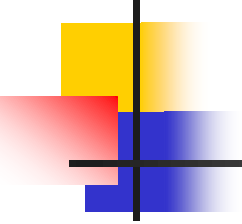
Introducción de otra/s variables independientes



La incorporación de una segunda variable independiente convierte la ecuación de la recta de predicción en la ecuación de un plano de predicción.

Identificación de un punto en el espacio según datos de tres variables





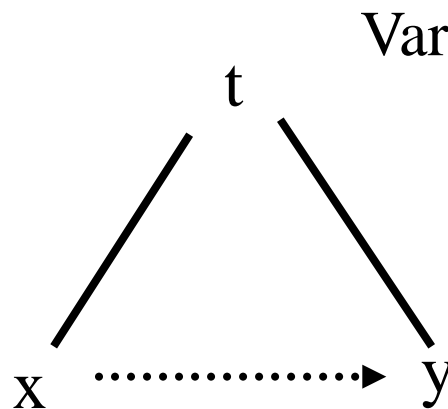
**Identificación de posibles interpretaciones
espurias. Búsqueda de explicación:
Variables de control**

Esquema multivariado. Introducción de variables de control



Relación original

La covariación observada no necesariamente es explicación



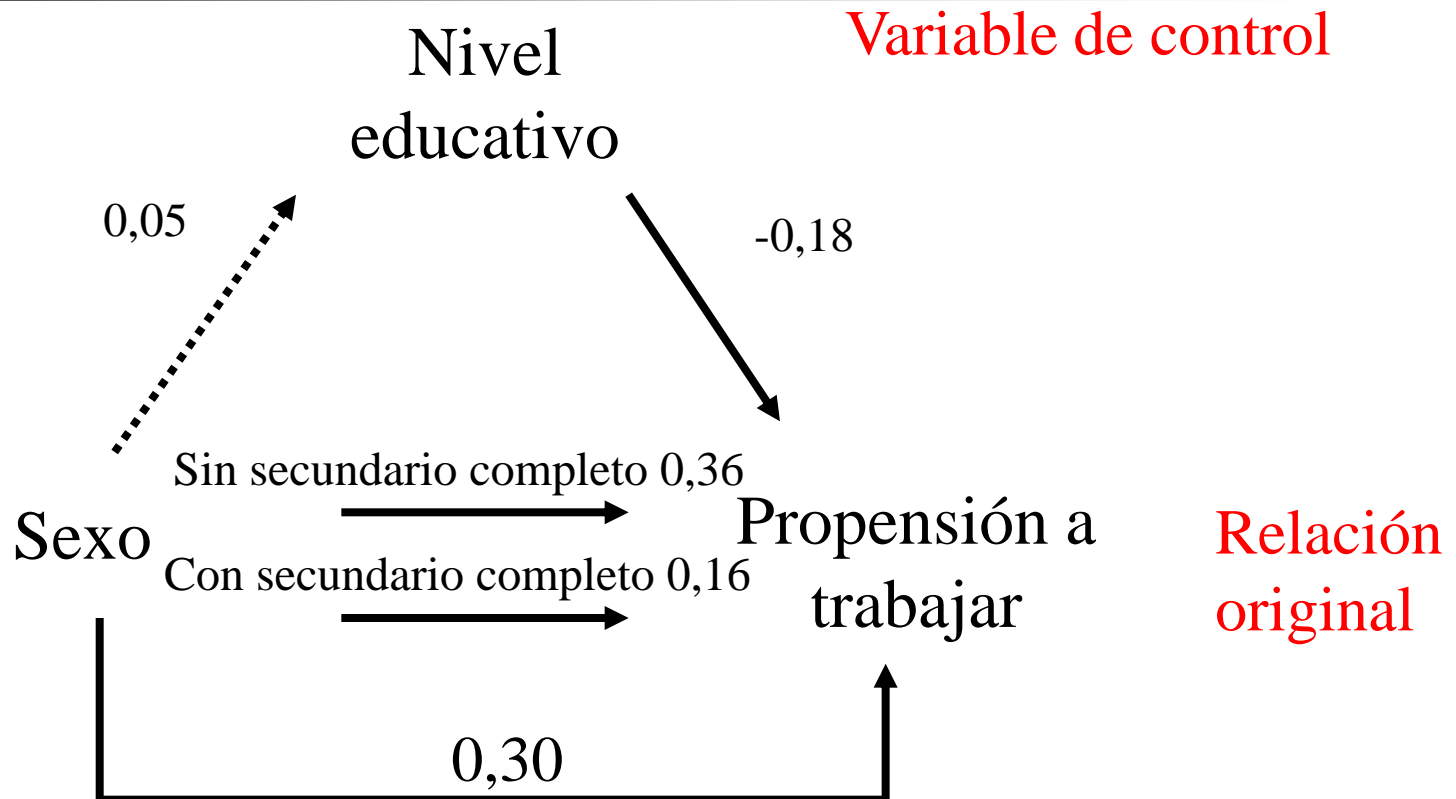
Variable de control

**Caso típico de
interpretación espuria**

Relación original

Esquema de modelo Lazarfeld

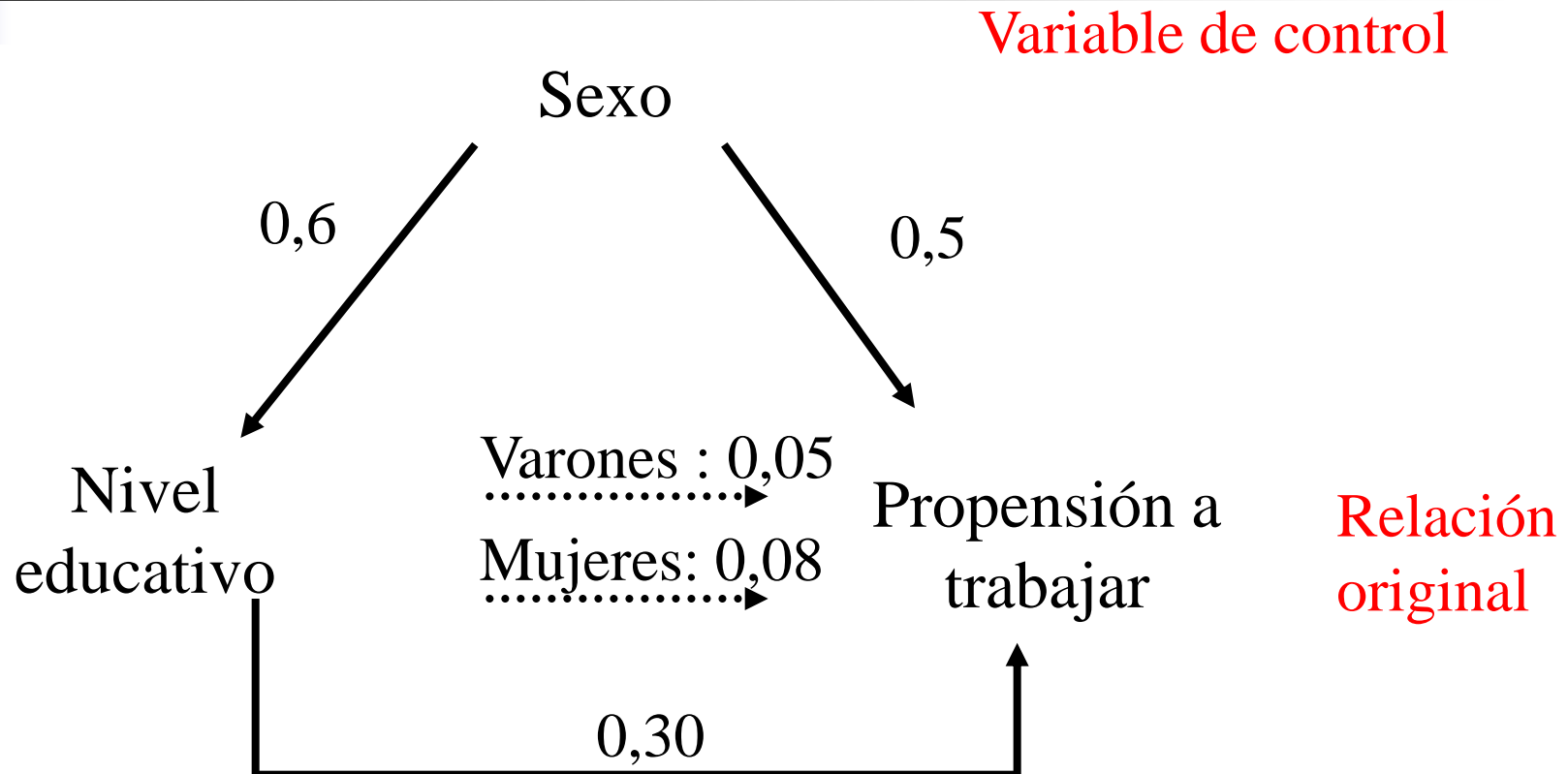
Coeficientes a analizar – Datos simulados:



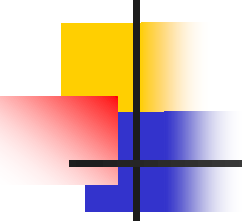
La variable de control especifica la relación. Se refuerza la relación original en el grupo de personas de bajo nivel educativo y disminuye en las de alto nivel educativo.

Caso de interpretación espuria

Coeficientes a analizar – Datos simulados:



La interpretación realizada era espuria. En realidad la propensión a trabajar depende del sexo de las personas.



Clasificación de unidades de análisis: Análisis de cluster

Ejemplo de Análisis de Cluster

Cluster	Aglomerado	Tasa de Actividad			Tasa de Empleo			Tasa de Desocupación		
		Media	Máx.	Mín.	Media	Máx.	Mín.	Media	Máx.	Mín.
1	Gran Buenos Aires	49,0			43,2			11,9		
	Gran La Plata	46,1			40,2			12,8		
	Bahía Blanca - Cerri	45,8			42,0			8,4		
	Gran Rosario	45,5			39,2			13,7		
	Gran Córdoba	45,3			41,2			9,2		
	Neuquen-Plottier	45,2			40,6			10,1		
	Ushuaia - Río Grande	47,7			43,2			9,5		
	Mar del Plata y Batán	48,4			43,1			11,0		
	Río Cuarto	43,7			39,4			9,7		
	Total		46,3	49,0	43,7	41,3	43,2	39,2	10,7	13,7
2	Gran Resistencia	35,4			33,4			5,6		
	Formosa	33,3			31,8			4,5		
	Total	34,4	35,4	33,3	32,6	33,4	31,8	5,1	5,6	4,5
3	Gran Santa Fe	39,2			36,4			7,2		
	Gran Paraná	41,3			38,1			7,7		
	Posadas	39,0			35,7			8,5		
	Corrientes	37,6			34,6			8,0		
	Concordia	40,0			35,9			10,1		
	Santiago del Estero- La Banda	40,5			36,8			9,2		
	Gran Catamarca	40,9			36,6			10,5		
	Salta	42,5			37,1			12,6		
	Gran San Juan	42,9			38,7			9,8		
	Gran Tucumán-Tafí Viejo	40,5			35,4			12,6		
Total		40,4	42,9	37,6	36,5	38,7	34,6	9,6	12,6	7,2
4	Comodoro Rivadavia- Rada Tilly	44,0			40,7			7,4		
	Gran Mendoza	45,8			42,6			6,9		
	Jujuy- Palpalá	43,0			39,9			7,2		
	Río Gallegos	44,5			43,8			1,7		
	La Rioja	43,1			40,0			7,2		
	San Luis - El Chorrillo	44,1			43,5			1,2		
	Santa Rosa - Toay	41,2			39,6			4,0		
	Total		43,7	45,8	41,2	41,4	43,8	39,6	5,1	7,4