



METODOLOGÍA Y TÉCNICAS DE INVESTIGACIÓN EN CIENCIAS SOCIALES

Titular: Agustín Salvia

Teórico clase 7 (1º parte)

Estadística univariada.

Eduardo Donza

Distribución de frecuencia: Estado civil de los asistentes a ver cierta película de un cine del GBA

Estado Civil	Frecuencia Absoluta (f)	Frecuencia Porcentual (f% = $f \cdot 100/n$)
Soltero	246	51,3
Casado/Unido de hecho	196	40,8
Separado/Divorciado	29	6,0
Viudo	9	1,9
Total (n)	480	100,0



Estadística descriptiva univariada

Estadística



Posibilita analizar gran cantidad de datos

Estadística descriptiva

- Conjunto de instrumentos y temas relacionados con la descripción de series de observaciones

Estadística inferencial

- Lógicas y procedimientos para la inducción de propiedades de una población en base a datos de una muestra



Estadística descriptiva

Medidas de tendencia central

- Determinan la posición del grueso de la distribución

Medidas de dispersión

- Miden el nivel de homogeneidad de la distribución de datos

Medidas de posición

- Permiten identificar la posición de una unidad de análisis en función de la distribución



Estadística descriptiva

Medidas de tendencia central

- Modo o Moda
- Mediana
- Media

Medidas de dispersión

- Rango
- Desvío estándar o típico
- Varianza
- Producto de proporciones ($p \cdot q$)
- Coeficiente de variación ($CV = \text{desvío estándar} / \text{media} * 100$)

Medidas de posición

- Clasificación en cuartiles / quintiles / deciles / percentiles / etc.

Distribución de frecuencia

Variable con categorías

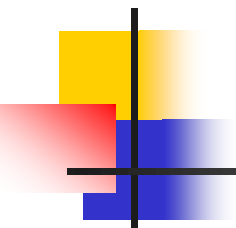
Formato típico de salida de paquete estadístico de análisis de datos

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

Sector de Inserción

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Sector Público	735034	5,6	5,6	5,6
	Sector Formal	2571964	19,7	19,7	25,4
	Sector Informal	2505374	19,2	19,2	44,6
	Desocupados/Inactivos	7215833	55,4	55,4	100,0
	Total	13028205	100,0	100,0	
Perdidos	Sistema	6419	,0		
Total		13034624	100,0		

Medidas de estadística descriptiva



Sector de Inserción					
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válidos	Sector Público	735034	5,6	5,6	5,6
	Sector Formal	2571964	19,7	19,7	25,4
	Sector Informal	2505374	19,2	19,2	44,6
	Desocupados/Inactivos	7215833	55,4	55,4	100,0
	Total	13028205	100,0	100,0	
Perdidos	Sistema	6419	,0		
Total		13034624	100,0		

Estadísticos		
Sector de Inserción		
N	Válidos	13028205
	Perdidos	6419
Media		3,2436
Mediana		4,0000
Moda		4,00
Desv. típ.		,95792
Varianza		,918
Rango		3,00
Percentiles	10	2,0000
	20	2,0000
	30	3,0000
	40	3,0000
	50	4,0000
	60	4,0000
	70	4,0000
	80	4,0000
	90	4,0000

No corresponden

Medidas de estadística descriptiva

Estadísticos

Monto de ingreso de la ocupación principal percibido en ese mes

N	Válidos	5564144
	Perdidos	0
Media		2304,34
Mediana		2000,00
Moda		2000
Desv. típ.		1942,553
Varianza		3773510,329
Rango		29970
Percentiles	10	500,00
	20	1000,00
	30	1200,00
	40	1500,00
	50	2000,00
	60	2200,00
	70	2700,00
	80	3200,00
	90	4500,00

Ejemplo de aplicación: Experimento sobre tiempos de reacción (TR) en centésimas de segundo de una persona para 20 presencias sucesivas de un estímulo

Nº de orden	Serie 1	Serie 2
1	20	32
2	15	40
3	18	33
4	25	37
5	17	35
6	32	29
7	18	42
8	17	62
9	19	50
10	23	39
11	19	45
12	21	47
13	15	52
14	22	37
15	17	38
16	17	39
17	21	40
18	19	41
19	17	42
20	23	39

Serie 1: TR de una persona que ha reaccionado 20 veces al encenderse una luz roja.

Serie 2: TR de la persona que debe reaccionar sólo cuando aparece una luz roja pero puede encenderse una luz roja, verde o amarilla.

Medidas de tiempo de reacción

	Serie 1	Serie 2
Media	19,75	40,95
Desvío	3,96	7,53
CV	20,05%	18,39%

Datos de ingresos según deciles de población

Cuadro 1. Población según escala de ingreso per cápita familiar. Total 31 aglomerados urbanos. Tercer trimestre de 2019

Decil	Escala de ingreso		Población ⁽²⁾		Ingreso per cápita familiar				
	Desde	Hasta	Población por decil	Porcentaje de personas	Ingreso total por decil (en miles)	Porcentaje del ingreso	Ingreso medio por decil	Ingreso medio por estrato	Mediana por decil
	\$	\$		%	\$	%	\$	\$	\$
1	0	3.833	2.798.947	10,0	6.657.071	1,4	2.378		2.500
2	3.833	5.800	2.800.163	10,0	13.445.037	2,9	4.802		4.800
3	5.800	7.667	2.798.250	10,0	18.817.514	4,1	6.725		6.700
4	7.667	9.891	2.798.334	10,0	24.481.789	5,3	8.749	5.663	8.750
5	9.891	12.000	2.798.886	10,0	30.276.561	6,5	10.817		10.867
6	12.000	14.560	2.799.195	10,0	36.838.181	7,9	13.160		13.167
7	14.571	17.833	2.799.855	10,0	45.265.921	9,8	16.167		16.200
8	17.833	23.000	2.797.773	10,0	56.485.042	12,2	20.189	15.083	20.000
9	23.025	33.267	2.798.858	10,0	77.488.475	16,7	27.686		27.500
10	33.267	422.000	2.798.750	10,0	154.047.545	33,2	55.042	41.363	45.000
Población total ⁽¹⁾			27.989.011	100,0	463.803.137	100,0	16.571	16.571	12.000

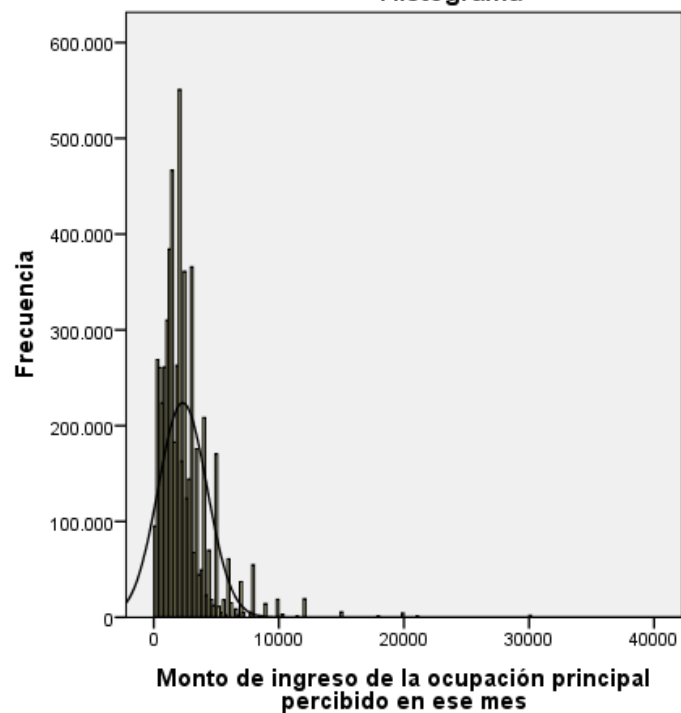


Gráficos

Forma de distribución

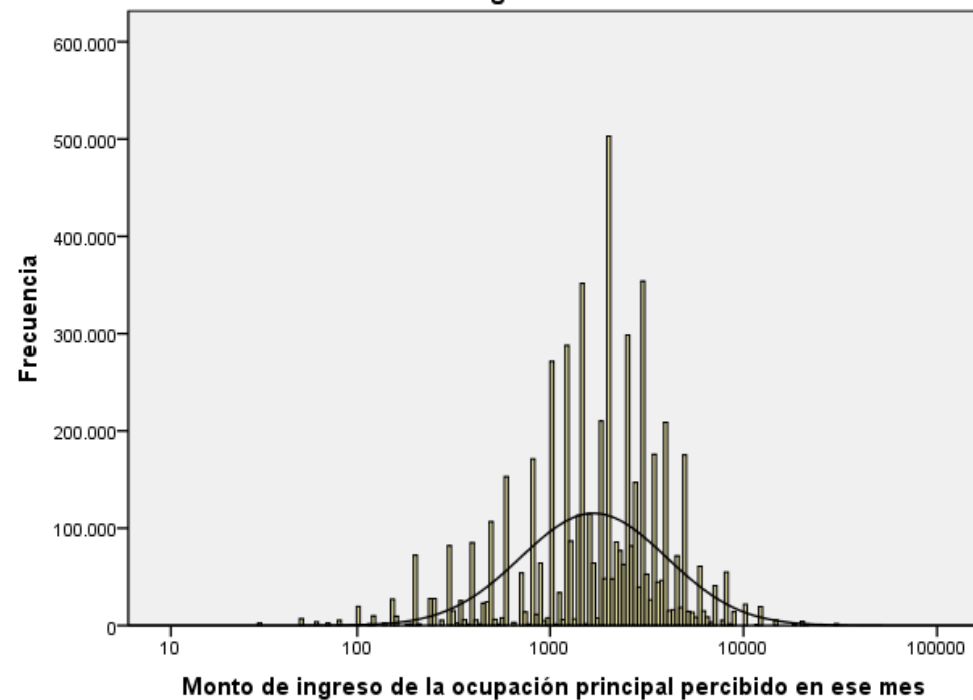


Histograma



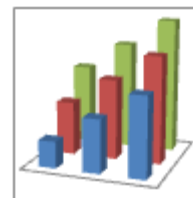
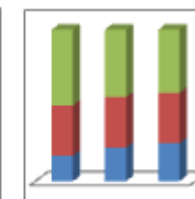
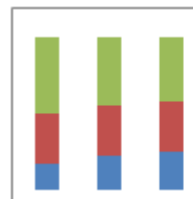
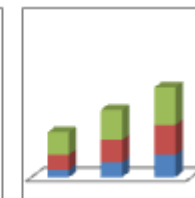
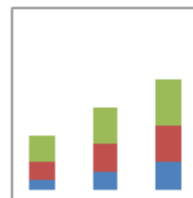
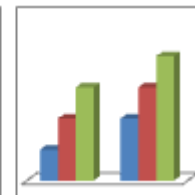
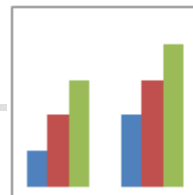
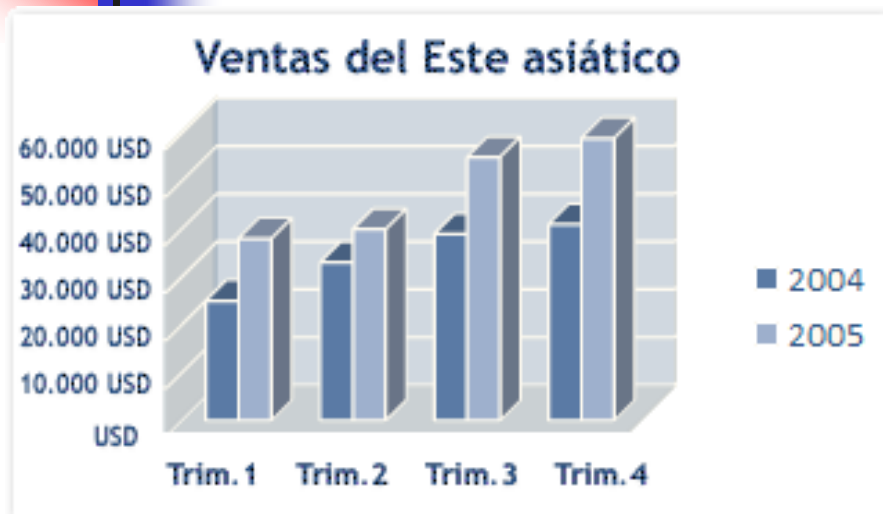
Casos ponderados por Ponderación

Histograma

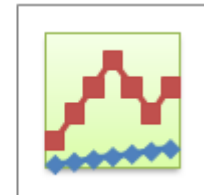


Casos ponderados por Ponderación

Gráficos de columnas

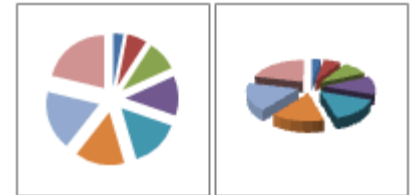
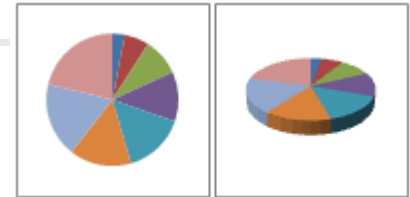
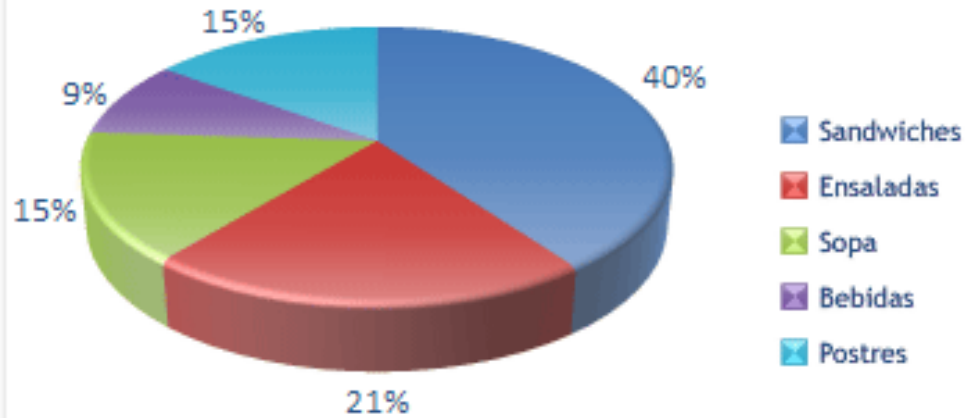


Gráficos de líneas

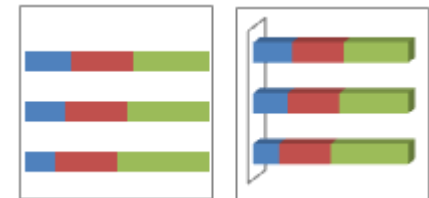
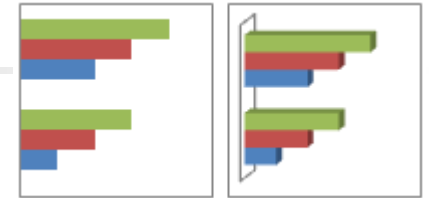
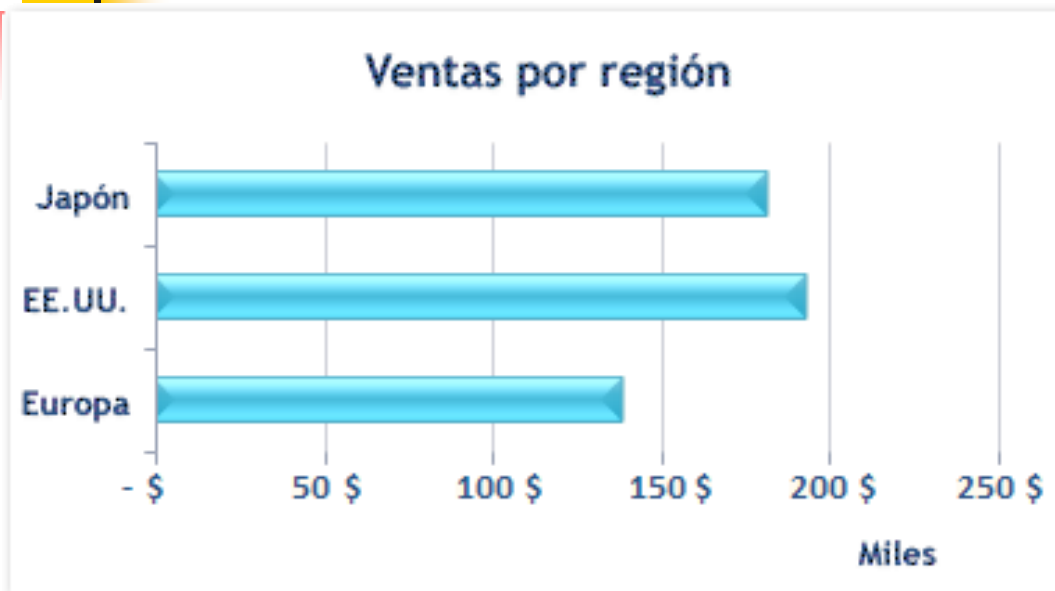


Gráficos circulares

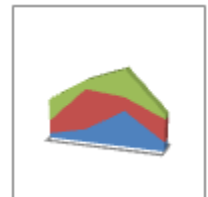
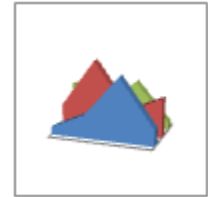
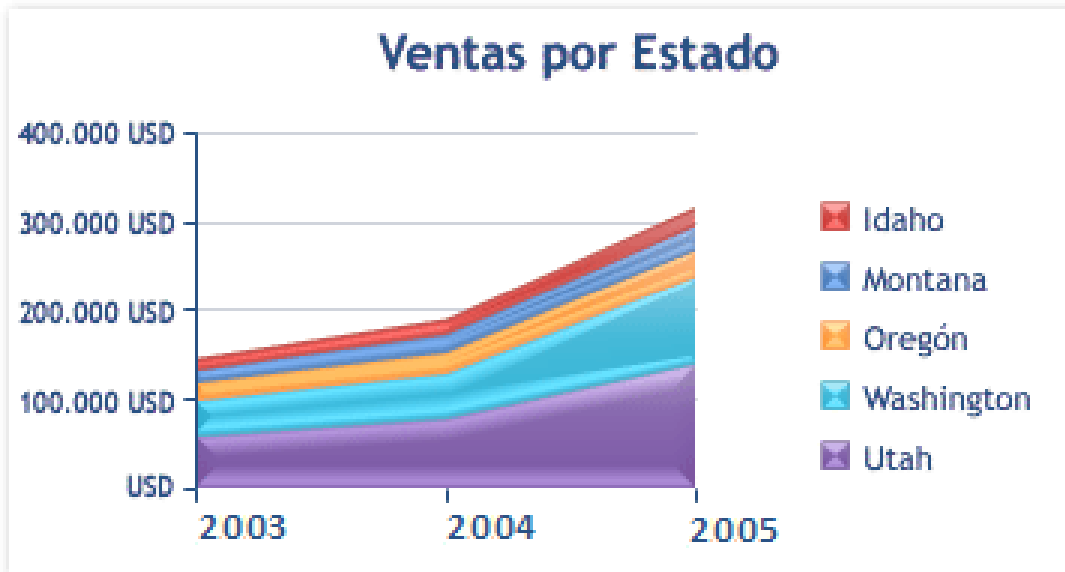
Ventas de comida



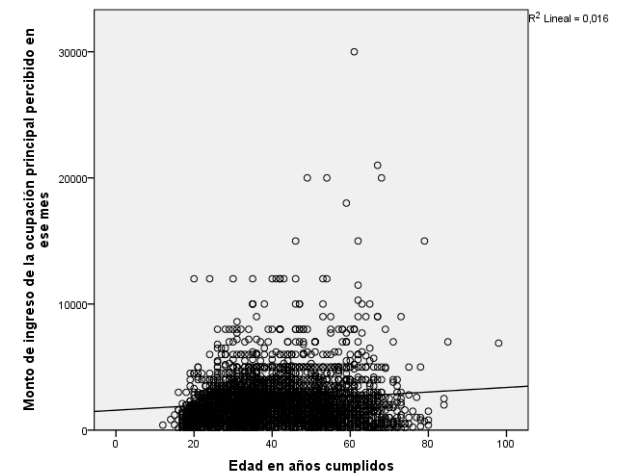
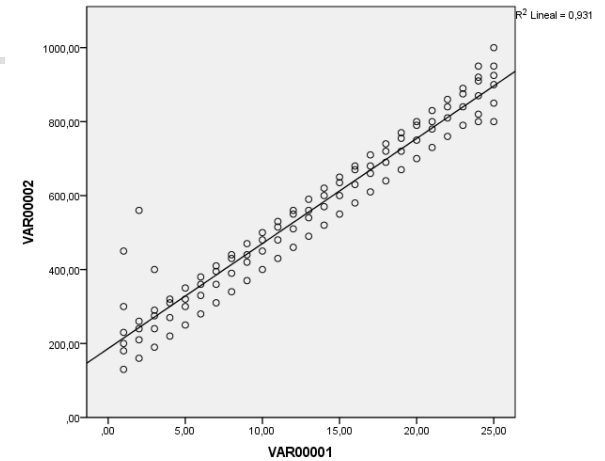
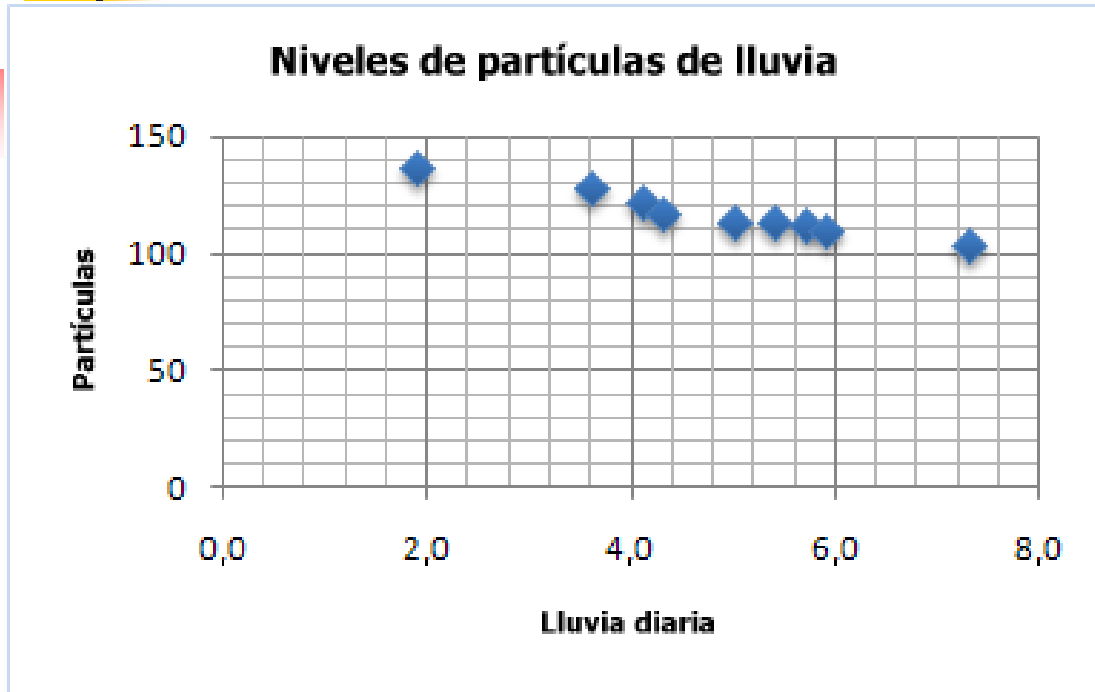
Gráficos de barras



Gráficos de área

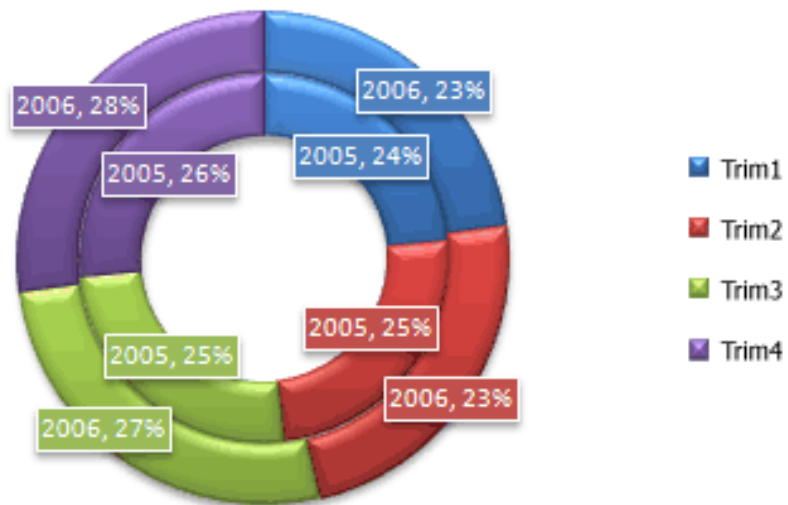


Gráficos de tipo XY (Dispersión)

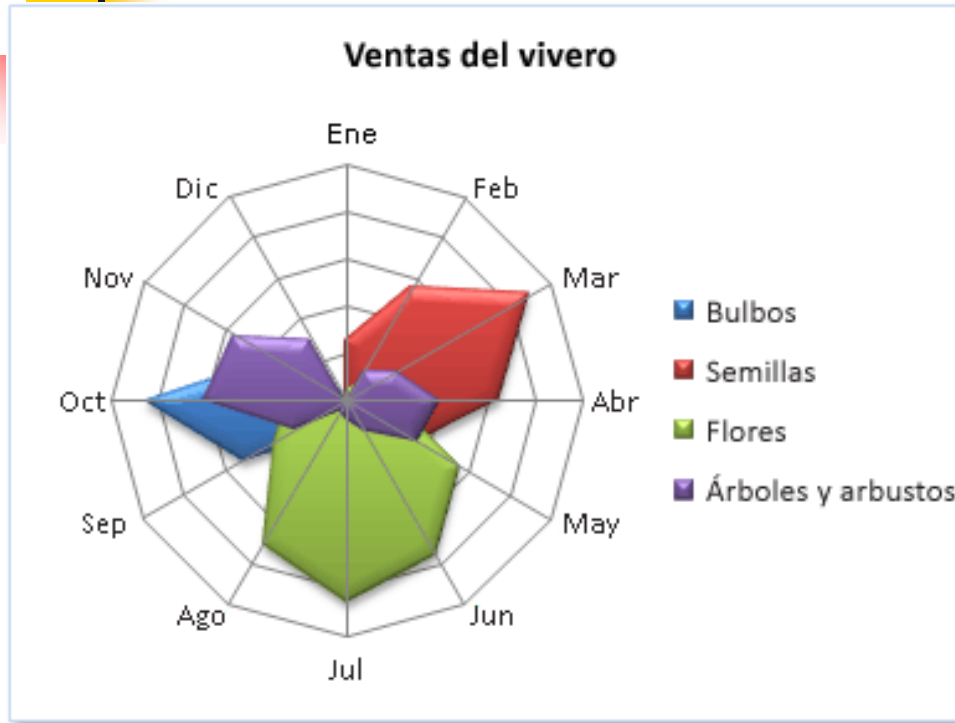


Gráficos de anillos

Ventas trimestrales



Gráficos radiales





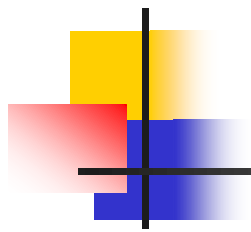
METODOLOGÍA Y TÉCNICAS DE INVESTIGACIÓN EN CIENCIAS SOCIALES

Titular: Agustín Salvia

Clase 7 (2º parte)

**CONCEPTO DE COVARIANZA, RELACIONES
BIVARIDAS. ANÁLISIS POR MEDIO DE CUADROS DE
CONTINGENCIA. CORRELACIÓN Y REGRESIÓN**

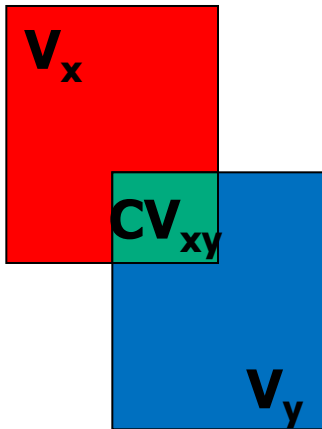
Eduardo Donza



CONCEPTO DE COVARIANZA, RELACIONES BIVARIADAS. ANÁLISIS POR MEDIO DE CUADROS DE CONTINGENCIA.

Conceptos de covarianza

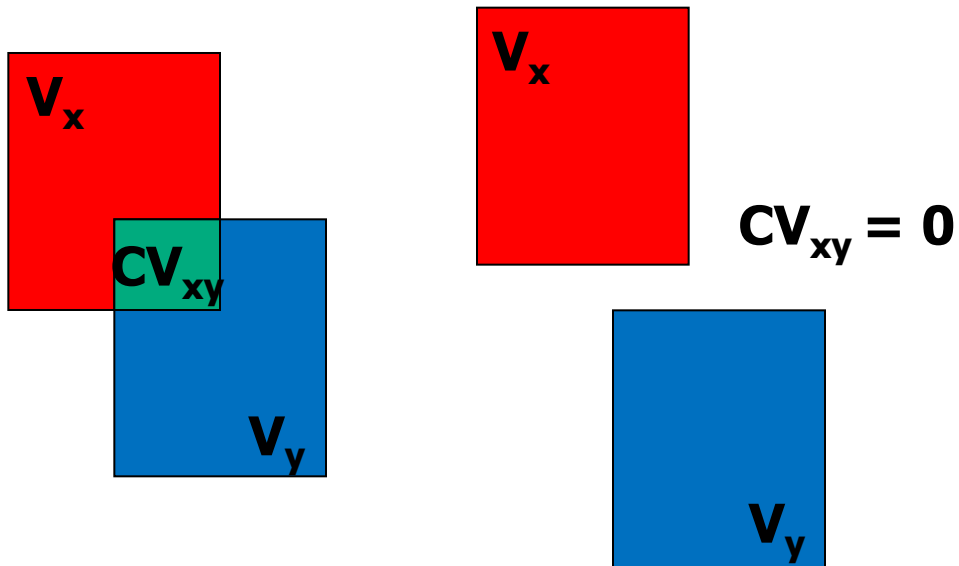
Covarianza: varianza en conjunto entre las dos variables



**Relación intermedia
entre las variables**

Conceptos de covarianza

Covarianza: varianza en conjunto entre las dos variables

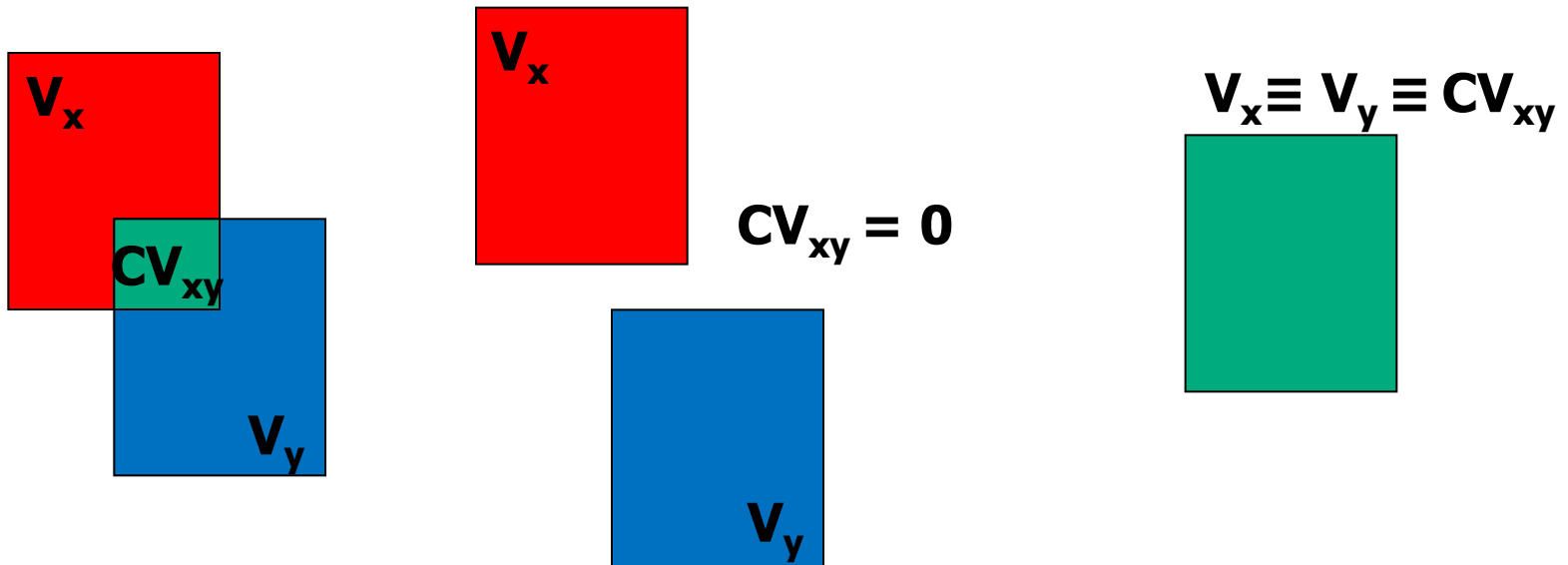


**Relación intermedia
entre las variables**

**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**

Conceptos a tener en cuenta

Concepto de covarianza → Contrastación de hipótesis

Relación entre variables

- Fuerza (Intensidad)
- Sentido
- Forma
- Grado

Tipos de hipótesis

- Diagonales (H1)
- Rinconales (H2)

H1:	B	A
B	■	□
A	□	■

H2:	V	M
Si	■	□
No	□	□

Posibles resultado al analizar la covarianza

- Intermedia
- Nula
- Total

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

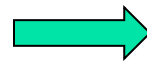


VI y VD: Cualitativas



Cuadros bivariados (%)
Coeficientes de asociación
Test de ji cuadrado

VI y VD: Cuantitativa



Correlación
Análisis de regresión



RELACIÓN ENTRE VARIABLES CUALITATIVAS: ASOCIACIÓN



Concepto de cuadro bivariado

Para variables con nivel
de medición
nominal u ordinal

Variable x

Variable y

	x_1	x_2	x_3	
y_1	Frecuencias condicionales			Marginal 1
y_2				Marginal 2
y_3				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**



Cuadro bivariado

Pertenecer a la audiencia del programa de televisión según sexo

GBA - Mayo 2017

-Cantidad de personas-

Ve el programa de televisión * Sexo Crosstabulation

Count

		Sexo		Total
		Varón	Mujer	
Ve el programa de televisión	Si	300	310	610
	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

			Sexo		Total
			Varón	Mujer	
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



Cuadro bivariado para verificar hipótesis

**Ver el
programa
de
televisión**

Sexo

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

Cuadro bivariado para verificar hipótesis

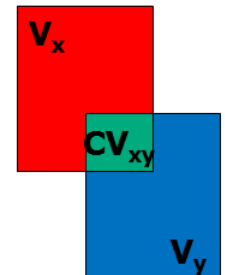
**Ver el
programa
de
televisión**

Sexo

	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



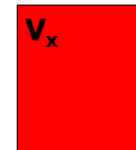
Cuadro bivariado para verificar hipótesis

**Ver el
programa
de
televisión**

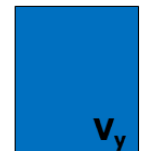
Sexo		
	Varón	Mujer
Miran el programa de televisión	60%	60%
No miran el programa de televisión	40%	40%
	100%	100%

$d\% = 0\%$

Independencia
estadística entre las
variables



$CV_{xy} = 0$



Cuadro bivariado para verificar hipótesis

**Ver el
programa
de
televisión**

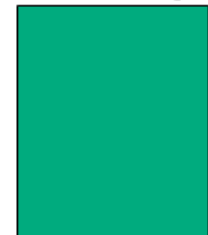
Sexo

	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

$$V_x \equiv V_y \equiv CV_{xy}$$





Cuadros bivariados para verificar hipótesis

Reglas para el procedimiento

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

Cuadro bivariado para verificar hipótesis

Condición de actividad por sexo

GBA / EPH 2º trim. de 2010

-En porcentaje-

Tabla de contingencia Condición de actividad * Sexo

% dentro de Sexo

		Sexo		Total
		Varón	Mujer	
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

Cuadro bivariado para verificar hipótesis

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

		Sexo		Total
		Varón	Mujer	
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

--- Se completara la clase que viene --