



METODOLOGÍA Y TÉCNICAS DE INVESTIGACIÓN EN CIENCIAS SOCIALES

Titular: Agustín Salvia

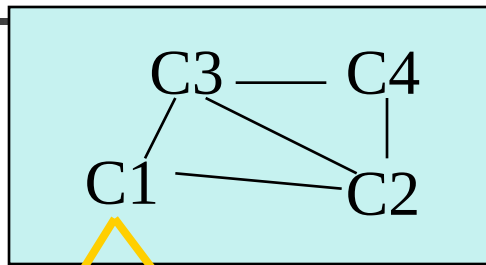
Teórico clase 8

**Unidades de análisis. Base de datos. Variables.
Estadística univariada. Gráficos**

Eduardo Donza

Proceso de operacionalización

Abstracción
inicial /
Situación /
Fenómeno de
interés



Marco teórico

- Conceptos
- Relaciones entre conceptos
- Hipótesis

Dimensiones de conceptos complejos

Subdimensiones

Variables

Posibilidad de generar índices que expresen el concepto

Indicadores

Dimensión 1

Dimensión 2

Subd 1 Subd 2

Subd 1 Subd 2

V1

V2

V3

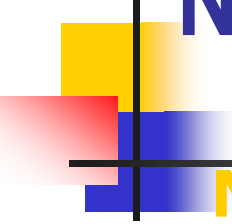
V4

I1

I2

I3

I4



Niveles de medición de las variables

Nominal

- Clasifican las unidades de análisis

Ordinal

- Clasifican y ordenan las unidades de análisis

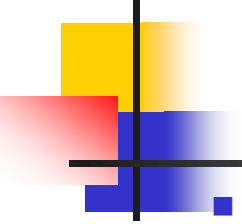
Intervalar

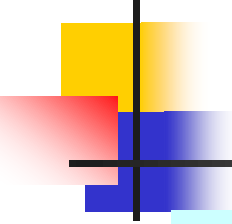
- Clasifican, ordenan las unidades de análisis y se puede medir distancia entre los valores que asumen

De razón

- Clasifican, ordenan las unidades de análisis, se puede medir distancia entre los valores que asumen y poseen un cero absoluto

Obtención de los diversos niveles de medición de las variables

- 
- En función de la naturaleza propia de los conceptos, el proceso de medición implicará considerar distintas reglas o propiedades formales que caracterizarán al conjunto de números o conjunto de resultados de la asignación de valores numéricos o símbolos. Obtenemos así distintos niveles de medición.
 - Este aspecto de la medición es importante en tanto que el nivel es uno de los determinantes del contenido informativo del dato y, al mismo tiempo, condiciona el tratamiento de los mismos y las técnicas de análisis.
 - El orden de los niveles de medición responde a una ordenación que expresa una jerarquía creciente de operaciones matemáticas y algebraicas que pueden realizarse a partir de una serie de reglas o propiedades formales poseídas de forma acumulativa: clasificatorias, de ordenación, de distancia y de origen.



LAS VARIABLES ESTADÍSTICAS

Escalas de medida

NOMINAL O DE CLASIFICACIÓN

Sexo, ciudad, situación laboral, religión, etc.

DE ORDEN JERÁRQUICO

Clase social, nivel educativo, escalas de actitud, etc.

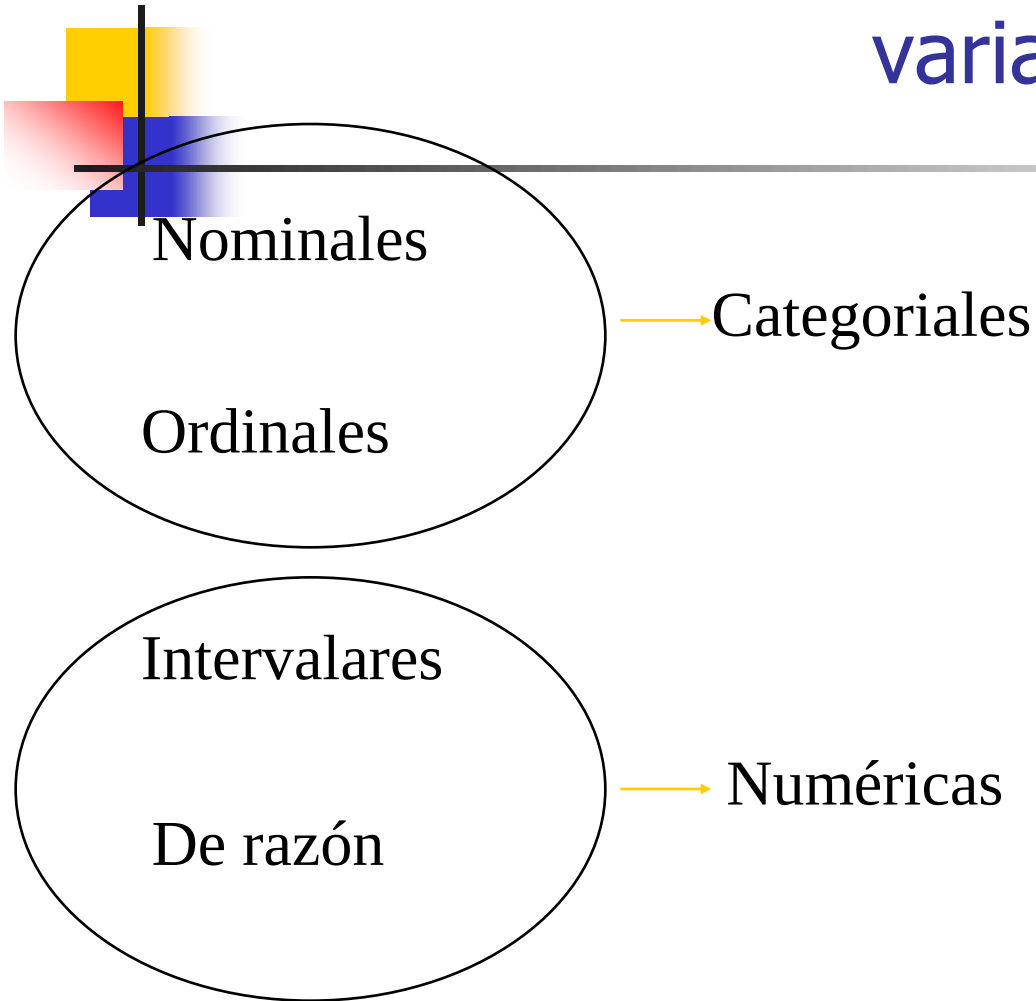
INTERVALARES

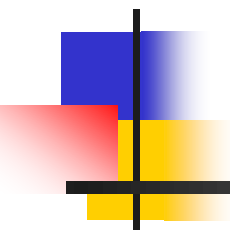
Fecha del calendario, factoriales, test, etc.

RAZÓN

Nº de hijos, ingresos, antigüedad, edad, etc.

Agrupación de niveles de medición de las variables





Bases de datos

Matriz o base de datos

Variables

Unidades de análisis

Nº de orden	Sexo	Edad	Lugar de residencia	Estado civil
1	2	58	2	2
2	1	10	1	1
.....				
480	1	85	2	4

Datos simulados

Sexo: 1. Femenino / 2. Masculino / 3. Otro

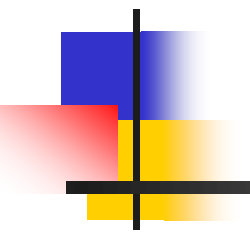
Edad: -- Directamente asume un valor ---

Lugar Residencia: 1. Ciudad de Buenos Aires / 2. Conurbano bonaerense / 3. Otro

Estado civil: 1. Soltera/o 2. Casada/o - Unida/o de hecho 3. Separada/o - Divorciada/o 4. Viuda/o

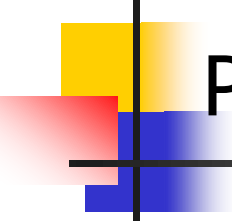
Distribución de frecuencia: Estado civil de los asistentes a ver cierta película de un cine del GBA

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



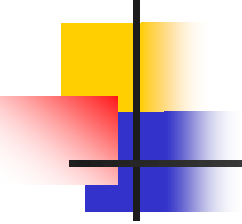
Posibilita analizar gran cantidad de datos

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

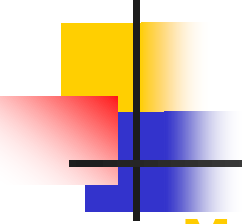
- Intentan reflejar un valor típico para describir un conjunto de casos

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

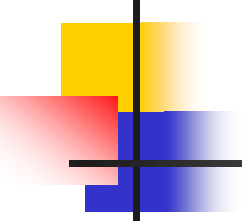
- Rango
- Desvío estándar o típico
- Varianza
- Producto de proporciones ($p \cdot q$)
- Coeficiente de variación ($CV = \text{desvío estándar} / \text{media}$)

Medidas de posición

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

Estadística descriptiva

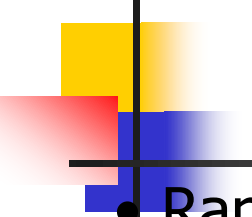
Medidas de tendencia central



- Moda o modo: es la categoría o el valor que ocurre más frecuentemente
- Mediana: es la categoría o el valor que corresponde al caso donde se acumula el primer cincuenta por ciento de los casos de una serie de observaciones ordenadas
- Media: es el punto de equilibrio de una serie de observaciones numéricas

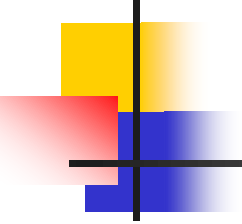
Estadística descriptiva

Medidas de dispersión

- 
- Rango ó Recorrido: es la diferencia entre el valor máximo y el mínimo
 - Varianza: es el promedio de las diferencias al cuadrado entre cada puntuación y la media
 - Desvío estándar: es la raíz cuadrada de la varianza
 - Producto de las proporciones: es el producto de la proporción (o porcentaje) de una categoría por la otra proporción (o porcentaje)
 - Coeficiente de Variabilidad: es la medida de la variación relativa. Se calcula como el desvío típico dividido la media

Estadística descriptiva

Medidas de posición

- 
- Percentil: es el valor en una distribución ordenada que corresponde al caso por debajo del cual queda un porcentaje determinado de casos (por ej, P.95 representa el valor del caso que tiene un 95% de casos inferiores). Sólo puede aplicarse en distribuciones ordenadas, no es válido para variables nominales.
 - Cuartil: es la medida de posición que corresponde cuando se divide una distribución ordenada de casos en cuatro partes equivalentes (Q1, Q2, Q3 y Q4) El Q4 coincide con el máximo.
 - Quintiles: distribución ordenada dividida en cinco partes.
 - Deciles: distribución ordenada dividida en diez partes.

Distribución de frecuencia

Variable con categorías

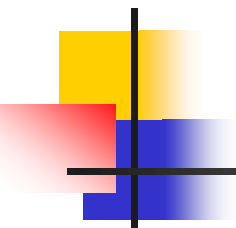
Formato típico de salida de paquete estadístico de análisis de datos

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

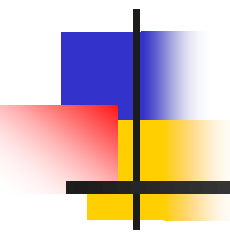
	Serie 1	Serie 2
Media	19,75	40,95
Desvío	3,96	7,53
CV	20,05%	18,39%

Fuente: Reuchlin, M. (1980). Compendio de Estadística. Método Conceptual. 1ª edición en lengua castellana. Madrid. Pablo del Río Editor.

Ejemplo de aplicación de deciles

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

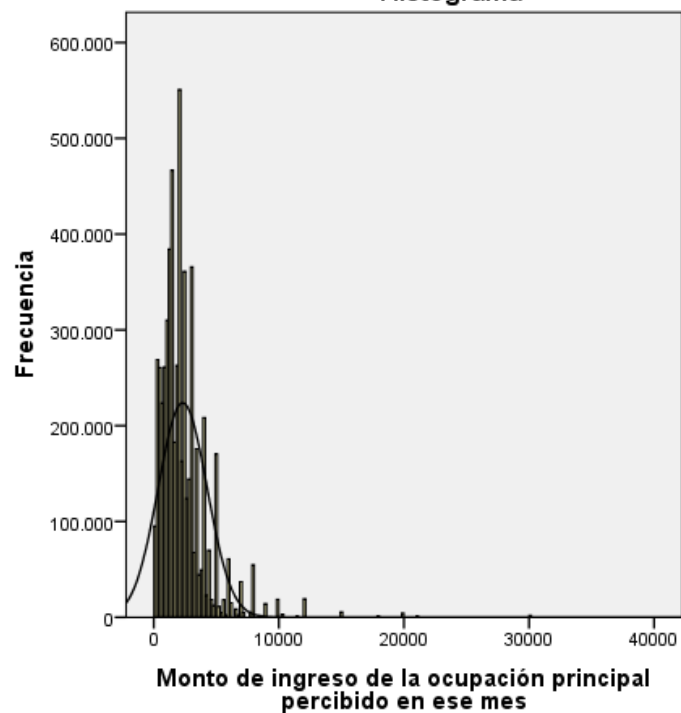


Gráficos

Forma de distribución

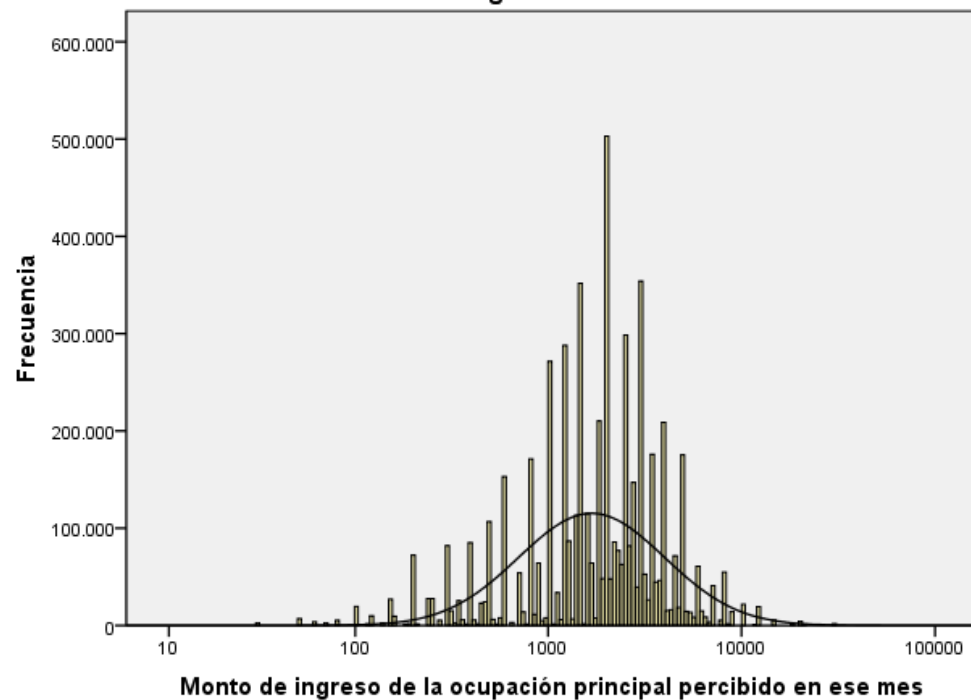


Histograma



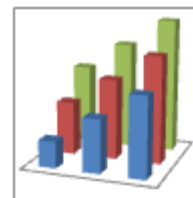
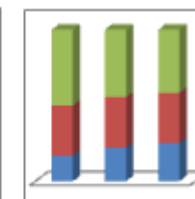
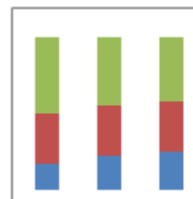
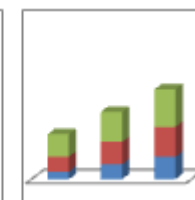
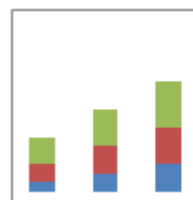
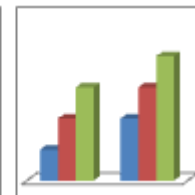
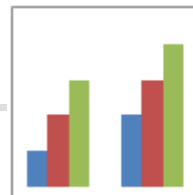
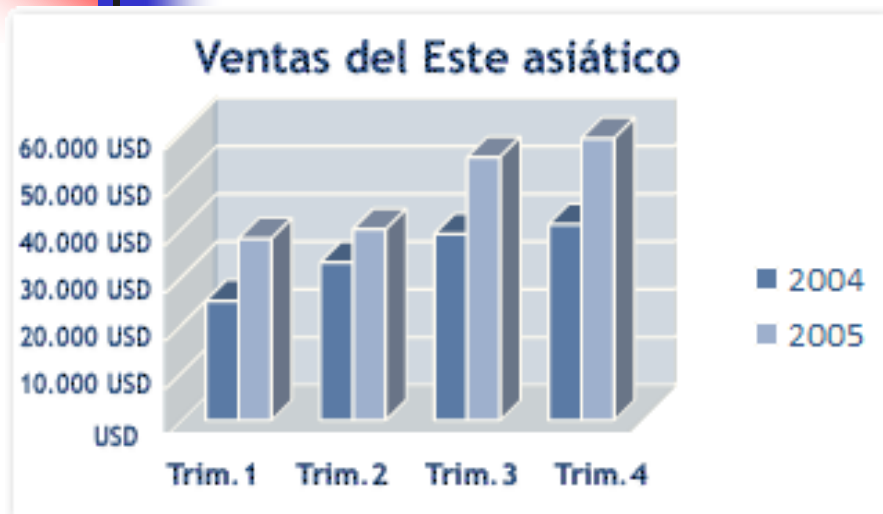
Casos ponderados por Ponderación

Histograma

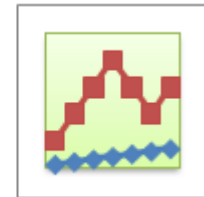


Casos ponderados por Ponderación

Gráficos de columnas

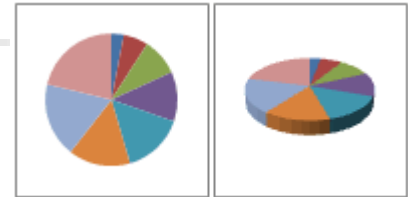
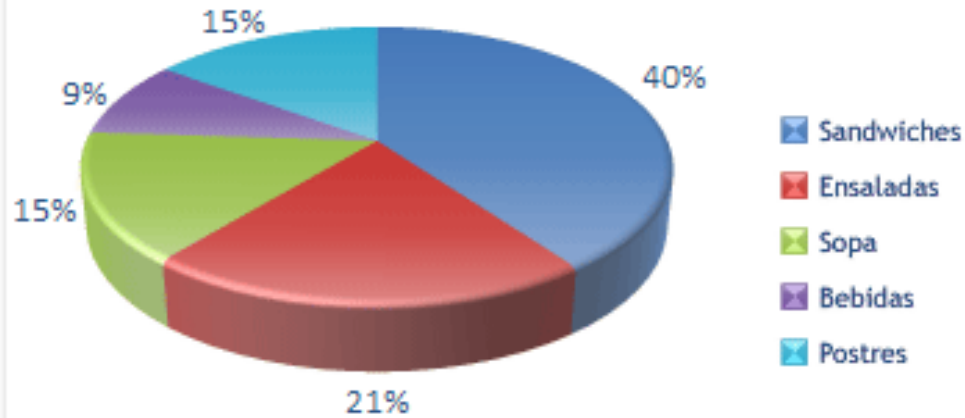


Gráficos de líneas

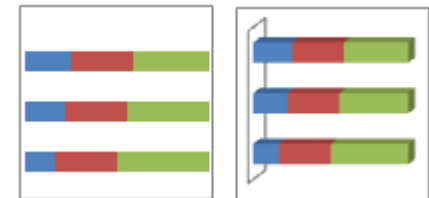
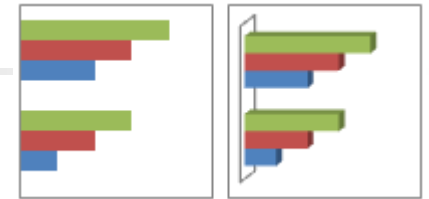
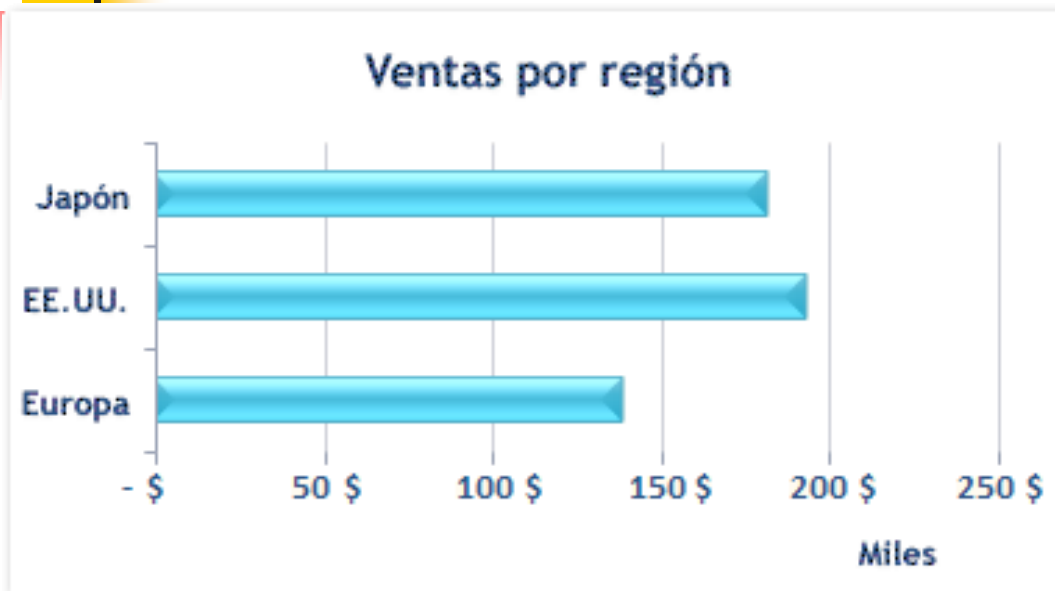


Gráficos circulares

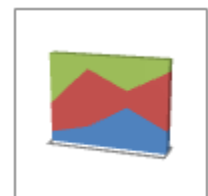
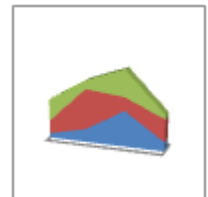
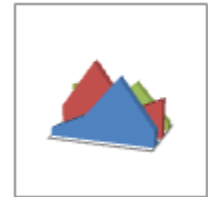
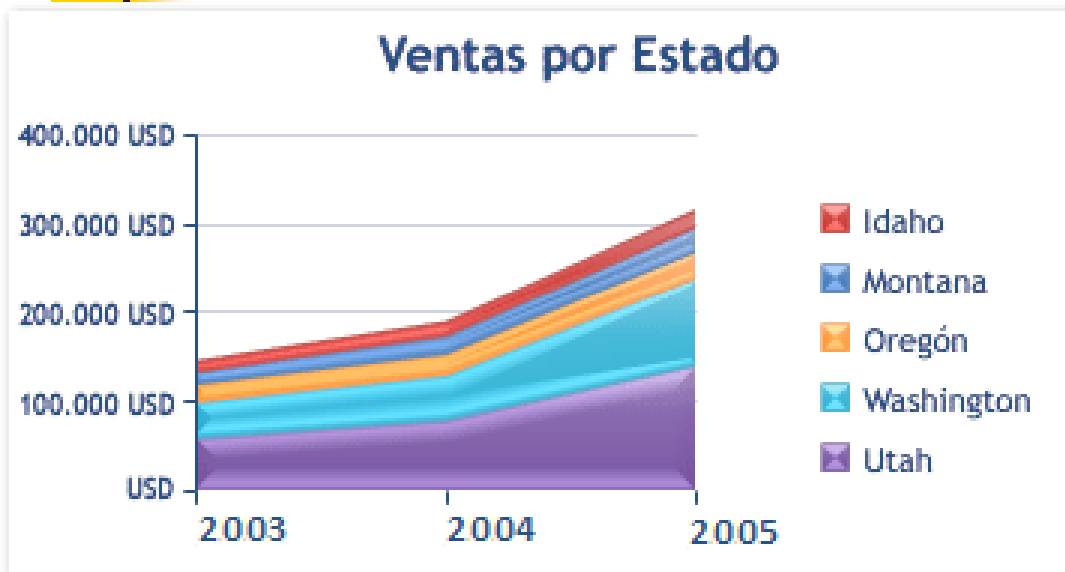
Ventas de comida



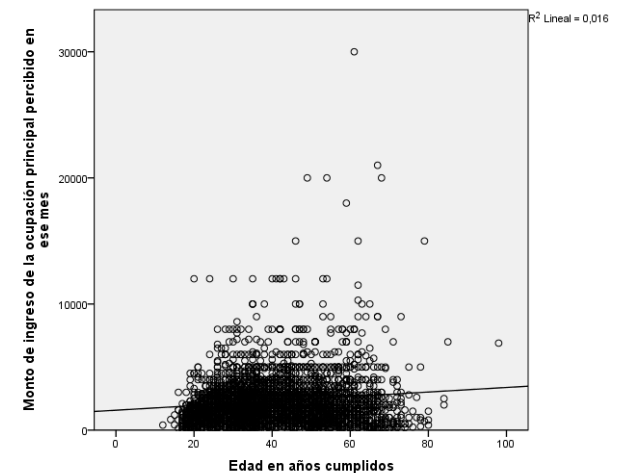
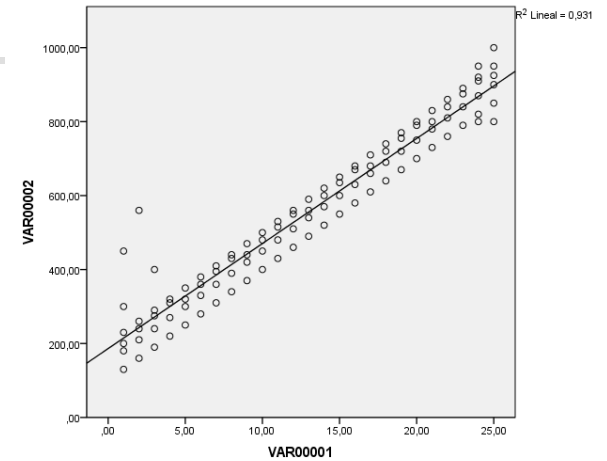
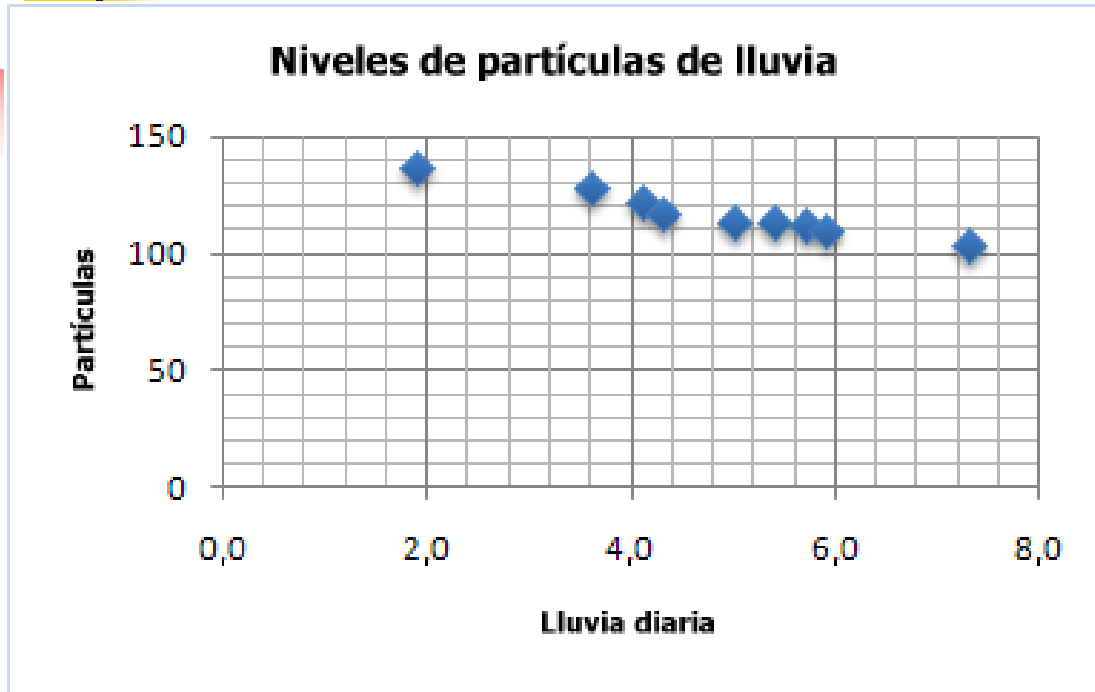
Gráficos de barras



Gráficos de área

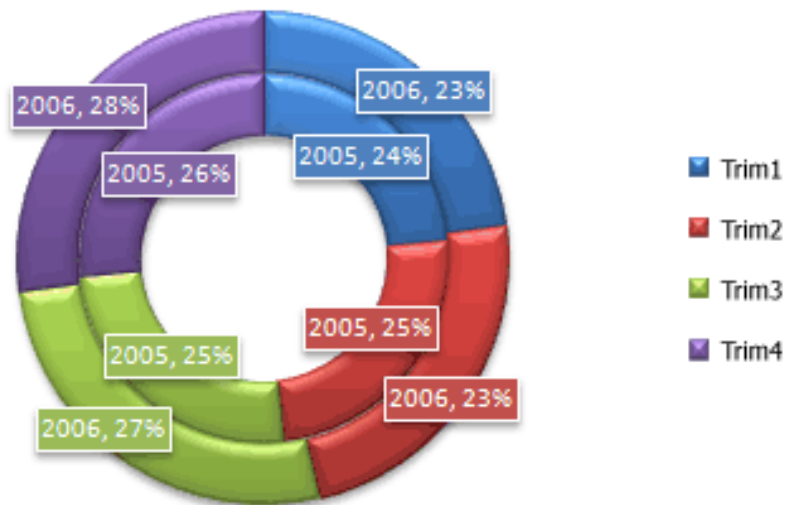


Gráficos de tipo XY (Dispersión)



Gráficos de anillos

Ventas trimestrales



Gráficos radiales

