

ANÁLISIS DE DATOS CUANTITATIVOS

TEMAS A ABORDAR

- ❖ **El dato: definición; fuentes.**
- ❖ **Matriz de datos**
- ❖ **Tabulación: las tablas de contingencia. Distribución de frecuencias. Cálculo de porcentajes.**
- ❖ **Análisis:**
univariados/bivariados/multivariados.
- ❖ **Medidas de tendencia central y de orden**
- ❖ **Representación gráfica de los datos**

La palabra dato tiene su origen etimológico en el término latino "Datum" que significa "**lo dado**".

La mayoría de los autores asumen que el investigador desempeña un papel activo respecto de los datos: ***el dato es el resultado de un proceso de elaboración, es decir, el dato hay que construirlo***

El dato obtenido es el resultado de una interacción entre el investigador, con sus supuestos básicos subyacentes, su enfoque del problema, los objetivos del estudio y su adhesión a algún paradigma teórico y/o metodológico.

LA VALIDEZ Y LA CONFIABILIDAD

La validez: refiere “al qué” , es decir se construye en relación con los conceptos relevantes

La confiabilidad: se apoya en los procedimientos que admiten el trabajo con éstos.

La confiabilidad se sustenta en **el “cómo”**, o sea, en la forma o lógica que asume la tarea, forma que deberá ser consistente, es decir, presentarse de manera similar en contextos semejantes.

Del instrumento, del observador y del informante.

Todo dato tiene una estructura compuesta por tres elementos: unidades de análisis, variables y valores.

Cualquier dato consistirá en:

- 1) Una unidad de análisis que...
- 2) En una variable asumirá...
- 3) Un determinado valor

Así lo que constituye un dato son estos tres elementos considerados en forma conjunta a través de las relaciones que mantienen entre sí. (Galtung)

REFERENCIA DE LA MATRIZ DE DATOS

UA= UNIDAD DE ANÁLISIS

V= VARIABLES

V1**SEXO:**

1. Masculino
2. Femenino

V2**EDAD**

(Pregunta Abierta)

V3**ESTADO CIVIL:**

1. Soltero
2. Casado/Unido
3. Separado/Divorciado
4. Viudo
99. Ns./Nc.

V4**¿PODRÍA DECIRME QUE LUGAR OCUPA EN SU FAMILIA?**

1. Cabeza de familia
2. Esposa/o cabeza de familia
3. Hijo/a
4. Hermano/a de
5. Padre/Madre
6. Otro especificar

V5**NIVEL DE ESTUDIOS:**

1. Sin Estudios
2. Primario Incompleto
3. Primario Completo
4. Secundario Incompleto
5. Secundario Completo
6. Terciario Incompleto
7. Universitario Incompleto
8. Terciario Completo
9. Universitario Completo o +
99. No sabe/No contesta

V6**SITUACIÓN OCUPACIONAL:**

1. Trabaja
2. Desocupado
3. Estudia
4. Jubilado o Pensionado
5. Rentista
6. Ama de Casa

MODELO DE MATRIZ DE DATOS

UNIDAD DE	VARIABLES						
ANÁLISIS	V 1	V 2	V 3	V 4	V 5	V 6	VN
UA1	1	20	1	2	4	3	
UA2	1	28	1	1	5	1	
UA3	2	46	2	1	6	6	
UA4	1	34	3	4	7	1	
UA5	2	29	2	2	9	1	
UA6	2	19	1	3	4	3	
UA7	2	54	2	2	4	5	
UA8	1	67	4	5	2	4	
UA9	1	32	2	1	9	1	
UA10	2	23	3	4	5	6	
UA11	1	19	1	3	2	2	
UA12	2	18	1	3	5	3	
UA13	1	36	3	1	8	1	
UA14	2	63	4	1	3	5	
UAN							

Una Matriz de Datos es una cuadrícula compuesta por filas y columnas, donde cada celda es la intersección de una fila y una columna.

La forma misma de la Matriz y la definición, conducen a los 3 Principios de la recolección de datos establecidos por Galtung, J.:

1) Principio de Comparabilidad

2) Principio de Clasificación

3) Principio de Integridad

1) Principio de comparabilidad:

Cada combinación de Unidad de Análisis/Variable debe tener sentido, esto es, debe ser verdadero o falso que dicha combinación brinde un determinado valor.

2) Principio de clasificación: implica que el sistema de categorías de una variable debe cumplir con dos requisitos lógicos: debe ser *exhaustivo* y *excluyente*.

- ***Exhaustivo*** en el sentido que ninguna unidad quede fuera del sistema de categorías.
- ***Excluyente***, de manera que cada unidad asuma un valor y sólo uno en cada variable.

3) Principio de Integridad: *no dejar ninguna celda vacía.* Lo que antecede significa que hay que intentar que la cantidad de celdas sin información, es decir, "No sabe"/"No contesta", se mantenga lo más baja posible.

TABULACIÓN DE LOS DATOS

Luego de confeccionar la matriz de datos, se procede a la tabulación de los mismos.

La tabulación es el proceso mediante el cual los datos recopilados se organizan y concentran, con base a determinadas ideas o hipótesis, en tablas o cuadros para su tratamiento estadístico

“Plan de tabulación”: determinar de antemano qué resultados de las variables se van a presentar y cuáles relaciones entre las mismas se van a analizar, a fin de brindar respuesta al problema y los objetivos formulados.

PLAN DE TABULADOS DIAGNOSTICO DE COMUNICACIÓN SOCIAL. PROMIN. FLORENCIO VARELA

- ✓ Población total por intervalos de edad según sexo
- ✓ Población por intervalos de edad según lugar de nacimiento
- ✓ Nivel de instrucción de la población encuestada según edad y sexo
- ✓ Meses de vida y número de controles

ANÁLISIS E INTERPRETACIÓN DE LOS DATOS

*El **análisis** consiste en separar los elementos básicos de la información y examinarlos con el propósito de responder a las distintas cuestiones planteadas en la investigación.*

*La **interpretación** es el proceso mental mediante el cual se trata de encontrar un significado más amplio de la información empírica recabada. (Rojas Soriano, R.)*

ANALISIS UNIVARIADO

La estadística descriptiva es la que permite, mediante el cálculo de ciertas medidas, caracterizar el comportamiento de un grupo o población. La estadística descriptiva me permite contar con medidas resumen de todos los datos.

1) Lo que expresan las *medidas de tendencia central* es la forma en que se comporta el grueso de la población: **modo o moda, mediana y media aritmética**

UNIVARIADO

Inmigrantes provocan inseguridad

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	En desacuerdo	302	70,2	70,2	70,2
	Algo de acuerdo	73	17,0	17,0	87,2
	De acuerdo	55	12,8	12,8	100,0
	Total	430	100,0	100,0	

Nivel de Culpabilizacion de Inmigrantes

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Bajo	250	58,1	58,1	58,1
	Medio	118	27,4	27,4	85,6
	Alto	62	14,4	14,4	100,0
	Total	430	100,0	100,0	

Categoría de Ocupacional del PSH

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Patrón	25	5,8	7,6	7,6
	Cuentapropista	20	4,7	6,1	13,7
	Empleado privado	183	42,6	55,6	69,3
	Empleado estatal	99	23,0	30,1	99,4
	Ns/Nc	2	,5	,6	100,0
	Total	329	76,5	100,0	
Missing	System	101	23,5		
Total		430	100,0		

LA MODA O MODO

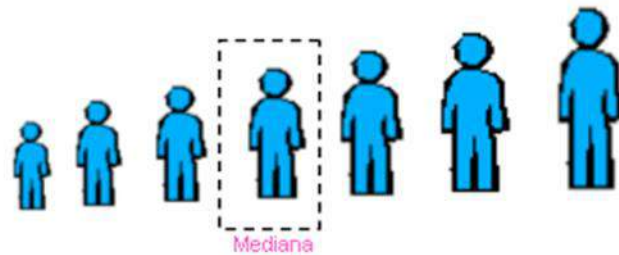
Es, por definición, la categoría que presenta la mayor concentración de frecuencia. **Es el valor que se presenta u ocurre con la mayor frecuencia.**

Ejemplo de edad:

80 personas tiene 25 años y 85 tienen 27 y no hay un año que tenga más personas, los 27 años de las 85 personas, ese es el modo.

MEDIANA

Es el punto o valor numérico que deja por debajo (y por encima) a la mitad de las puntuaciones de una distribución. Si digo la mitad, un primer cincuenta por ciento ¿Qué necesito? Un segundo cincuenta. Y el primer cincuenta lo tengo que definir de alguna manera ordenándolo, sino no tendría un primer cincuenta y un segundo cincuenta, entonces solo puedo aplicar la mediana a una variable que tenga como mínimo, un nivel de medición ordinal.



LA MEDIA ARITMÉTICA

Es el promedio, es lo mismo que se hace con las notas de los alumnos, si Uds. se sacan un 6 y luego un 8, saben que el promedio les da 7. Se suman las puntuaciones y se divide por el número de casos.

MEDIDAS DE ORDEN: cuartil, decil, percentil

Cuartiles: son los tres valores de la variable que dividen a un conjunto de datos ordenados en cuatro partes iguales. Q1, Q2 y Q3 determinan los valores correspondientes al 25%, al 50% y al 75% de los datos. Q2 coincide con la mediana.

1,	<u>2, 3,</u>	<u>4, 5,</u>	<u>6, 7,</u>	9
	2.5	4.5	6.5	
	↓	↓	↓	
	Q ₁	Q ₂	Q ₃	

2,	3,	4,	5,	6,	7,	9
	↓		↓		↓	
	Q ₁		Q ₂		Q ₃	

Deciles: son los nueve valores que dividen la serie de datos en diez partes iguales.

Los deciles dan los valores correspondientes al 10%, al 20%... y al 90% de los datos. D5 coincide con la mediana.

Primer trimestre de 2011		GRAN LA PLATA	E P H		
Grupo decílico	Mínimo	Máximo	Población en miles	% de población	Ingresos en miles de pesos
1	0	500	76	10	23.442
2	500	800	76	10	50.886
3	800	1.175	75	10	73.733
4	1.175	1.500	76	10	101.154
5	1.500	1.875	76	10	126.546
6	1.875	2.200	76	10	153.680
7	2.200	2.688	75	10	180.682
8	2.688	3.325	76	10	224.382
9	3.325	4.500	76	10	291.477
10	4.500	12.000	75	10	496.926
Total	0	12.000	756	100	1.722.909

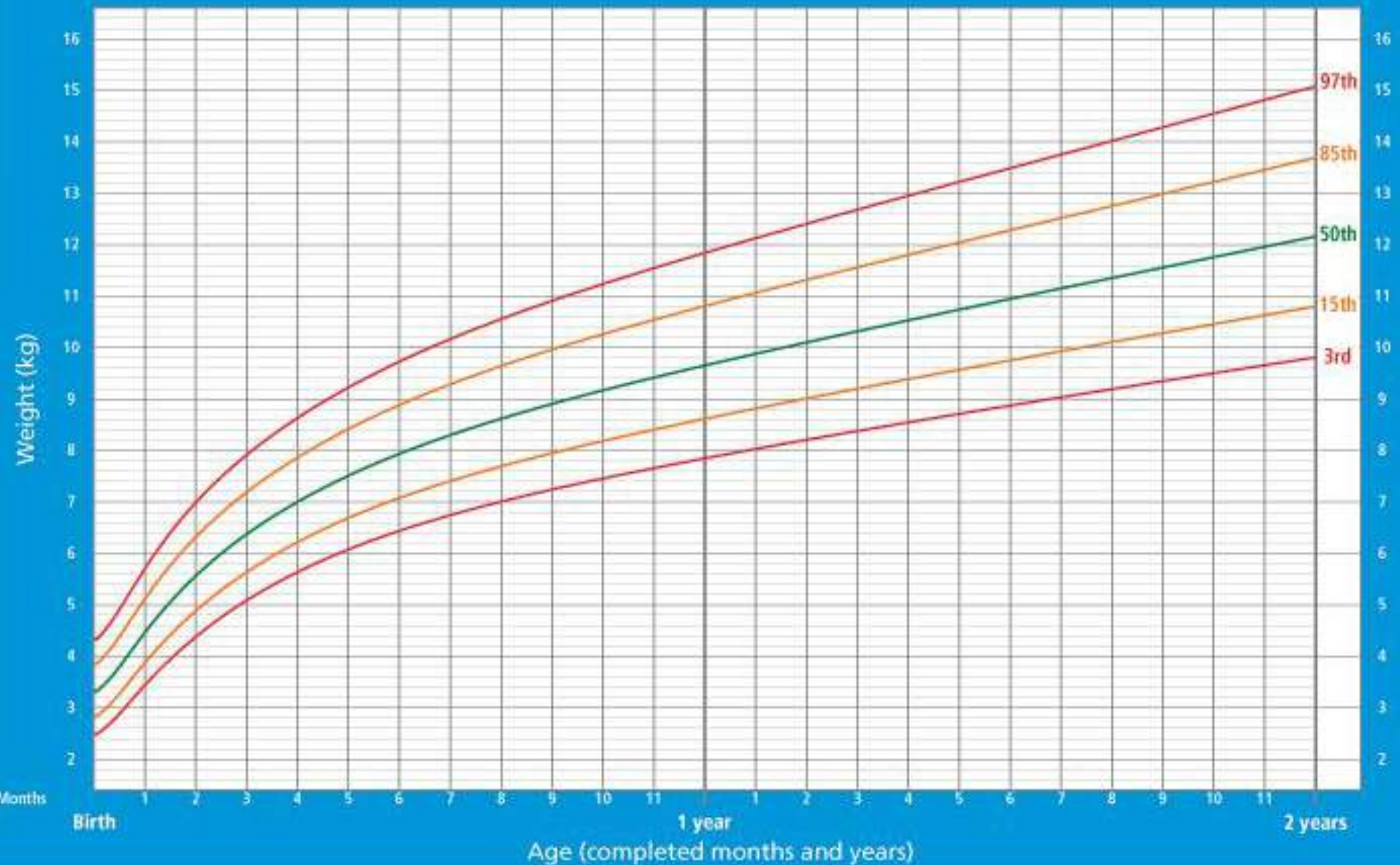
Percentiles: son los 99 valores que dividen la serie de datos en 100 partes iguales.

Los percentiles dan los valores correspondientes al 1%, al 2%... y al 99% de los datos. P50 coincide con la mediana.

Weight-for-age BOYS



Birth to 2 years (percentiles)



Qué significa que un niño esté en el percentil 10 de peso?

Que si lo comparamos con 100 niños de su misma edad, habrá sólo 10 que pesen menos, mientras que los otros 90 pesarán más que él.

Lo mismo vale para la talla. Si decimos que un niño está en el percentil 75 de talla, estamos diciendo que comparado con cien niños, 75 medirían menos y 25 más.

ANALISIS BIVARIADO

Con el resultado del recuento de los casos incluidos en la matriz, se ha elaborado una Tabla de Contingencia que contiene cifras absolutas, la misma nos va a permitir iniciar un análisis bivariante y/o multivariante de los datos.

TABLAS DE CONTINGENCIAS. DISTRIBUCIÓN DE FRECUENCIAS. CÁLCULO DE PORCENTAJES.

Es el resultado del cruce (o tabulación simultánea) de dos o más variables que también reciben los nombres de clasificación cruzada o tabulación cruzada.

MEDIDAS DE ASOCIACIÓN

El análisis de asociación es un instrumento estadístico concebido para analizar relaciones entre variables ligadas a las relaciones entre conceptos planteadas en el campo teórico.

Características de una asociación de dos variables:

- a) Existencia o no de una asociación
- b) La fuerza de la asociación
- c) La dirección de la asociación
- d) La naturaleza de la asociación

El resultado de los mismos varía entre -1 a 1 pasando por 0 (cero). 1 (uno) es el grado máximo de asociación, 0 (cero) indica independencia, o sea, falta de asociación y -1 (uno) es el grado máximo de asociación inverso a la hipótesis.

EJEMPLO: El grado de exposición a los medios masivos de comunicación de los habitantes de la ciudad de Buenos Aires.

Las preguntas iniciales son las siguientes:

- ❑ *¿Cuál es el grado de exposición a los medios masivos de comunicación de los habitantes de la ciudad de Buenos Aires en el año 1998?*
- ❑ *¿Cuáles son los factores que inciden en el grado de exposición a los medios masivos de comunicación de los habitantes de la ciudad de Buenos Aires?*

Objetivo general: Medir el grado de exposición a los medios masivos de comunicación de los habitantes de la ciudad y entre los objetivos específicos: Determinar el grado de exposición a los medios masivos según nivel socioeconómico.

Hipótesis:

"A mayor Nivel Socioeconómico de los habitantes de la ciudad de Buenos Aires menor grado de exposición a los medios masivos de comunicación y a la inversa."

La variable Nivel Socioeconómico se la considera como independiente (X) y el grado de exposición a los medios como dependiente (Y).

TABLA N° 3

GRADO DE EXPOSICIÓN A LOS MEDIOS SEGÚN NIVEL SOCIOECONÓMICO

GRADO DE EXPOSICIÓN A LOS MEDIOS	NIVEL SOCIOECONÓMICO				TOTAL		Dif. %
	MEDIO BAJO	%	MEDIO ALTO	%			
MEDIO BAJA	150	15	150	50	300	15 - 50 =	-35
MEDIO ALTA	850	85	150	50	1000	85 - 50 =	35
TOTAL	1000	100	300	100	1300		Sumatoria= 0

Mientras un 15 % de **nivel socioeconómico medio bajo** tiene una **exposición a los medios media baja**, un 50 % de **nivel socioeconómico medio alto** tiene una **exposición media baja.**"

Aquí es donde surge la diferencia de 35 puntos, entre ambos estratos y por ende se verifica la hipótesis de trabajo.

En efecto, a mayor nivel socioeconómico menor exposición a los medios y a la inversa.

Al leer la tabla se puede observar que las mayores frecuencias, tanto en términos absolutos como relativos, se concentran en la diagonal comprendida por las celdas **"b"** y **"c"**.

El primer paso de cualquier análisis es verificar *si la tabla "cierra"*, vale decir, si se cumplen las relaciones aritméticas que debe satisfacer cada cifra; en caso contrario es evidente que se ha producido algún error en la tabulación.

Los valores de la tabla están expresados en cifras absolutas, lo cual dificulta efectuar el análisis, ya que no se pueden realizar comparaciones al no existir una base común, por lo tanto, ***conviene calcular los porcentajes y obtener la Diferencia Porcentual y/o aplicar algún Coeficiente de Asociación.***

¿Cómo se lee y analiza una tabla de contingencia?

- 1) Toda tabla debe tener **su número y el título** correspondiente, indicando las unidades o cifras en que se expresan los datos. Por convención, se ubica primero a la variable dependiente y luego la variable independiente. En este caso: "**Grado de exposición a los medios según Nivel Socioeconómico.**"
- 2) Identificar claramente cuáles son las **unidades de análisis, las variables y sus respectivas categorías** y los niveles de medición de las mismas.
- 3) Calcular **los porcentajes y las diferencias porcentuales**. Para ello hay que determinar cuál será la base de los porcentajes.

Tipos de programas de TV preferidos según sexo
Gran Buenos Aires – Año 2003

			SEXO		Total
			Masculino	Femenino	
TIPOS DE PROGRAMA DE TV PREFERIDOS	Noticieros de TV en general	Cantidad	5050	5578	10628
		% por fila	47,5%	52,5%	100,0%
		% por columna	70,2%	77,4%	73,8%
	Musicales	Cantidad	776	910	1686
		% por fila	46,0%	54,0%	100,0%
		% por columna	10,8%	12,6%	11,7%
	Deportes o periodismo deportivo	Cantidad	1322	478	1800
		% por fila	73,4%	26,6%	100,0%
		% por columna	18,4%	6,6%	12,5%
	Teleteatro/Series de TV	Cantidad	12	146	158
		% por fila	7,6%	92,4%	100,0%
		% por columna	,2%	2,0%	1,1%
	Periodísticos y debates sobre temas de actualidad	Cantidad	12	48	60
		% por fila	20,0%	80,0%	100,0%
		% por columna	,2%	,7%	,4%
	Cine	Cantidad	21	35	56
		% por fila	37,5%	62,5%	100,0%
		% por columna	,3%	,5%	,4%
Otros	Cantidad		11	11	
	% por fila		100,0%	100,0%	
	% por columna		,2%	,1%	
Total	Cantidad	7193	7206	14399	
	% por fila	50,0%	50,0%	100,0%	
	% por columna	100,0%	100,0%	100,0%	

BIVARIADO

Nivel de Discriminación de Inmigrantes * Sexo

Crosstab

			Sexo		Total
			Varón	Mujer	
Nivel de Discriminación de Inmigrantes	Bajo	Count	108	125	233
		% within Nivel de Discriminación de Inmigrantes	46,4%	53,6%	100,0%
		% within Sexo	51,7%	56,6%	54,2%
	Medio	Count	55	66	121
		% within Nivel de Discriminación de Inmigrantes	45,5%	54,5%	100,0%
		% within Sexo	26,3%	29,9%	28,1%
	Alto	Count	46	30	76
		% within Nivel de Discriminación de Inmigrantes	60,5%	39,5%	100,0%
		% within Sexo	22,0%	13,6%	17,7%
Total	Count	209	221	430	
	% within Nivel de Discriminación de Inmigrantes	48,6%	51,4%	100,0%	
	% within Sexo	100,0%	100,0%	100,0%	

VISITA AL DENTISTA	SEXO					
	HOMBRES %		MUJERES %		TOTAL %	
SI	15	45.5	22	73.3	37	58.7
NO	18	54.5	8	26.7	26	41.3
TOTAL	33	100.0	30	100.0	63	100.0

Si observamos con atención, la diferencia porcentual entre los hombres que si visitan al dentista y las mujeres que lo hacen es mayor de 10 puntos porcentuales negativos ($45.5 - 73.3 = -27.8$), lo cual indica que probablemente sí existe asociación entre el sexo y el hecho de visitar al dentista por lo menos una vez al año.

Este procedimiento de análisis porcentual se denomina como la ***distribución de la variable dependiente en la variable independiente*** cuando se trata de probar una hipótesis de causalidad.

REPRESENTACIÓN GRÁFICA DE LOS DATOS

TIPO DE VARIABLE

TIPO DE GRÁFICO

NOMINAL U ORDINAL

GRÁFICO DE BARRAS AGRUPADO

CUANTITATIVA DISCRETA

GRÁFICO DE BARRAS

POLÍGONO DE FRECUENCIAS

INTERVALO O RAZÓN

HISTOGRAMA

CONTINUAS

POLÍGONO DE FRECUENCIAS

DIAGRAMA DE DISPERSIÓN

CUANTITATIVA O

DIAGRAMA DE SECTORES O GRÁFICO DE

CUALITATIVA

TORTA

PICTOGRAMAS Y CARTOGRAMAS

PIRÁMIDES DE POBLACIÓN

GRÁFICO DE BARRAS

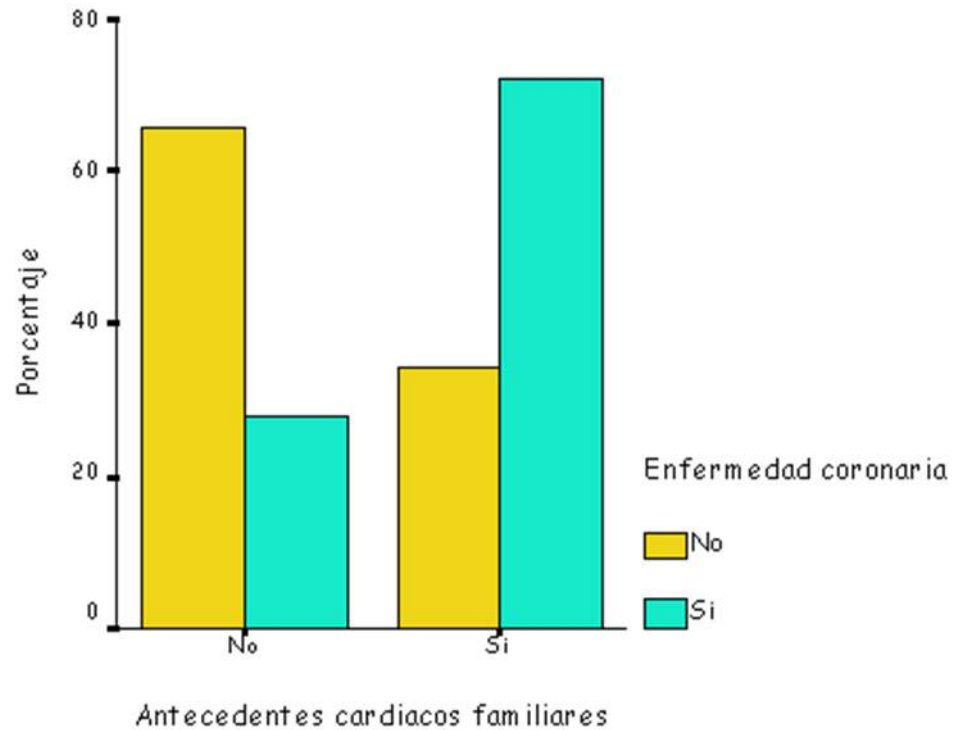
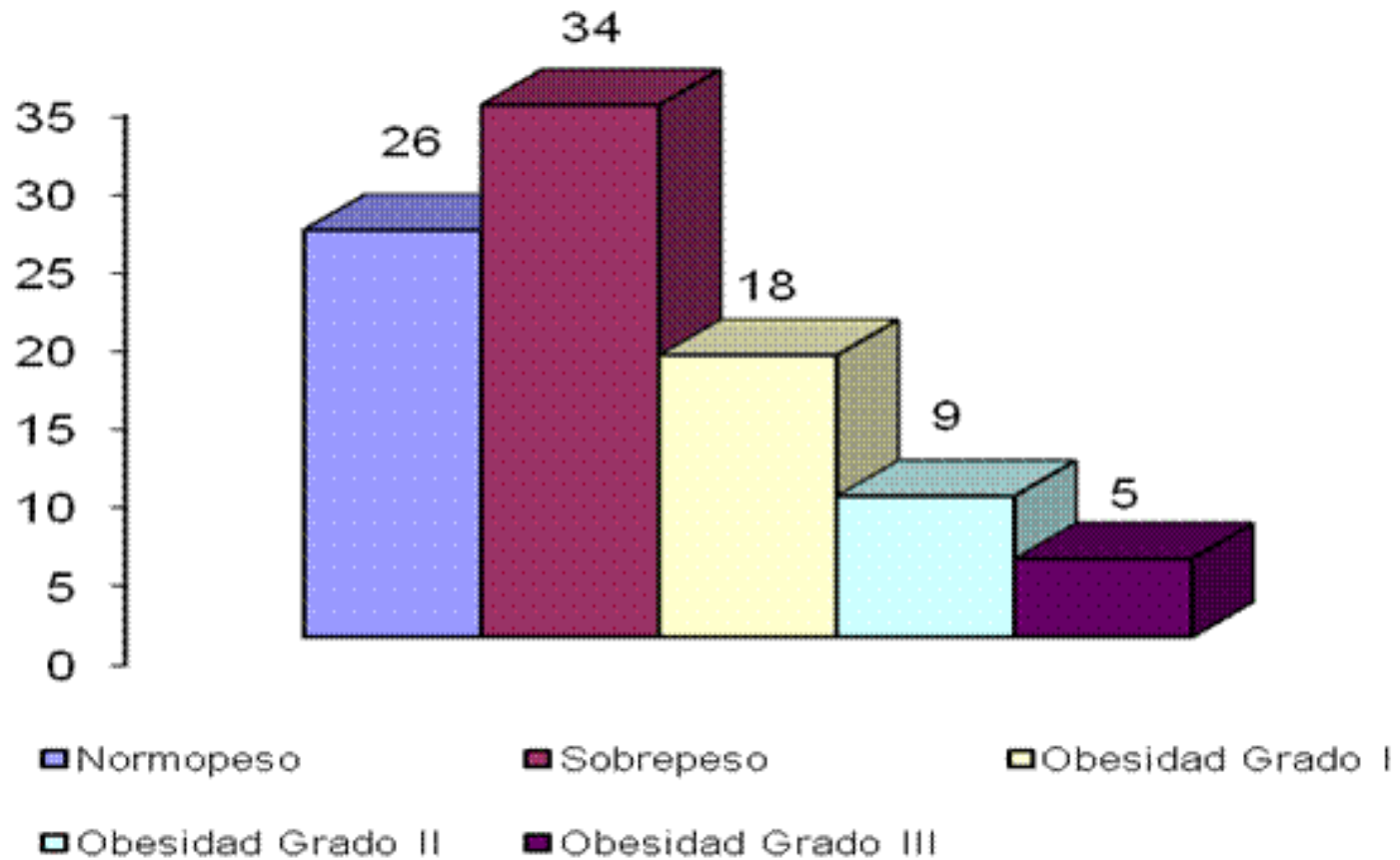
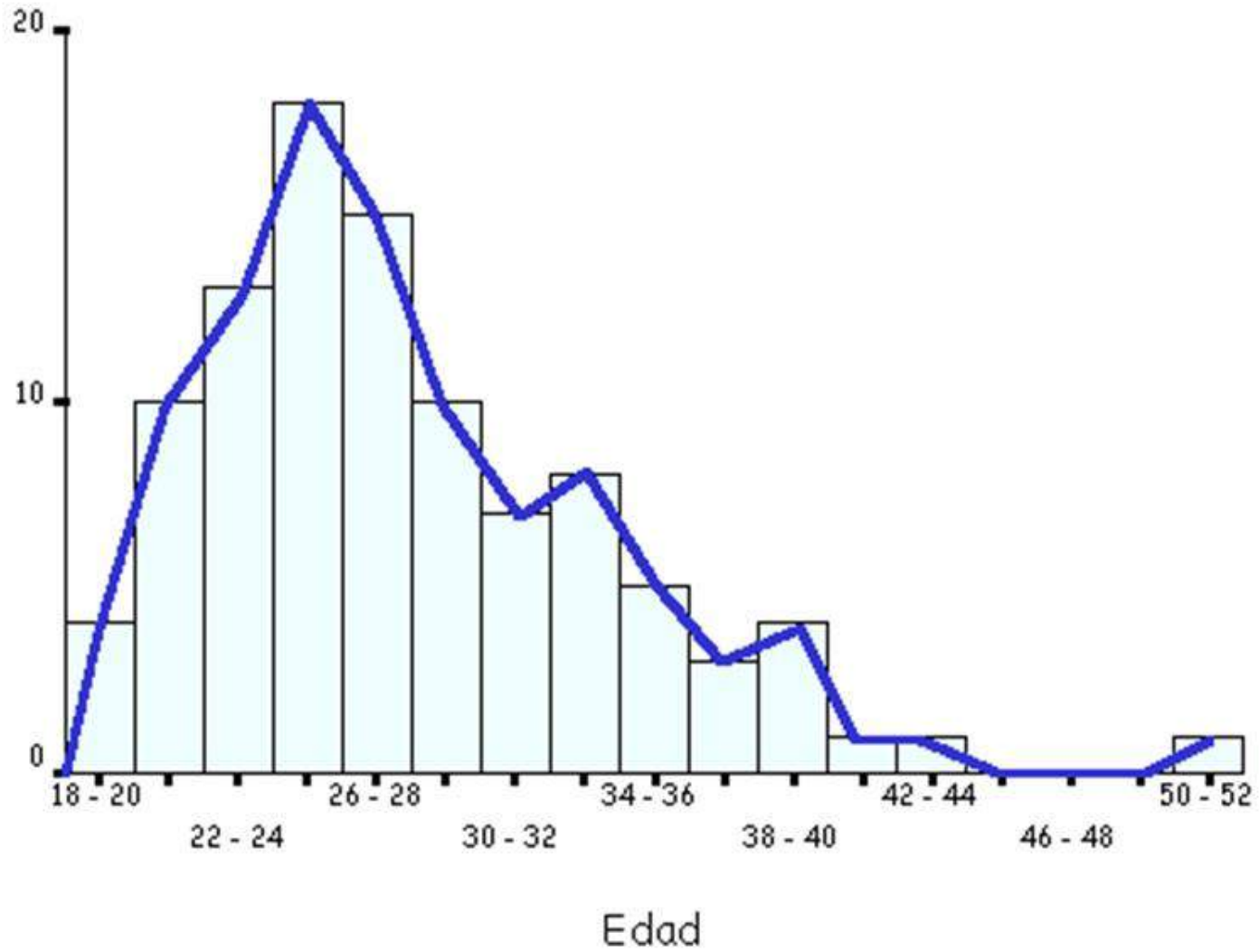


GRÁFICO DE BARRAS

ESTADO NUTRICIONAL



POLÍGONO DE FRECUENCIAS



HISTOGRAMA

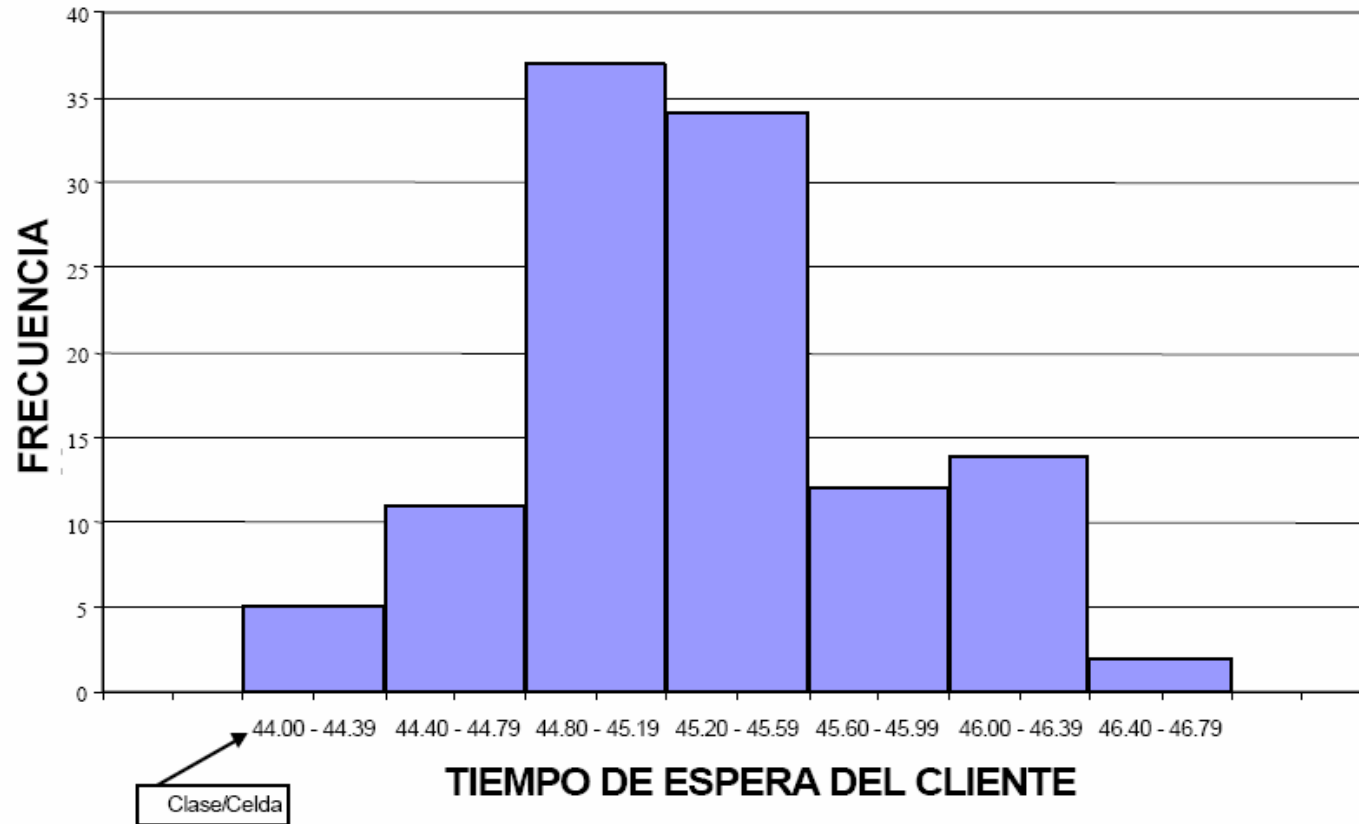


DIAGRAMA DE DISPERSION

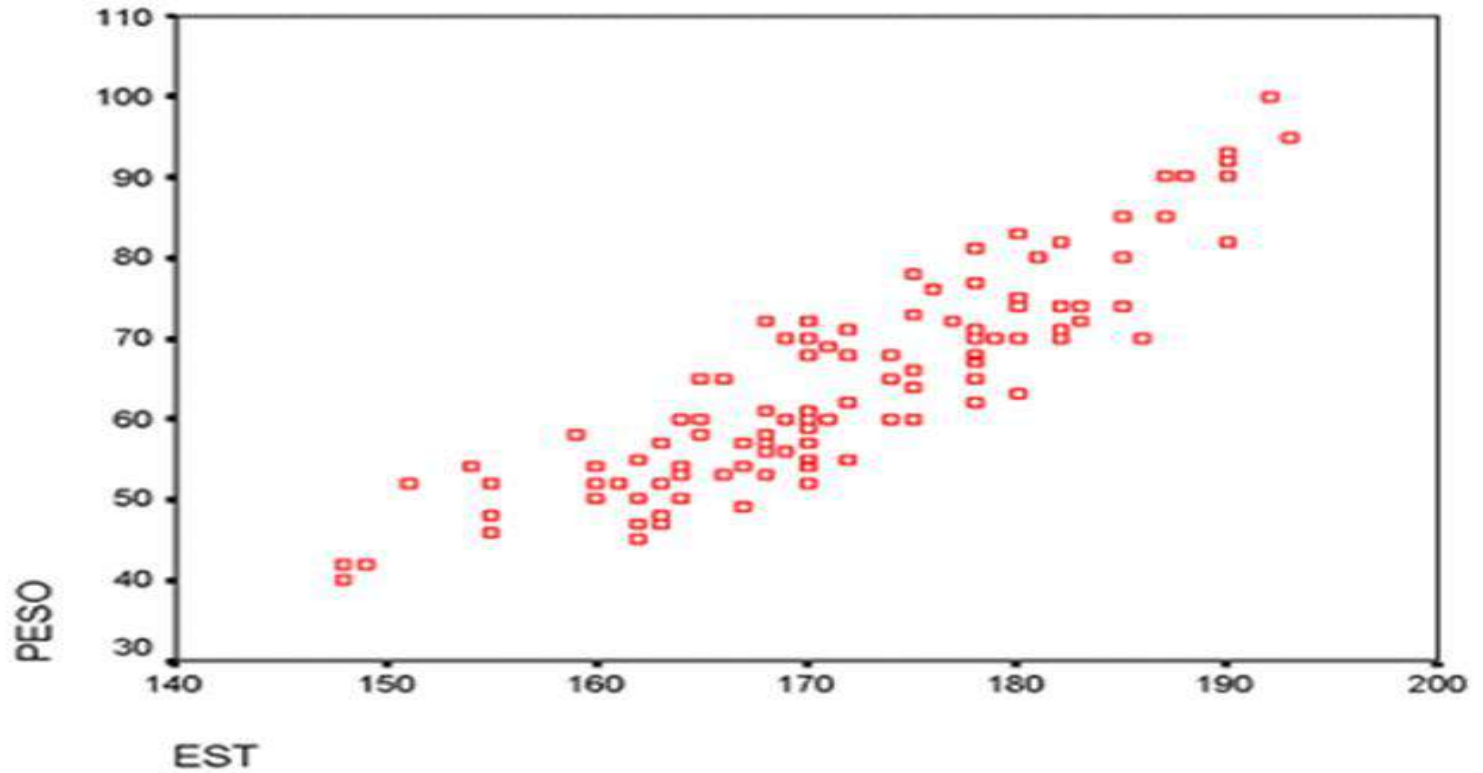
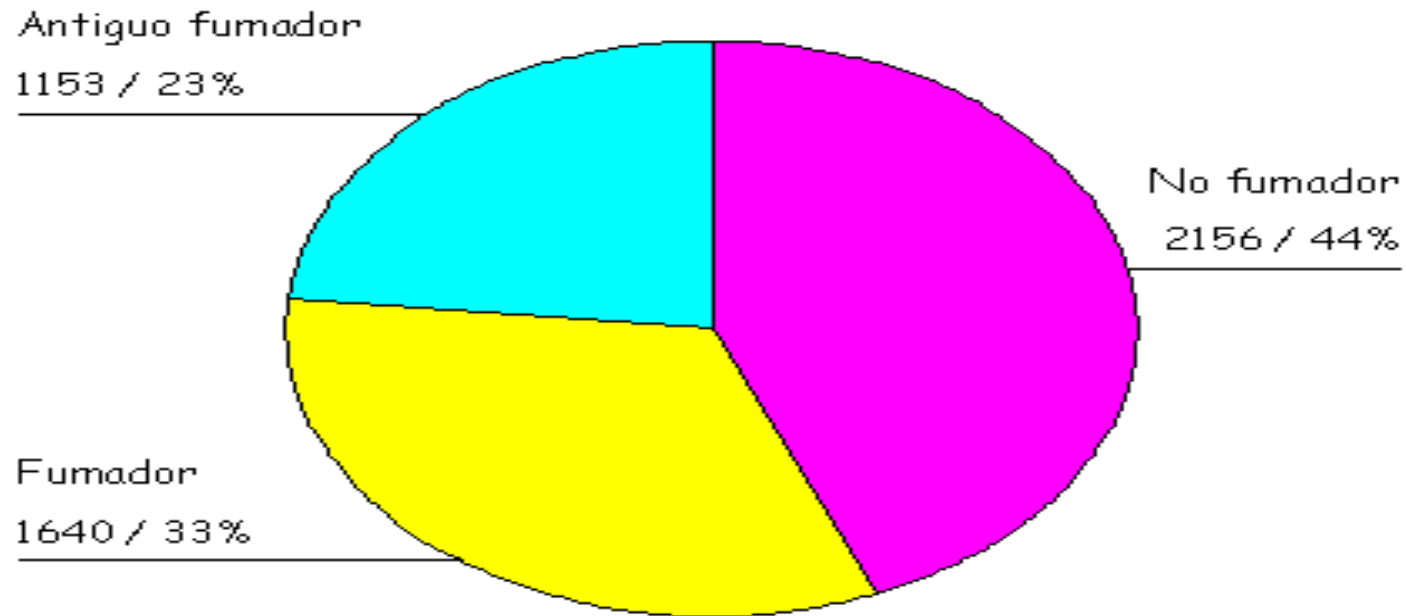
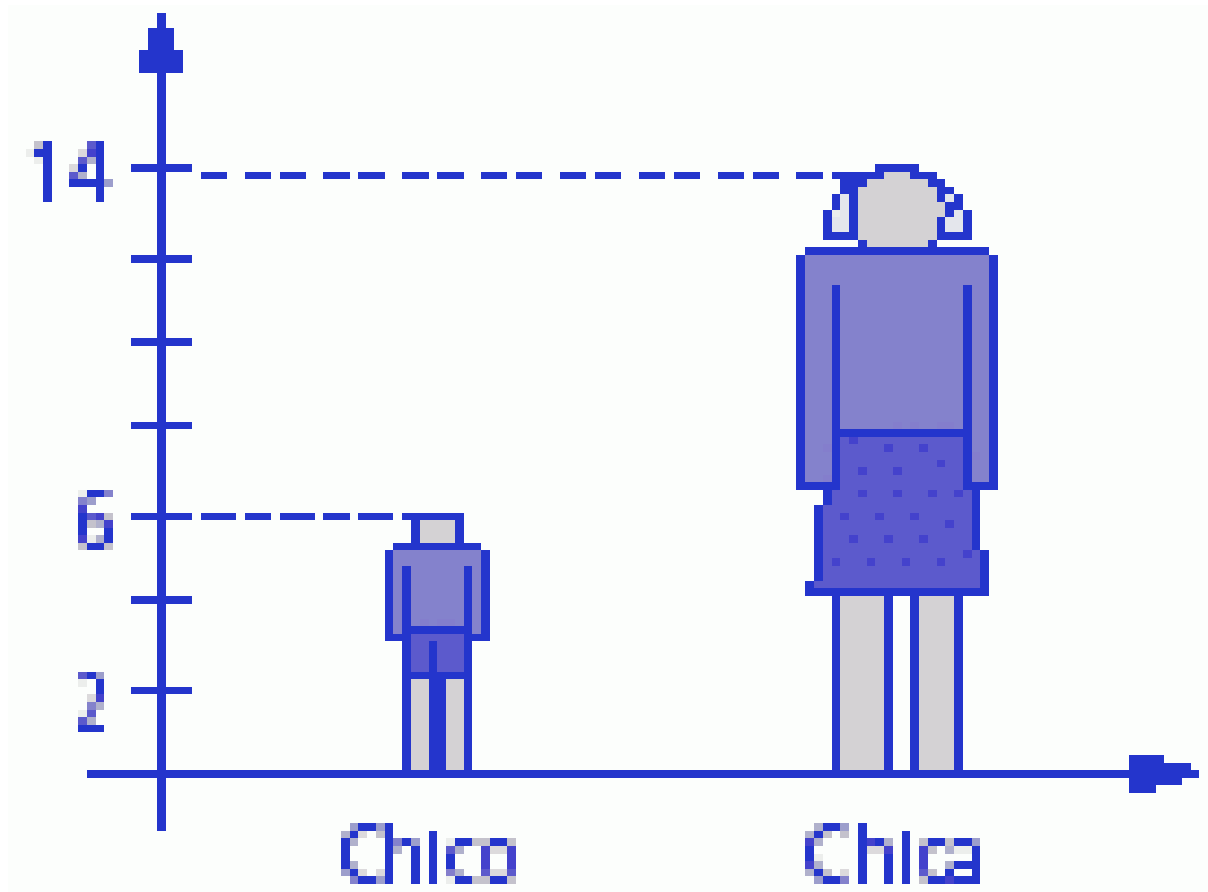


DIAGRAMA DE SECTORES O GRÁFICO DE TORTA

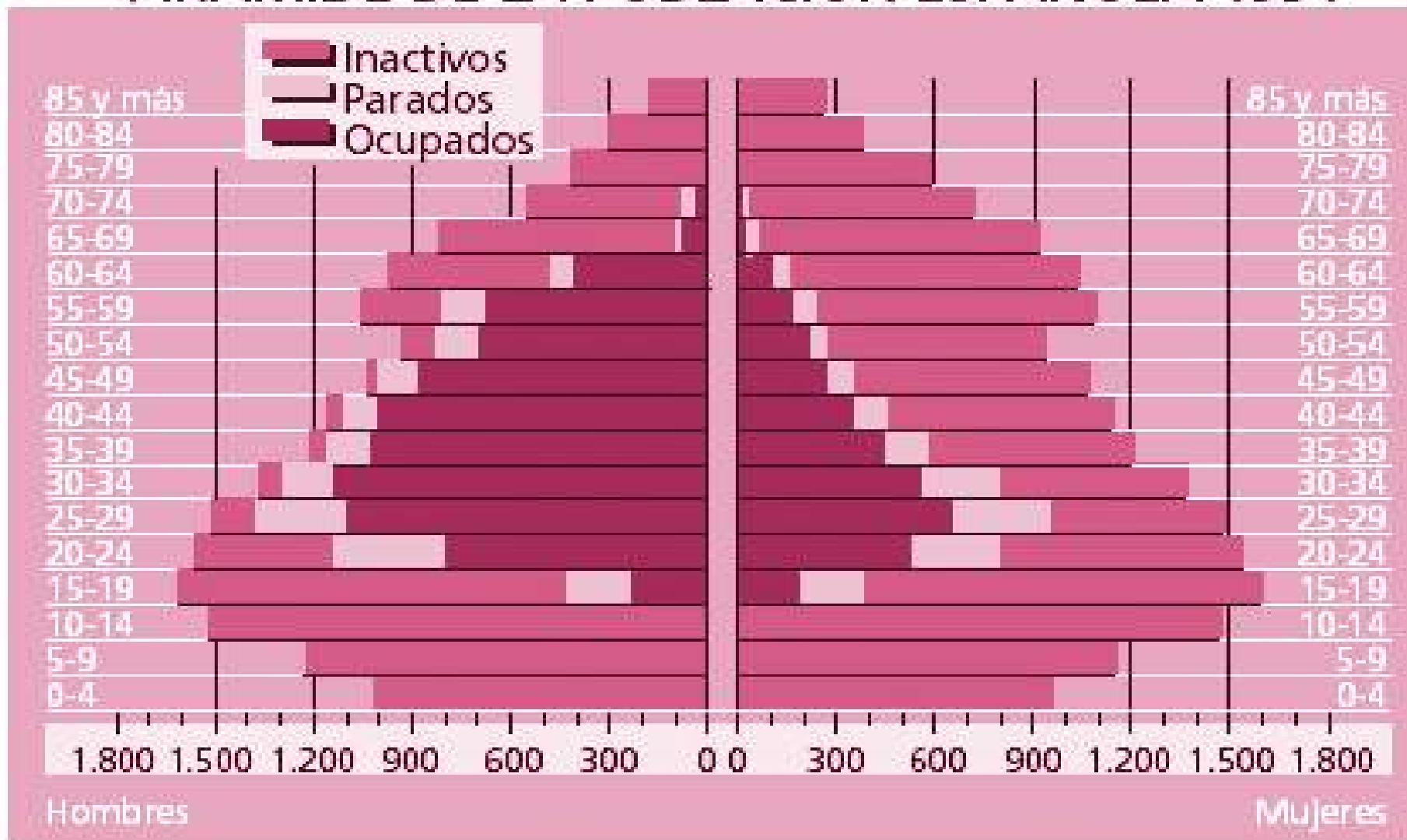


PICTOGRAMAS Y CARTOGRAMAS



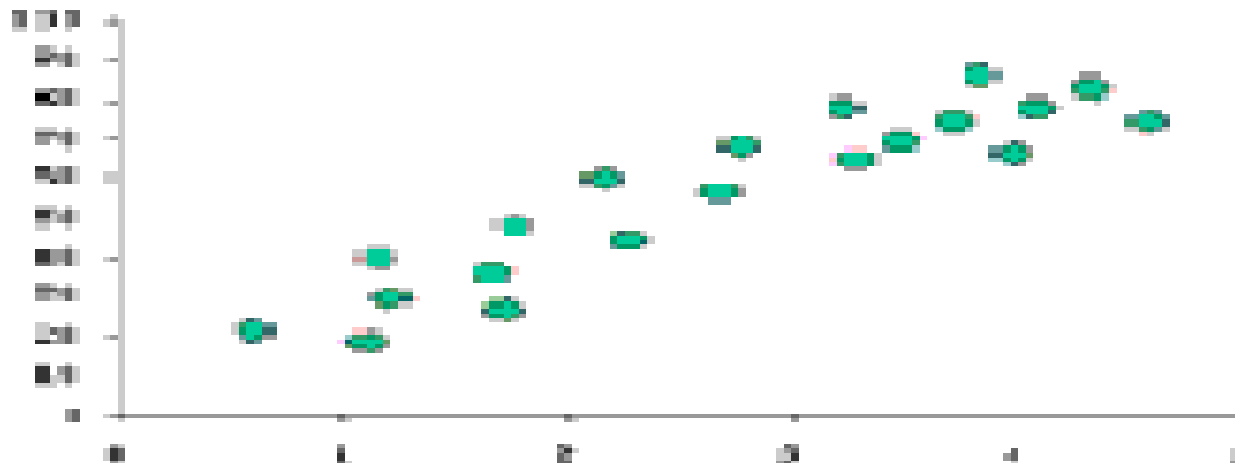
PIRAMIDES DE POBLACIÓN

PIRÁMIDE DE LA POBLACIÓN ESPAÑOLA 1991



LÍNEA DE TENDENCIA

RELACION LINEAL



SIN RELACION

