

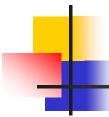
## METODOLOGÍA Y TÉCNICAS DE INVESTIGACIÓN EN CIENCIAS SOCIALES

**Titular: Agustín Salvia** 

Clase 8 (1º parte)

## RELACIONES BIVARIDAS. ANÁLISIS POR MEDIO DE CUADROS DE CONTINGENCIA. CORRELACIÓN Y REGRESIÓN

**Eduardo Donza** 

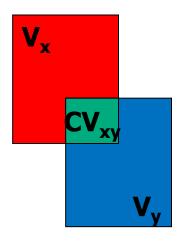


# CONCEPTO DE COVARIANZA, RELACIONES BIVARIDAS. ANÁLISIS POR MEDIO DE CUADROS DE CONTINGENCIA.

### Conceptos de covarianza



## Covarianza: varianza en conjunto entre las dos variables

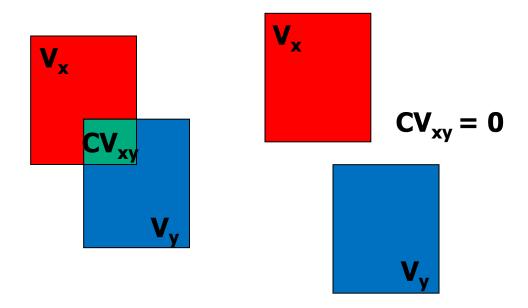


Relación intermedia entre las variables

### Conceptos de covarianza



## Covarianza: varianza en conjunto entre las dos variables



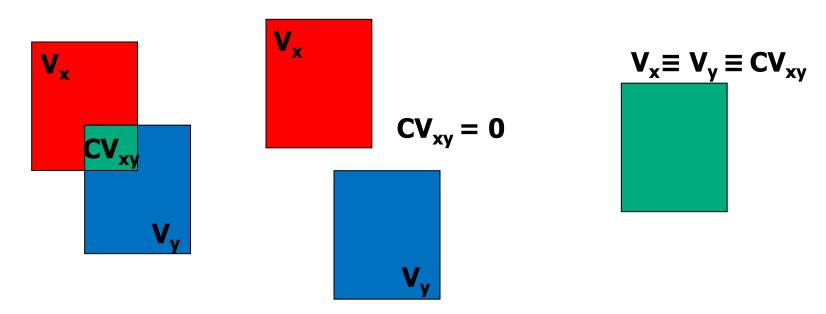
Relación intermedia entre las variables

Independencia estadística

### Conceptos de covarianza



## Covarianza: varianza en conjunto entre las dos variables



Relación intermedia entre las variables

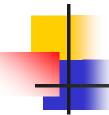
Independencia estadística

Relación perfecta entre las variables

### Clasificación de las técnicas estadísticas bivariadas que se expondrán en el curso

Cuadros bivariados (%)
VI y VD: Cualitativas Coeficientes de asociación
Test de ji cuadrado

VI y VD: Cuantitativa Correlación
Análisis de regresión



### RELACIÓN ENTRE VARIABLES CUALITATIVAS: ASOCIACIÓN



Para variables con nivel Variable x

de medición

nominal u

Variable y

ordinal	X <sub>1</sub>	X <sub>2</sub>	<b>X</b> <sub>3</sub>	
<b>y</b> <sub>1</sub>	Frecuencia	s condicion	ales	Marginal 1
<b>y</b> <sub>2</sub>				Marginal 2
У3				Marginal 3
	Subtotal 1	Subtotal 2	Subtotal 3	Total



### Usos de cuadros bivariados

 Para describir a la población según características de dos variables

Para contrastar hipótesis



## Pertenecer a la audiencia del programa de televisión según sexo GBA - Mayo 2017

-Cantidad de personas-

#### Ve el programa de televisión \* Sexo Crosstabulation

#### Count

		Se		
		Varón	Mujer	Total
Ve el programa de televisión	Si	300	310	610
television	No	100	300	400
Total		400	610	1010

Fuente: datos simulados.



# Para describir a la población según características de dos variables

### Cuadro bivariado para analizar

**datos**Pertenecer a la audiencia del programa de televisión según sexo
GBA - Mayo 2017

-En porcentajes-

#### Ve el programa de televisión \* Sexo Crosstabulation

			Se	XO	
			Varón	Mujer	Total
Ve el programa de televisión	Si	% within Ve el programa de televisión	49,2%	50,8%	100,0%
		% within Sexo	75,0%	50,8%	60,4%
		% of Total	29,7%	30,7%	60,4%
	No	% within Ve el programa de televisión	25,0%	75,0%	100,0%
		% within Sexo	25,0%	49,2%	39,6%
		% of Total	9,9%	29,7%	39,6%
Total		% within Ve el programa de televisión	39,6%	60,4%	100,0%
		% within Sexo	100,0%	100,0%	100,0%
		% of Total	39,6%	60,4%	100,0%

Fuente: datos simulados.



## Para contrastar hipótesis



#### Sexo

Ver el programa de televisión

	Varón	Mujer
Miran el programa de televisión	90%	20%
No miran el programa de televisión	10%	80%
	100%	100%



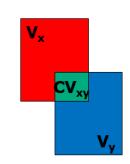
#### Sexo

Ver el programa de televisión

	Varón	Mujer
Miran el programa de televisión	90%	20%
No miran el programa de televisión	10%	80%
	100%	100%

d% = 70%

Relación intermedia entre las variables





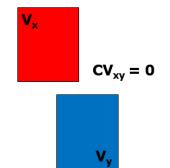
#### Sexo

Ver el programa de televisión

	Varón	Mujer
Miran el programa de televisión	60%	60%
No miran el programa de televisión	40%	40%
	100%	100%

d% = 0%

Independencia estadística entre las variables





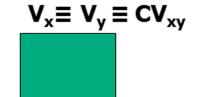
#### Sexo

Ver el programa de televisión

	Varón	Mujer
Miran el programa de televisión	100%	0%
No miran el programa de televisión	0%	100%
	100%	100%

$$d\% = 100\%$$

Relación perfecta entre las variables



### Reglas para el procedimiento

- Colocar la variable independiente en el cabezal del cuadro
- 2. Si son variables ordinales, verificar divergencia o convergencia de las categorías
- Realizar porcentaje por columnas
- Comparar por filas

Condición de actividad por sexo GBA / EPH 2º trim. de 2010

-En porcentaje-

#### Tabla de contingencia Condición de actividad \* Sexo

% dentro de Sexo

		Sexo		
		Varón	Mujer	Total
Condición de actividad	ocupado	93,1%	90,3%	91,9%
	desocupado	6,9%	9,7%	8,1%
Total		100,0%	100,0%	100,0%

d% = 2.8%

d% = -2.8%

#### Pasos:

- Var. Independiente en el cabezal
- Orden de categorías
- Porcentajes por columnas
- Comparar por fila

Sector de inserción de la población según sexo GBA / EPH 2º trim. de 2010

-En porcentaje-

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

% dentro de Sexo

		Se	хо	
		Varón	Mujer	Total
Sector de Inserción	Sector Público	10,4%	15,7%	12,6%
	Sector Formal	49,8%	36,5%	44,2%
	Sector Informal	39,8%	47,8%	43,1%
Total		100,0%	100,0%	100,0%

d%=-5,3%

d%=13,3%

d% = -8.0%

#### Pasos:

- Var. Independiente en el cabezal
- Orden de categorías
- Porcentajes por columnas
- Comparar por fila

La d% no es medida resumen de fuerza de la relación en cuadros de más de 2 col. x 2 filas

## Procedimientos a utilizar para la verificación de hipótesis

#### **Procedimientos:**

Lectura de porcentajes

Coeficientes de asociación

Pruebas de independencia estadística

## Procedimientos a utilizar para la verificación de hipótesis

#### Coeficientes de asociación:

- Miden la fuerza de la relación entre las variables
- Algunos coeficientes miden también el sentido de la relación (aplicable solo cuando ambas variables poseen nivel de medición ordinal).

## Procedimientos a utilizar para la verificación de hipótesis

### Pruebas de independencia estadística:

- La mas aplicada es la de ji (chi) cuadrado.
- Determinan el nivel de confianza con que se puede aseverar 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 cualitativas

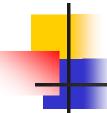
La lectura de porcentajes confirma la concentración en las celdas verificadoras.

Phi = + 0,40

La fuerza expresada por el coeficiente de asociación es moderada y el sentido es el propuesto por la hipótesis.

Significancia = 3%

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.



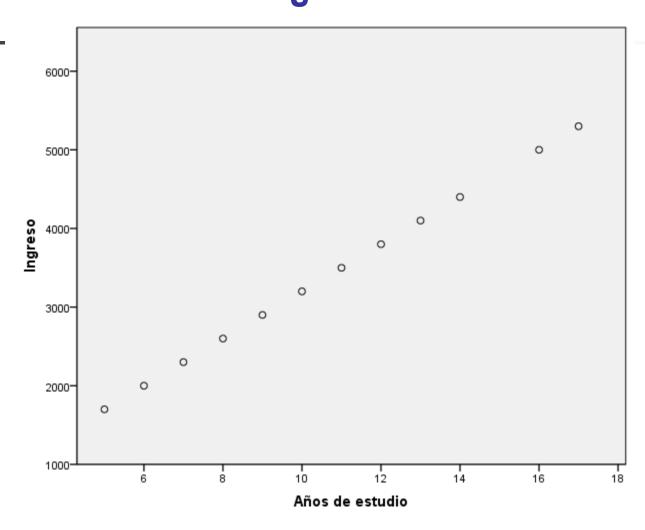
### **CORRELACIÓN Y REGRESIÓN**

## Datos de variables años de estudio e ingresos

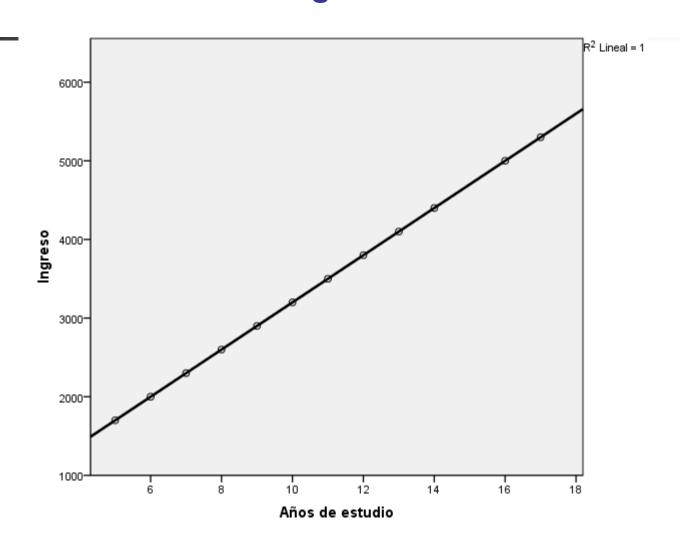
## Nivel de medición numérico

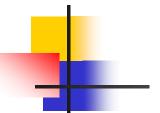
Años de estudio (años)	Ingresos (\$)
5	1.700
6	2.000
7	2.300
8	2.600
9	2.900
10	3.200
11	3.500
12	3.800
13	4.100
14	4.400
16	5.000
17	5.300

## Diagrama de dispersión años de estudio e ingresos

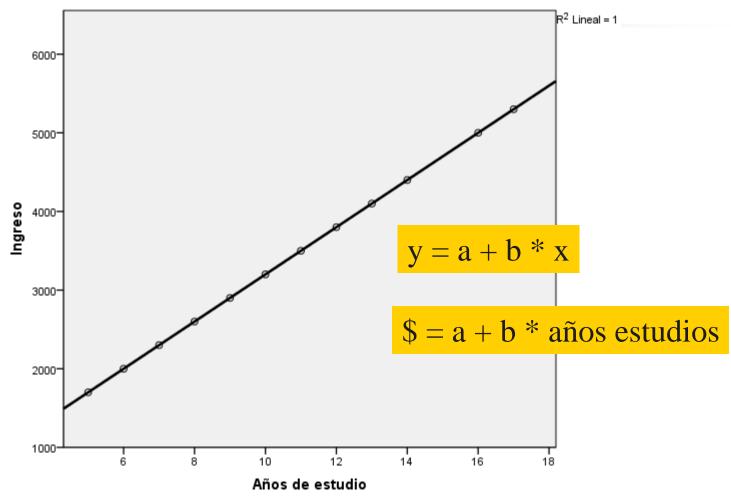


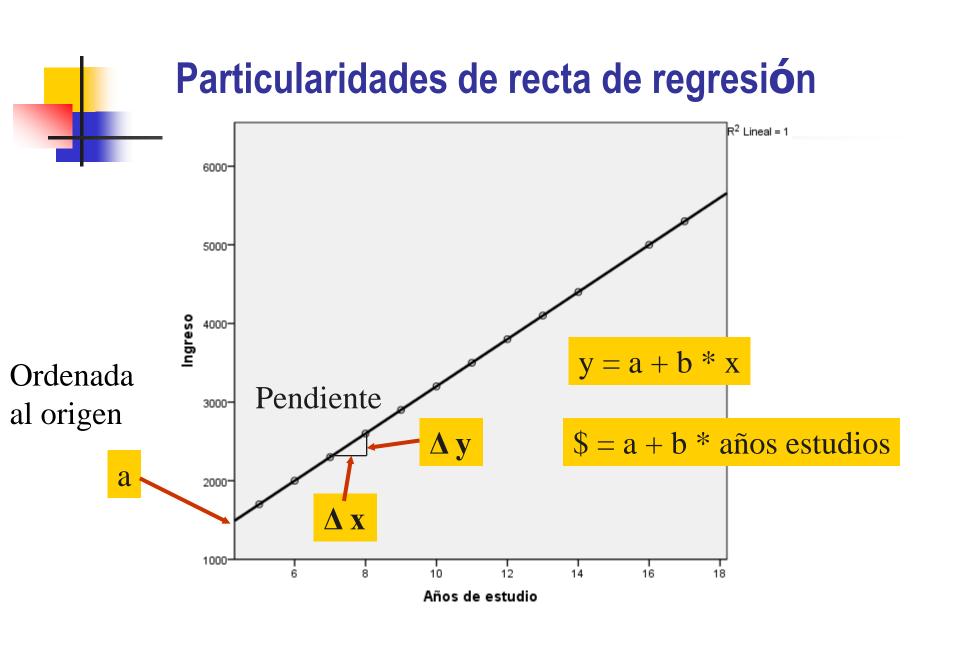
## Diagrama de dispersión años de estudio e ingresos



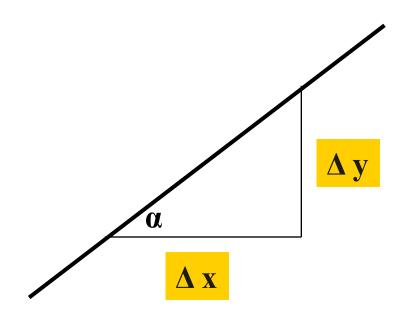


### Recta de regresión



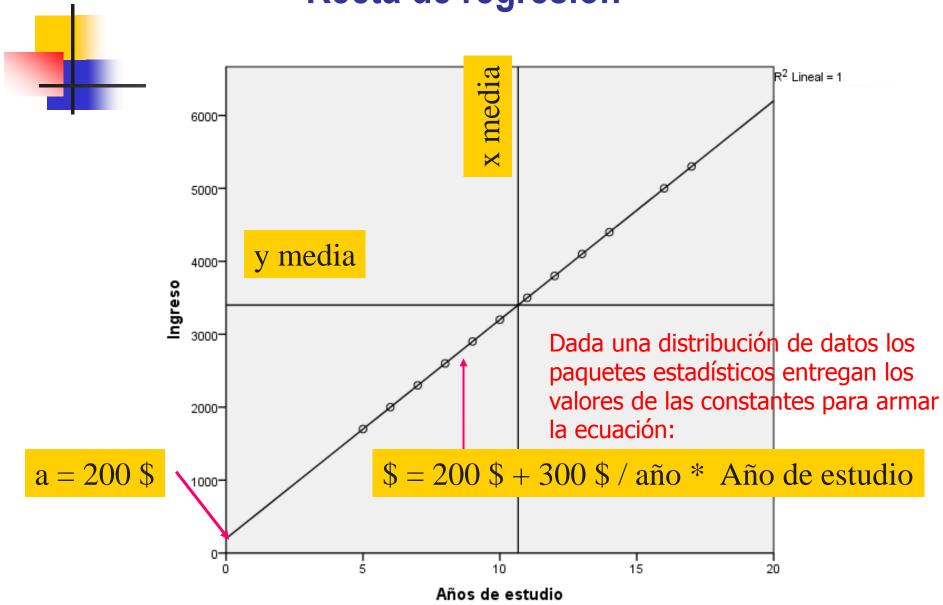


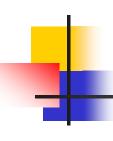
## Pendiente de recta de regresión



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

### Recta de regresión





### Predicción por medio de la ecuación

$$$ = 200 $ + 300 $ / año * Años de estudio$$

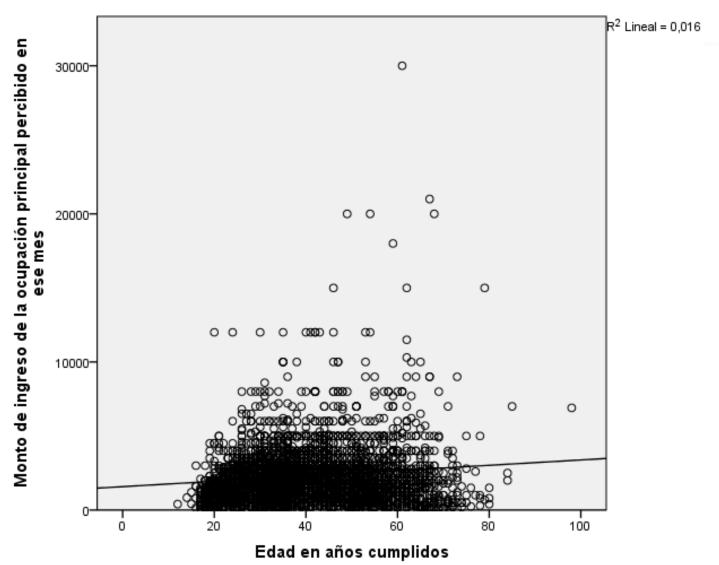


$$$ = 200 + 300 / anos * 15 anos$$

$$$ = 200 $ + 4500 $$$

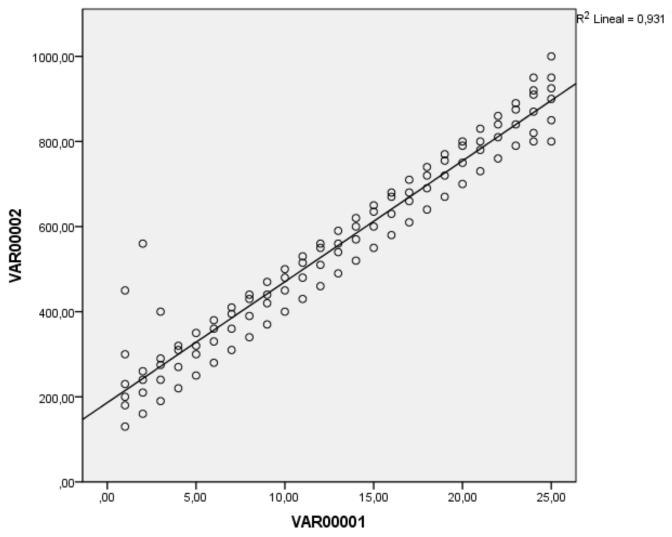
## Dispersión de casos reales

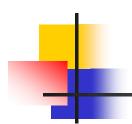




## Recta de regresión / Técnica de mínimos cuadrados







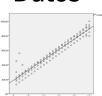
### Correlación y regresión

#### **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





$$r = + 0,40$$

$$r^2 = 0.16$$

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 8 (3º parte)

## NOCIONES BÁSICAS DE ANÁLISIS MULTIVARIADO

**Eduardo Donza** 



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