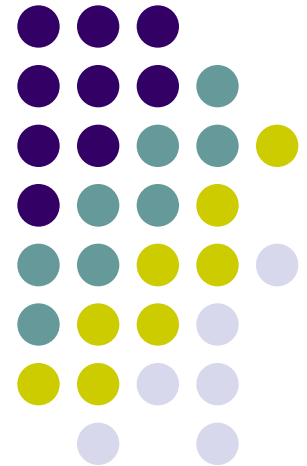
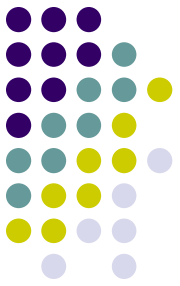


# Estadística Básica. Aplicación con SPSS

Sabina Pérez Vicente  
Unidad de Calidad  
APES Hospital Costa del Sol  
[sabina.perez.exts@juntadeandalucia.es](mailto:sabina.perez.exts@juntadeandalucia.es)

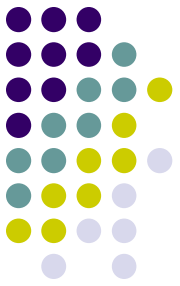


# Contenidos

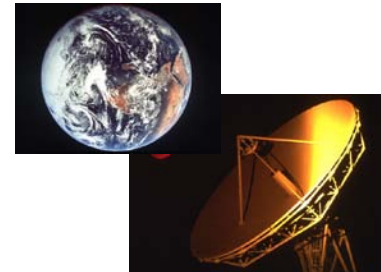
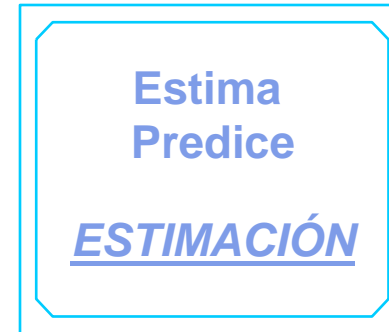


1. Introducción a la estadística y al paquete estadístico SPSS
2. Estadística descriptiva
  1. Tipos de variables
  2. Tablas de frecuencias
  3. Medidas de centralización y dispersión
  4. Representaciones gráficas
  5. Relación de 2 o más variables

# Clasificación de la estadística:

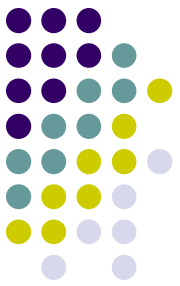


**Muestras**

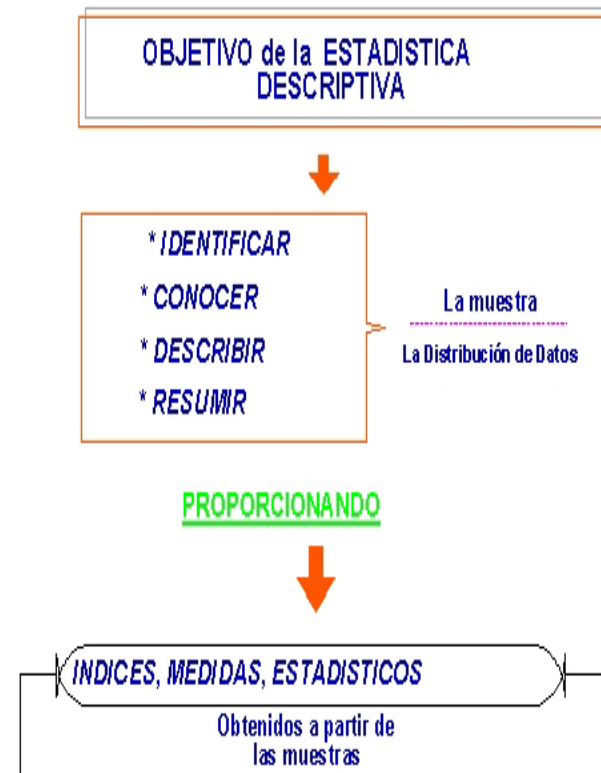


**Poblaciones**

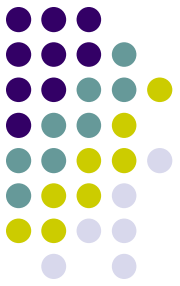
# Estadística descriptiva:



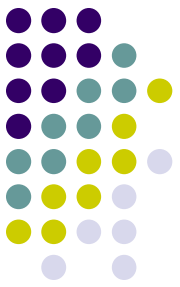
- La estadística descriptiva se encarga de la recogida y presentación de la información; nos indica como debe ser obtenida, señala las medidas necesarias para organizarla y sintetizarla, así como su representación gráfica.



# Definiciones:

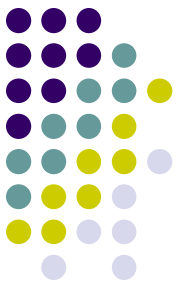


- **Población:** conjunto de sujetos de los que se desea extraer información
- **Muestra:** subconjunto de la población a la que tenemos acceso. Representativa de la población



# Definición operativa (1)

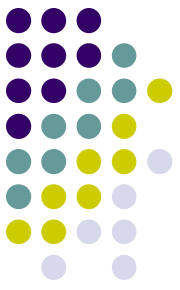
- Descripción detallada de las instrucciones necesarias para llevar a cabo las mediciones
- Debe incluir:
  - Escalas de medida o instrumentación
  - Forma de realizar las mediciones



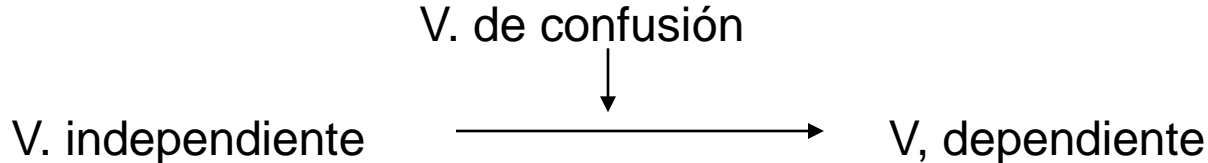
# Definición operativa (y 2)

- Justificación
  - Reproducibilidad del estudio
  - Homogeneización entre estudios
  - Comparación entre estudios
- Escalas validadas e instrumentos estandarizados
  - En mediciones estandarizadas no es necesario pero sí conveniente
  - Especificar en conceptos más variables

# 1ª Clasificación variables: Manipulación del investigador



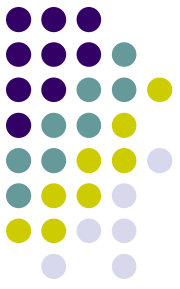
- Variables independientes: controladas por el investigador, variables de agrupación o variables demográficas
- Variables dependientes: variables desenlace o variables de respuesta



- Variables confusoras: intervienen en la relación de dos variables modificando la asociación existente entre éstas

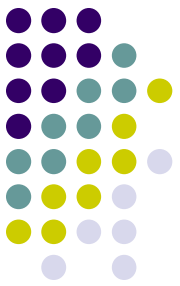


# 2ª Clasificación variables: Escala de medición

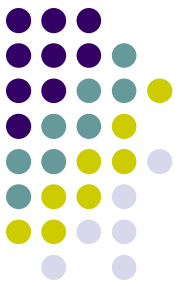


- **Cualitativas: miden cualidades**
  - Nominales: ninguna ordenación (grupo sanguíneo)
  - Ordinales: existe ordenación (nivel de satisfacción)
- **Cuantitativas: miden cantidades**
  - Discretas: números enteros ( $n^0$  caries)
  - Continuas: números con decimales (nivel de hb)

# Ejemplo: Pacientes con migraña



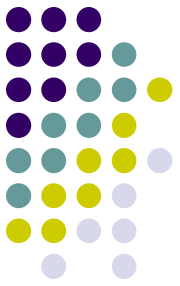
Variables	Tipo
■ Edad (edad en años del paciente)	■ Discreta
■ Productividad (% actividad laboral paciente migrañoso)	■ Continua
■ SEXO: 1_Varón, 2_Mujer	■ Dicotómica
■ Intensidad: 1_Leve, 2_Moderada, 3_Severa	■ Ordinal
■ Duración: 1_4-12H, 2_12-24H, 3_24-48H, 4_>48H	■ Ordinal
■ Localización: 1_Hemicraneal, 2_Holocraneal	■ Nominal
■ FBG: Fibrinógeno sérico	■ Continua
■ Leucocitos: recuento g. blancos	■ Discreta



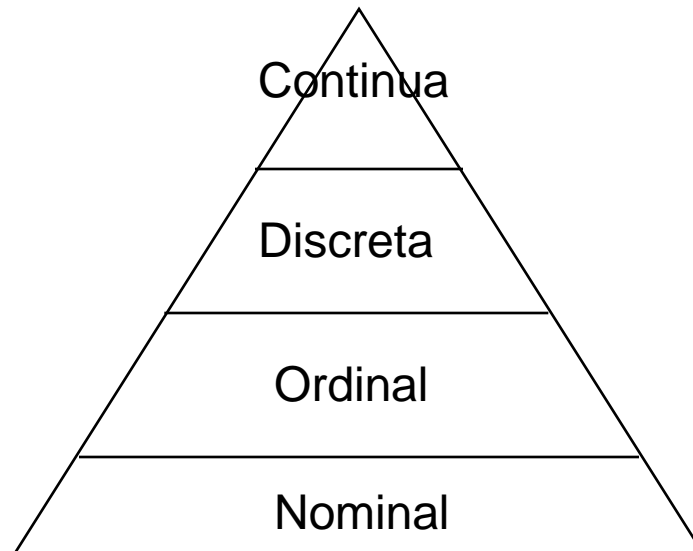
# Notas (1)

- V. numéricas se convierten en cualitativas agrupando. Ejemplo: niveles de sodio
  - $\leq 135$
  - 136-140
  - $> 140$
- V. cuantitativas mayor información que v. cualitativas
- V. cualitativas codificadas. Ejemplo: Intensidad de la migraña
  - 1 Leve
  - 2 Moderada
  - 3 Severa

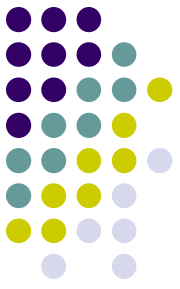
# Notas (2)



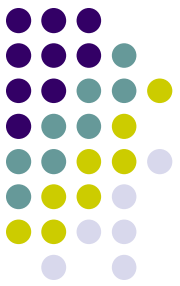
- V. Continuas se discretizan
- Grado en clasificación de variables



# Análisis según nº de variables



- Análisis univariante: cada variable de forma individual (descriptivo)
- Análisis bivariante: relación variables 2 a 2 (asociaciones y correlaciones)
- Análisis multivariante: relación simultánea de varias variables (factores de riesgo múltiples)



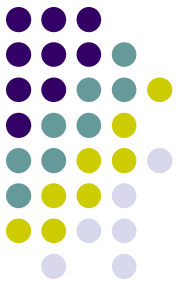
# 1. Construcción de un fichero de datos

- Base de datos (access, paradox, dbase, ...)
- Hoja de Cálculo (excel, lotus,...)
- Fichero de Texto (ascii)
- En el propio paquete estadístico

# Casos y variables (I)



- Casos: uno por fila
- Variables: una por columna
  - No deben contener caracteres especiales como ?, -, % ... (si se permite el guión bajo \_, acentos, eñe, punto)
  - No deben tener espacios en blanco
  - En algunos paquetes estadísticos el número máximo de caracteres admitidos para los nombres de las variables es de 8, truncando dichos nombres si se importan los datos desde otro programa que admita más caracteres
  - Los nombres de las variables no deben empezar por números.



# Casos y variables (II)

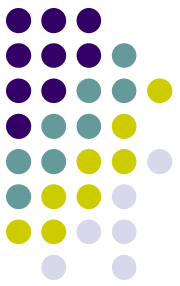
## Nombres válidos

- tas.1
- tas\_prev
- tas\_ini
- fechaingreso

## Nombres no válidos

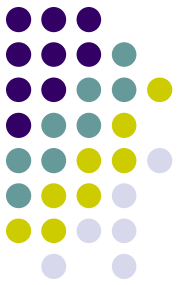
- 1.tas
- tas-prev
- tas ini
- fecha ingreso





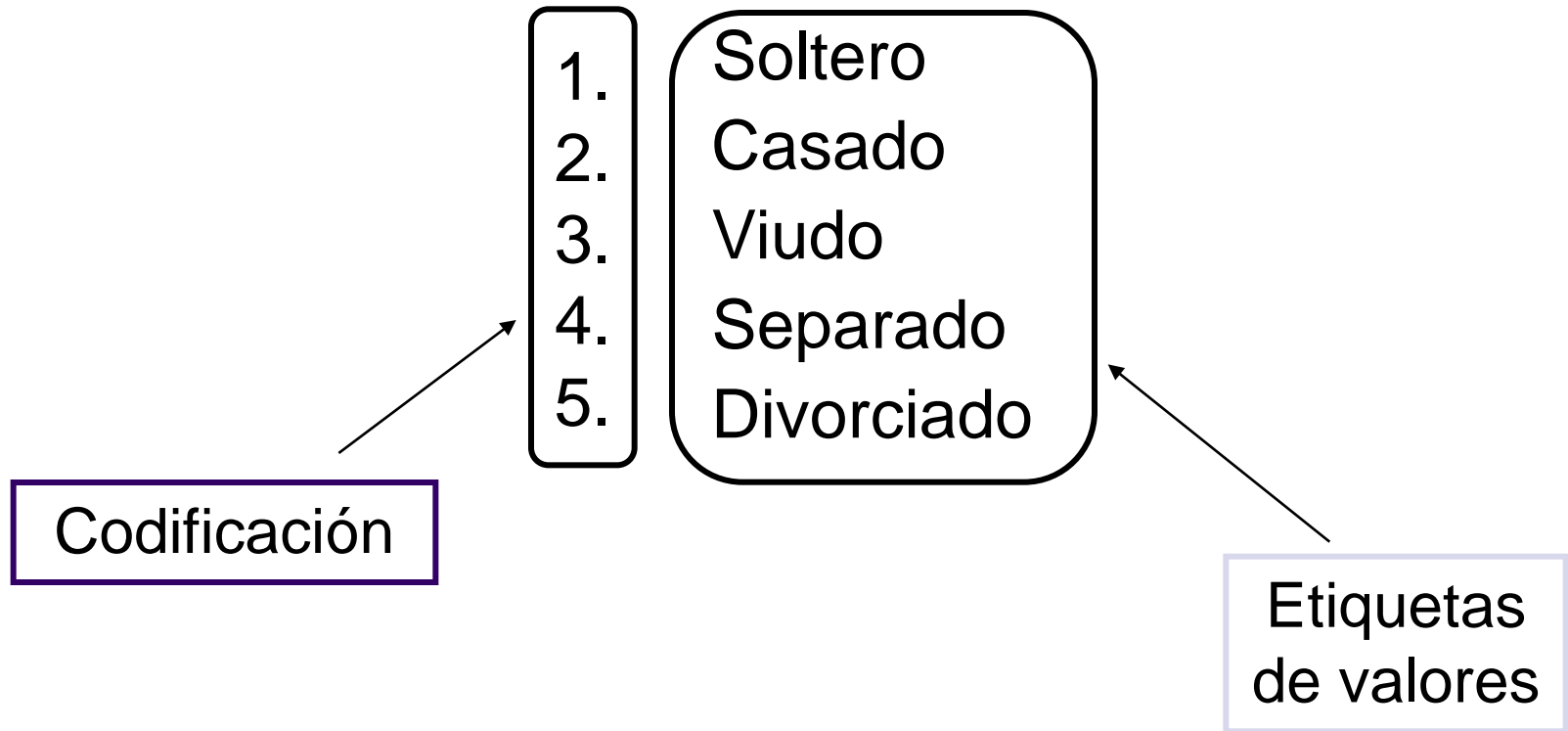
# Casos y variables (III)

- Mediciones en distintos grupos:
  - Añadir una nueva variable que indique el grupo de pertenencia
- Mediciones en el mismo grupo en distintos momentos
  - Añadir como variables distintas

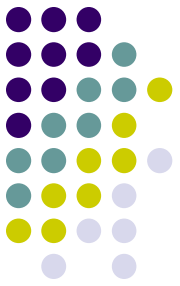


# Casos y variables (IV)

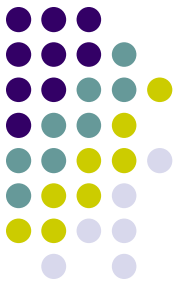
- Datos cualitativos codificados



# Casos y variables (V). Missing



- Dejar en blanco
- Codificar con algún valor fuera de rango
- (-9, -99,.....)



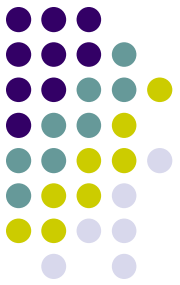
# Ejemplo

Edad	Fechingr	Tas_prev	Tas_3m	Grupo
40	11/2/01	150	140	1
49	14/3/01	140	140	1
50	19/5/01	130	-9	2
-9	20/6/01	120	120	2
48	30/4/01	-9	110	1
39	12/6/01	140	130	3
52	30/6/01	150	130	2



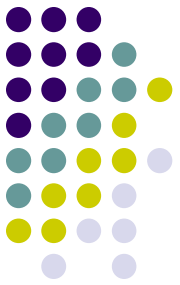
# Programa SPSS

- Programa modular
- Inicialmente creado para MSDOS
- Permite programar tareas periódicas
- Muy extendido en el ámbito sanitario
- Incorpora vistas de datos y variables
- Versión año 2008: 16.0
- Deficiente en algunos procedimientos



# Archivos de SPSS

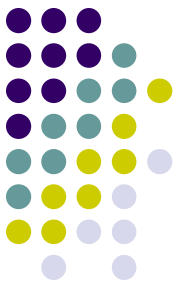
- Datos: fichero.sav
- Resultados: fichero.spo
- Sintaxis: fichero.sps
- Procesos: fichero.sbs
- Importa archivos realizados con otras aplicaciones (excel, access, dbase, ...)



# Etiquetas de variable

Pestaña vista de variables->Etiqueta

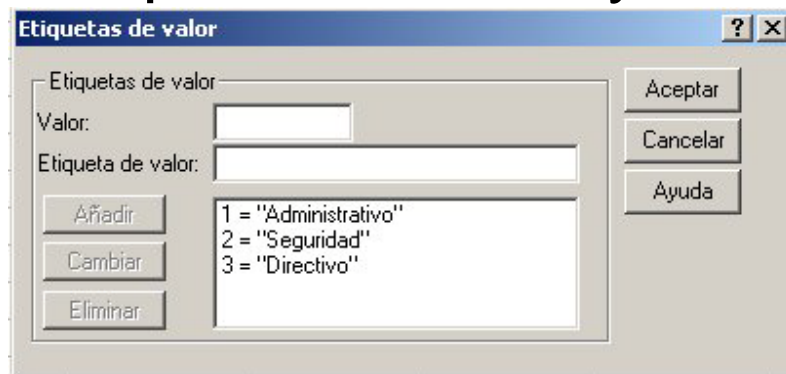
- Descripción de la variable. Puede contener caracteres especiales, espacios en blanco,...
- Ejemplo
  - Variable: `tas_prev`
  - Etiqueta: Tensión arterial sistólica previa



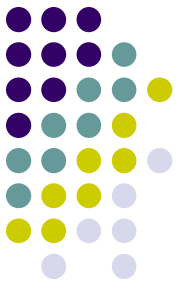
# Etiquetas de valor

Pestaña vista de variables->Valores

- Descripción de cada número en variables cualitativas
- Escribir el valor correspondiente
- Definir su etiqueta
- Pulsar Añadir para cada una y al terminar, Aceptar







# Recodificación

Transformar una variable cualquiera en otra categórica

- **Peso RN**

$\leq 2500$ gr.  $\longrightarrow$  1 (Bajo)

$> 2500$  gr.  $\longrightarrow$  2 (Normal)

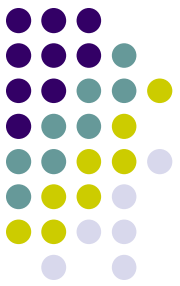
- **Dolor**

1 (Nulo)  $\longrightarrow$  1 No

2 (Leve)  $\longrightarrow$  2 Sí

3 (Moderado)  $\longrightarrow$  2 Sí

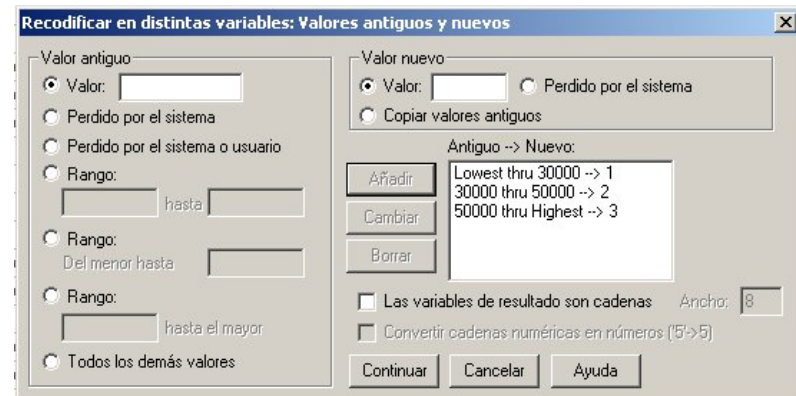
4 (Intenso)  $\longrightarrow$  2 Sí

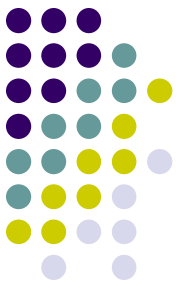


# Recodificación en SPSS

Transformar->Recodificar->En distintas variables

- Seleccionar nombre de variable a recodificar y pulsar en la flecha
- Definir nuevo nombre y pulsar cambiar
- Pinchar en valores antiguos y nuevos
- Para cada categoría que se construya seleccionar la opción correspondiente en valor antiguo, escribir el valor nuevo y pulsar Añadir
- Pulsar Continuar y Aceptar

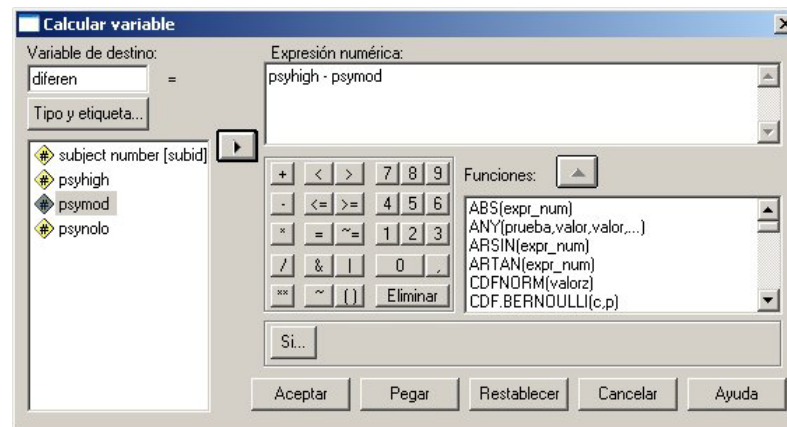


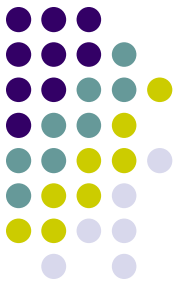


# Calcular variables

Transformar → Calcular

- Computa variables nuevas a partir de otras existentes
- Dar el nombre de la nueva variable, las variables originales y la operación a realizar entre ellas





# Importar datos

- Desde un fichero de texto
- Desde una hoja de cálculo
- Desde una base de datos

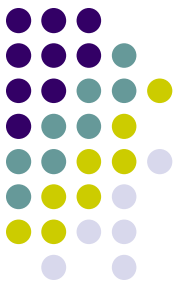


# Importar desde un fichero de texto

Archivo-→Abrir base de datos-→Nuevo

Archivo-→Abrir-→Datos

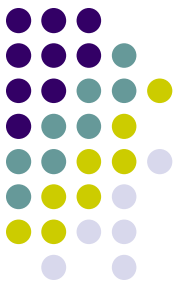
- Indicar si los datos están delimitados (comas, tabuladores, o están alineadas en columnas de ancho fijo)
- Especificar si los nombres de las variables están incluidos en la 1ª línea
- Definir en qué nº de línea comienza el primer caso
- Indicar el carácter delimitador de variables y texto



# Importar desde Excel

Archivo-→Abrir-→Datos

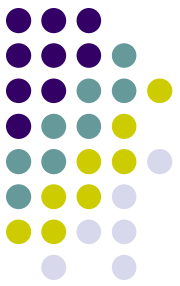
- En la ventana desplegable de *tipo* indicar Excel (\*.xls)
- Seleccionar la hoja de la cual queremos importar los datos
- Tomar el rango de celdas que aparece por defecto
- Si los nombres de las variables están escritos en la 1ª fila, marcar la casilla correspondiente



# Importar desde Access

Archivo->Abrir base de datos-> Nueva consulta

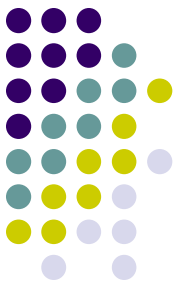
- Seleccionar la aplicación desde la que se quieren importar los datos (Access, Excel,...)
- Buscar la base de datos a importar
- Pinchar en la tabla o consulta que se quiere trasladar a SPSS
- En las demás opciones dar a la tecla *siguiente*



# 2. Estadística descriptiva

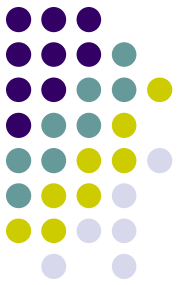
- 2.1 Presentación de datos
  - Tablas de frecuencias: frecuencias y %
  - Representaciones gráficas: barras, sectores, líneas...
- 2.2 Síntesis de datos
  - Medidas de posición
    - Centrales: Media, mediana, moda
    - No centrales: Percentiles
  - Medidas de dispersión
    - Rango (Máximo y mínimo)
    - Desviación típica o estándar





# Estadística descriptiva

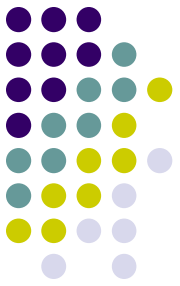
- La información que proporcionan los datos debe ser sintetizada, de forma que sea comprensible y de fácil manejo.
- **Objetivo:**  
Dar una visión global de cómo están distribuidas las variables sin necesidad de manejar todo el conjunto de datos
  - Tablas de frecuencias
  - Representaciones gráficas



# Cualitativa: frecuencias

Valor	Frecuencia	%
Asintomático	45	80.4
Angina	6	10.7
Insuf cardíaca	3	5.4
Shock	2	3.6
<b>Total</b>	<b>56</b>	<b>100.0</b>

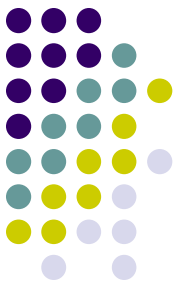
# Cualitativa: frecuencias (II)



## Proporción

Resulta de dividir el  $n^{\circ}$  de observaciones recogidas del estudio con una determinada característica entre el  $n^{\circ}$  total de observaciones; una proporción expresa una relación entre una parte y un todo (el numerador está incluido en el denominador). Sus valores van de 0 a 1 y no tiene unidades.

Para una más fácil interpretación y manejo, su valor suele multiplicarse por 100, y así queda expresado en forma de porcentaje.



# Cualitativa: frecuencias (III)

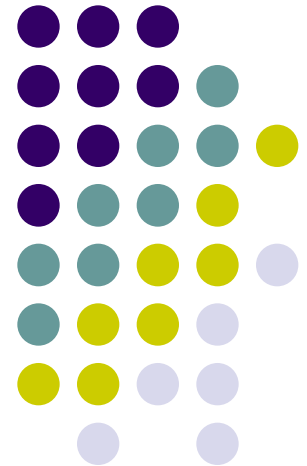
Analizar-→Estadísticos descriptivos-→Frecuencias

Intervalos	Frec.	%	Frec. acumulada	% acumulado
$\leq 138$	5	20.83	5	20.83
(138, 140]	7	29.17	12	50.00
(140, 142]	5	20.83	17	70.83
(142, 144]	3	12.50	20	83.33
$> 144$	4	16.67	24	100
<b>Total</b>	24	100	--	--

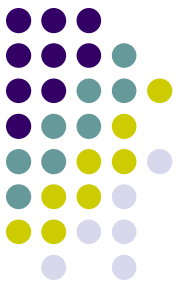
# Gráficos para variables cualitativas

---

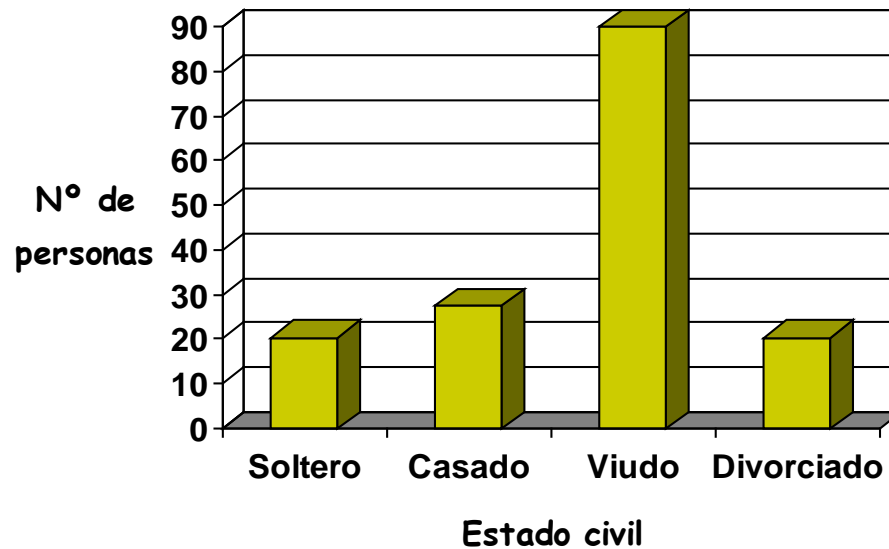
Diagramas de barras  
Diagramas de sectores  
Diagrama de Líneas



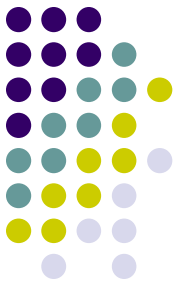
# Diagramas de barras



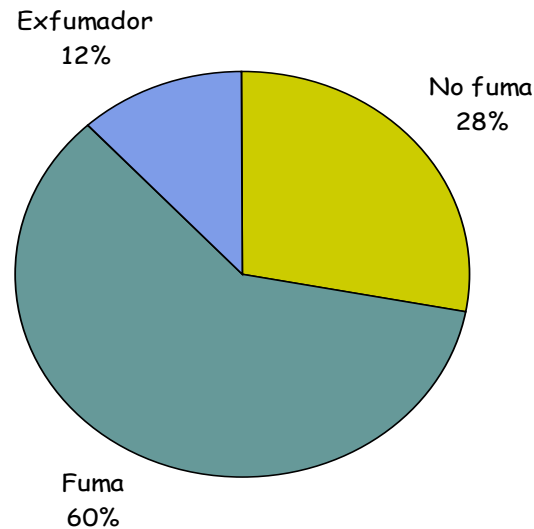
Representamos en el eje de ordenadas las modalidades y en abscisas las frecuencias absolutas o bien, las frecuencias relativas.

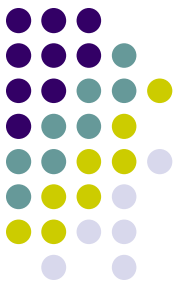


# Diagramas de sectores



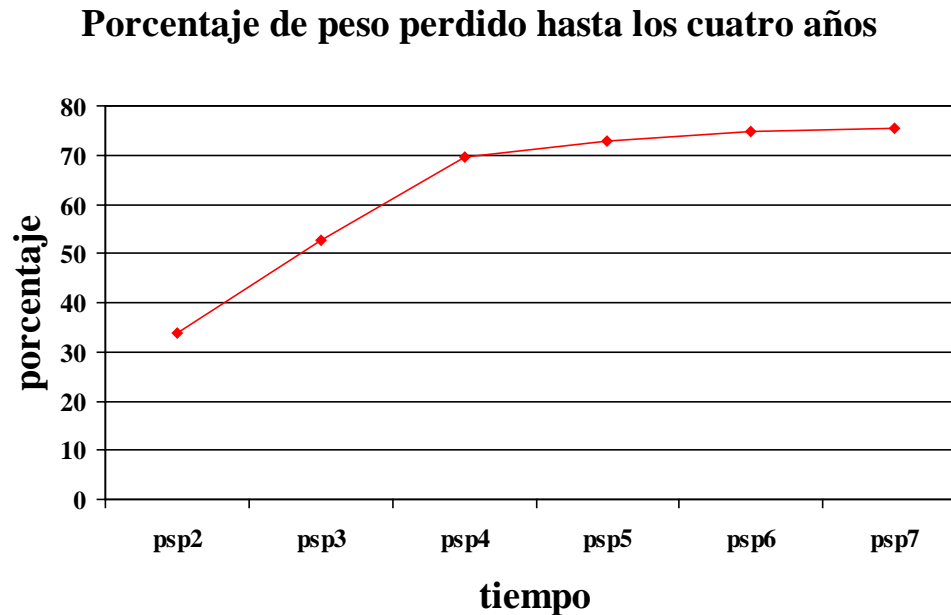
Se divide un círculo en tantas porciones como clases existan, de modo que a cada clase le corresponde un arco de círculo proporcional a su frecuencia absoluta o relativa.



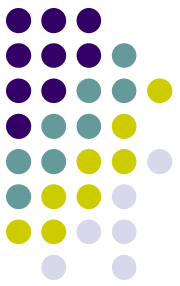


# Diagramas de líneas

Se representa en el *eje x* el tiempo de medición (meses, semestres, años,...) y en el *eje y* la frecuencia absoluta o el porcentaje





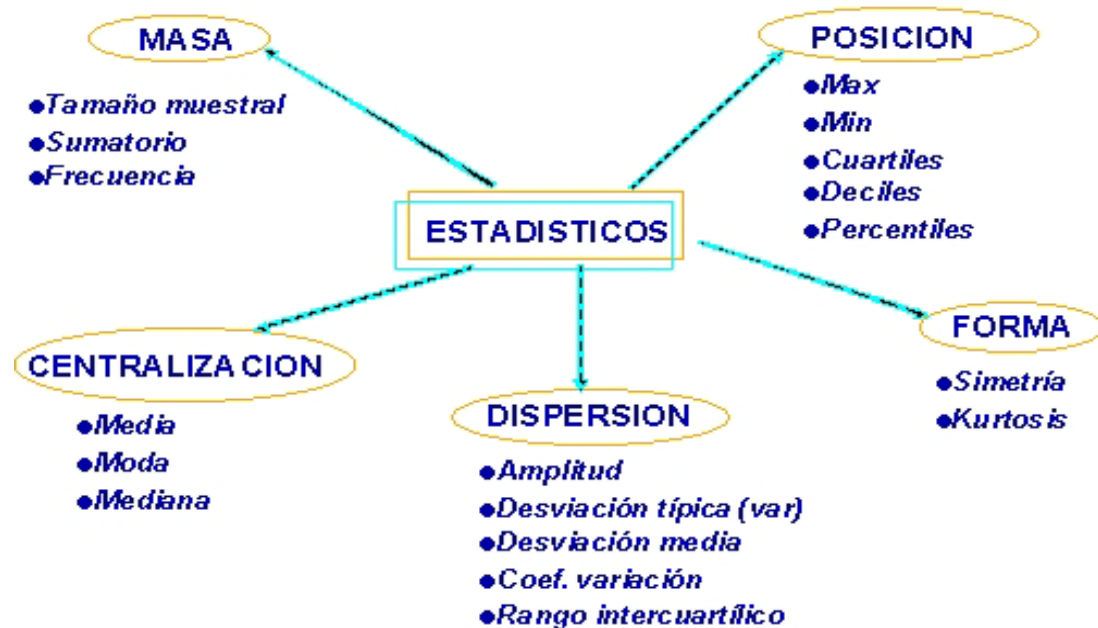


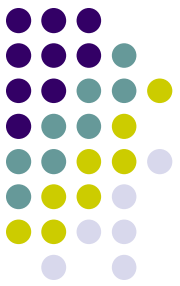
# Cuantitativa: síntesis de datos

Analizar -> Estadísticos descriptivos -> Frecuencias/Descriptivos/Explorar

- **Objetivo:**  
Reducir la información de los datos en torno a ciertos valores para tenerlos identificados con unas pocas medidas

## CLASIFICACION DE LOS ESTADISTICOS UNIVARIANTES





# Cuantitativa: síntesis de datos

**Variable:** edad de los pacientes (N, tamaño de la muestra)

## Medidas centrales de localización

- **Media:** valor promedio de la muestra
- **Mediana:** valor que se encuentra en la mitad
- **Moda:** valor más observado
- **Mínimo:** menor valor de la muestra
- **Máximo:** mayor valor de la muestra

## Medidas de dispersión

- **Amplitud:** diferencia entre máximo y mínimo
- **Desviación Estándar:** diferencia media de los datos respecto de la media.

## Medidas de localización

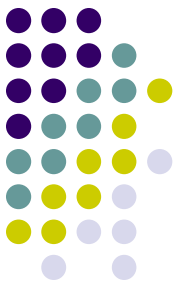
- **Cuartiles:** dividen la muestra en cuatro grupos de igual frecuencia
- **Centiles:** dividen la muestra en 100 grupos de igual frecuencia

## Medidas de Forma

- **Sesgo:** asimetría
- **Curtosis:** apuntamiento

EDAD
30
35
25
28
32
27
38
42
63
36
40
38
42
48
52
46
36
49
43
52
48
36

# Medidas de tendencia central I

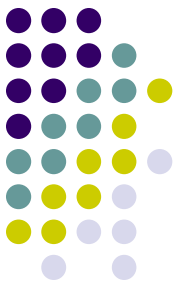


- **Media:**

Si tenemos  $X_1, X_2, \dots, X_n$  datos, se llama media muestral de los mismos a su media aritmética.

Se calcula sumando los valores de las observaciones individuales realizadas, y ese valor se divide por el número de observaciones:

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{\sum_{i=1}^n X_i}{n} = \frac{\sum X_i}{n}$$



# Medidas tendencia central II

- **Mediana:**

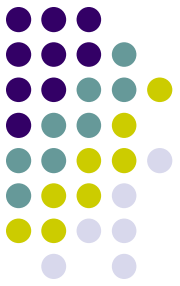
Expresa el valor que, una vez ordenados los valores de la variable, deja igual  $n^{\circ}$  de individuos u observaciones por encima que por debajo. Para calcularla debemos ordenar los valores obtenidos de las observaciones de mayor a menor (o viceversa) e identificar el valor central.

$$\hat{X} = \begin{cases} X_{\frac{n+1}{2}} & \text{si } n \text{ impar} \\ \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2} & \text{si } n \text{ par} \end{cases}$$

es fácilmente localizable en el valor central

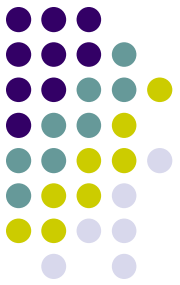
hay dos valores centrales y para calcular la mediana debemos hallar la media de esos dos valores.

# Medidas tendencia central III



- **Moda:**

El valor que más se repite (puede no existir y si existe puede no ser única).



# Ejemplo: días de estancia

1 1 2 2 5 4 1 2 3 5 4 1 2 3 1 4 2 1 4 1

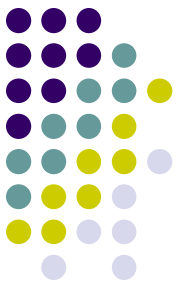
✓ Media:  $(1+1+2+\dots+4+1)/20=2.45$

✓ Mediana: 2



1 1 1 1 1 1 1 2 2 2 2 2 3 3 4 4 4 4 5 5

✓ Moda: 1

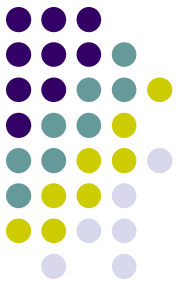


# Medidas de posición

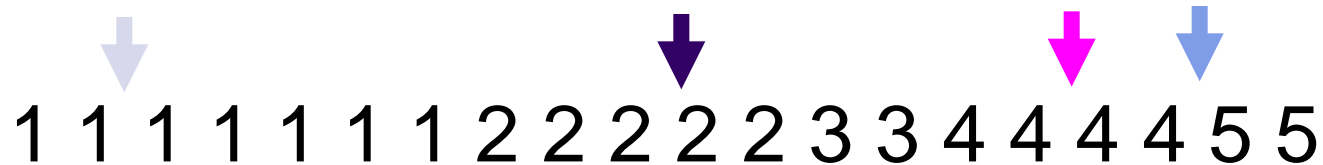
- Resumen en torno a un valor no central
- Los datos tienen que estar ordenados

## Medidas:

- Percentiles: 1,2,...,99, 100. Valor que deja un porcentaje determinado por debajo y el resto (hasta 100) por encima

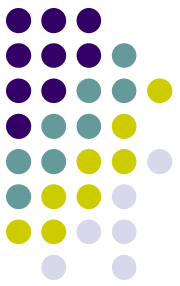


# Ejemplo



- Percentil 10: **1** (10% de datos por debajo, 90% por encima)
- Percentil 50: **2** (Mediana)
- Percentil 75: **4**
- Percentil 90: **4.5**





# Medidas de dispersión (Media)

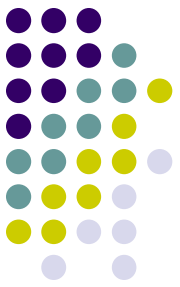
- **Desviación típica o estándar:**

Expresa la dispersión de los valores de la variable en torno a la media y se calcula obteniendo la raíz cuadrada de la **varianza ( $S^2$ )**.

La varianza se valora en unidades cuadradas de difícil interpretación. Si obtenemos su raíz cuadrada sus unidades son las mismas que la de la variable.

$$s = +\sqrt{S^2}$$
$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

# Medidas de dispersión (mediana)



- **Percentiles:**

Son cifras que indican los porcentajes de una distribución que son iguales o menores que ese valor. El percentil 50 coincide con la mediana.

Se denominan **cuartiles** a los percentiles 25, 50 y 75.

Ejemplo: si un pediatra dice que un niño mide 130 cm y está en el percentil 50, quiere decir que la mitad de los niños de esa población están por debajo de ese valor y ese niño está justo en el valor mediano de esa distribución.

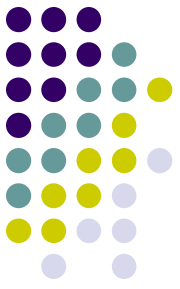
El **rango intercuartílico** viene definido por la diferencia entre los percentiles 25 y 75 e incluye el 50% de las observaciones.

# Uso de las medidas de tendencia central y dispersión



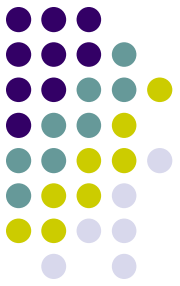
- Si la media y la mediana son iguales, la distribución es simétrica.
- Si la media es mayor que la mediana, la distribución está sesgada a la derecha.
- Si la media es menor que la mediana, la distribución está sesgada a la izquierda.
- La media se influye mucho por el peso de los valores extremos y la mediana no. Por ello conviene usar la media en las distribuciones simétricas y la mediana en las asimétricas.

# Uso de las medidas de tendencia central y dispersión

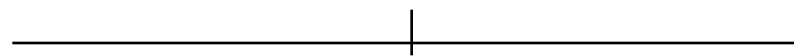
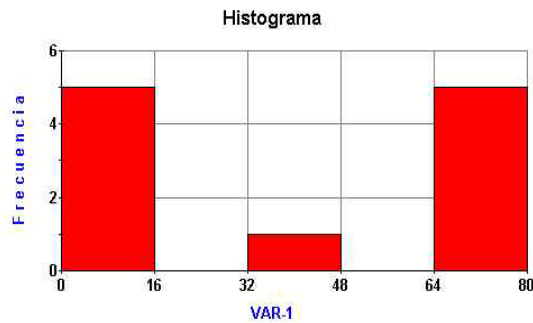


- La mediana se puede utilizar para variables cualitativas ordinales.
- La desviación estándar se usa cuando se calcula la media (distribuciones simétricas).
- Los percentiles suelen utilizarse cuando se utiliza la mediana.
- El rango (máximo-mínimo) se emplea cuando se quiere enfatizar los valores extremos.

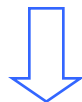
# Uso de las medidas de tendencia central y dispersión



Resumir la Información mediante la media sin indicar la desviación estándar es un importante error estadístico

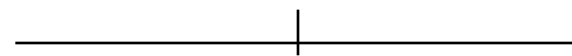
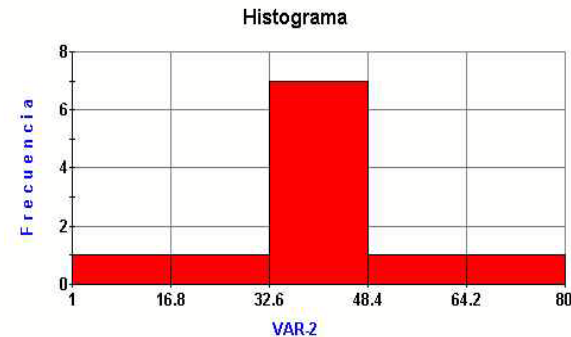


Datos: 0, 0, 0, 0, 0, 40, 80, 80, 80, 80



Media: 40

D.E.: 40



Datos: 0, 20, 40x7, 60, 80



Media: 40

D.E.: 20

# Gráficos para variables cuantitativas

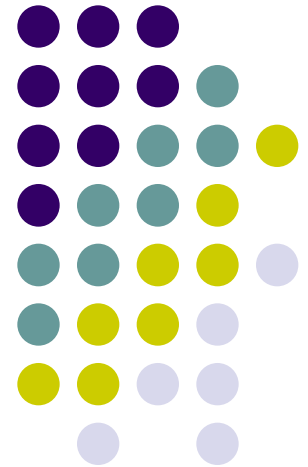
---

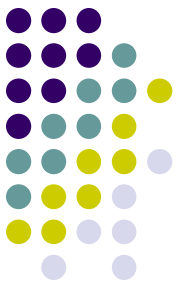
Histograma

Barras de error

Diagrama de cajas y bigotes

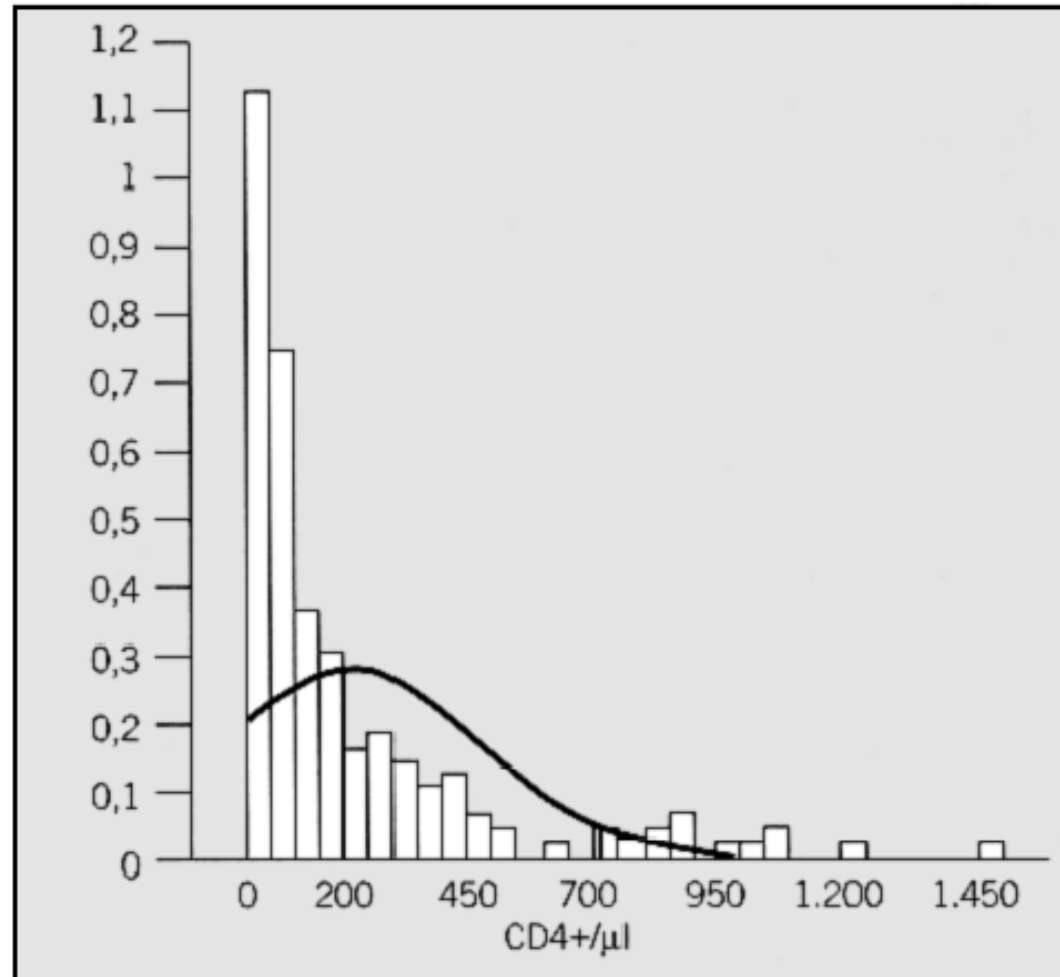
Diagrama de Tallo y hojas



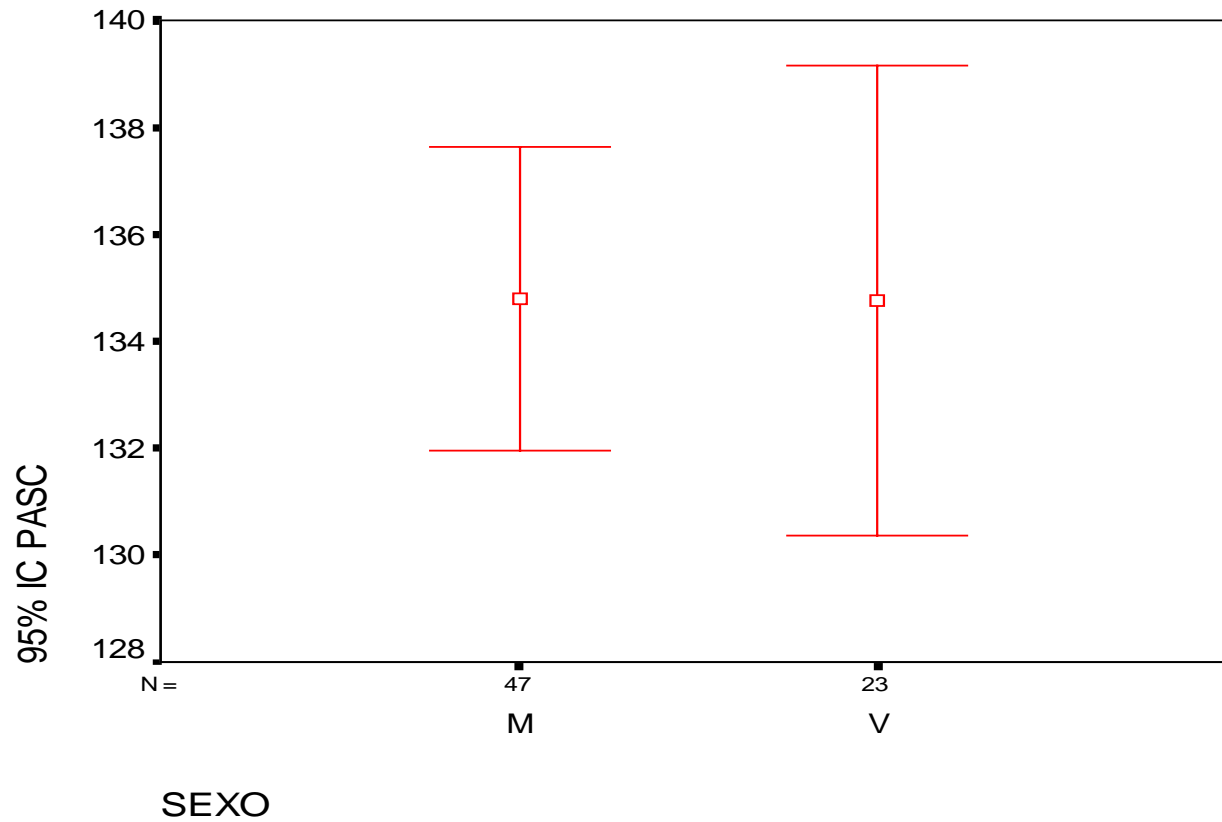
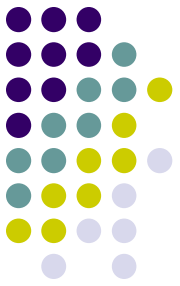


# Histograma

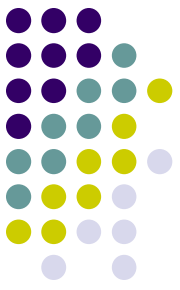
*Fig. 1. Histograma de la distribución de 181 pacientes con tuberculosis e infección por VIH en función del recuento de linfocitos CD4. La curva representa la distribución normal estimada al utilizar la media y desviación estándar de la muestra. El área delimitada por cada gráfico corresponde al número de pacientes.*



# Barras de error

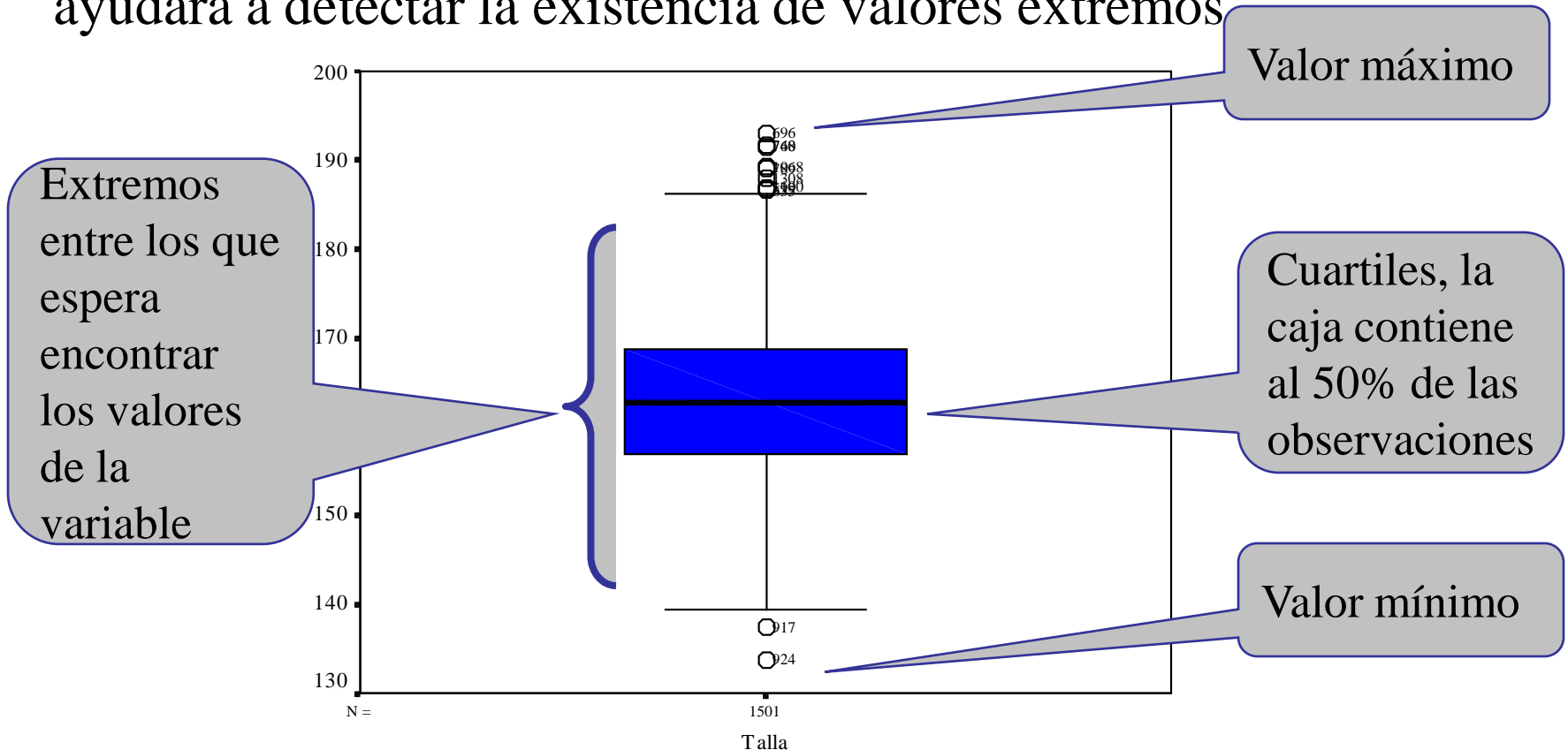


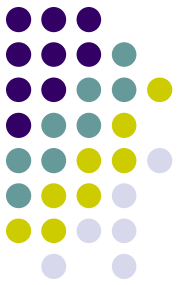




# Diagrama de caja y bigotes

El **diagrama de caja** es una representación gráfica que nos ayudará a detectar la existencia de valores extremos

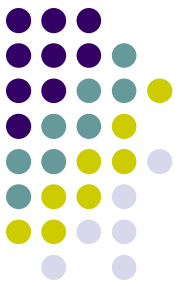




# Diagrama de tallo y hojas

## Edad de la madre

Frequency	Stem &	Leaf
,00	1 .	
6,00	1 .	444555
19,00	1 .	666666677777777777
26,00	1 .	88888888889999999999999999
30,00	2 .	000000000000000000111111111111
26,00	2 .	22222222222223333333333333
28,00	2 .	4444444444444455555555555555
11,00	2 .	66666666777
16,00	2 .	8888888889999999
12,00	3 .	00000011111
9,00	3 .	222222333
3,00	3 .	455
2,00	3 .	66
1,00	Extremes	(>=45)

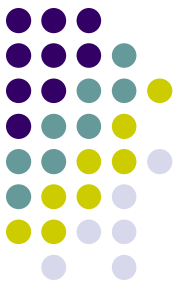


## 2 ó más variables cualitativas

Analizar-→Estadísticos descriptivos-→Tablas de contingencia

Consideramos cuatro tratamientos 1, 2, 3 y 4 para la curación de una enfermedad aplicados a sujetos distintos. Los resultados se midieron en tres categorías: Peor, Igual o Mejor, obteniendo los siguientes resultados:

Tto	Peor	Igual	Mejor	Total
1	7	28	115	150
2	15	20	85	120
3	10	30	90	130
4	5	40	115	160
Total	37	118	405	560

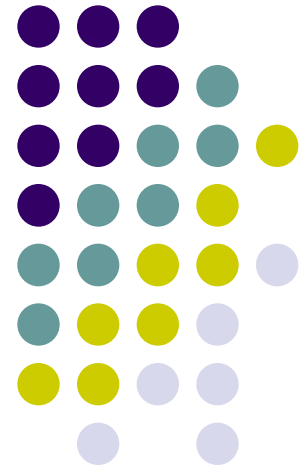


# Tablas de contingencia

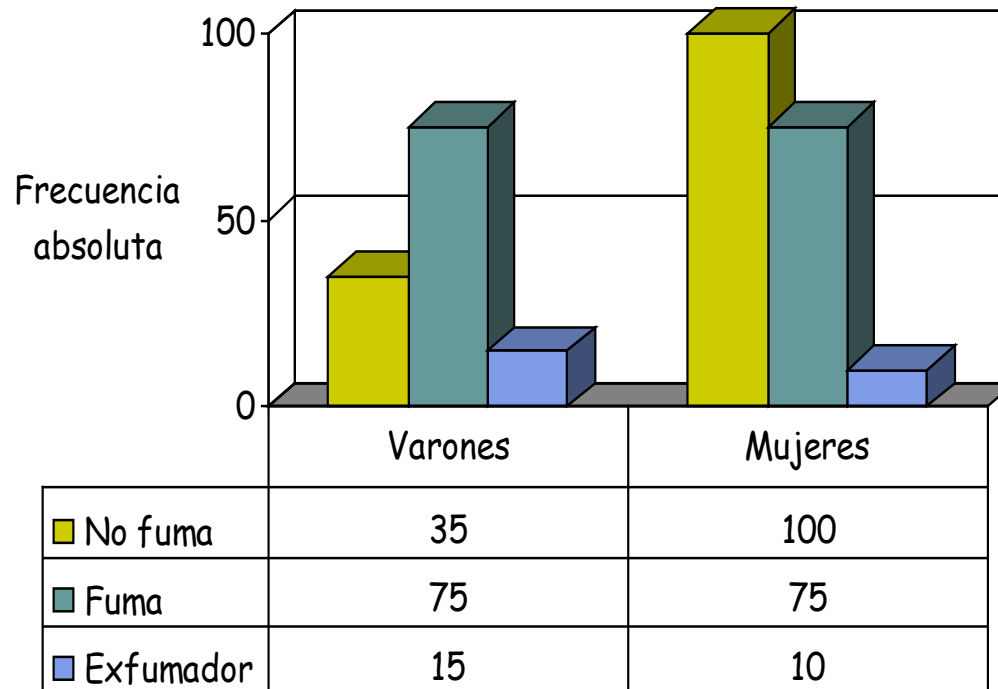
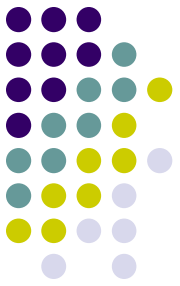
- En las tablas de contingencia se pueden calcular porcentajes de tres maneras distintas y, dependiendo de el experimento en cuestión, elegiremos una u otra:
  - por filas (% de sujetos que con cada tratamiento están peor, igual o mejor)
  - por columnas (% de personas que dentro de cada categoría de respuesta, se les aplicó uno u otro tratamiento)
  - referido al total (% de individuos que se encuentran en cada celda con respecto a todos los sujetos)

# Gráficos variables cualitativa\*cualitativa

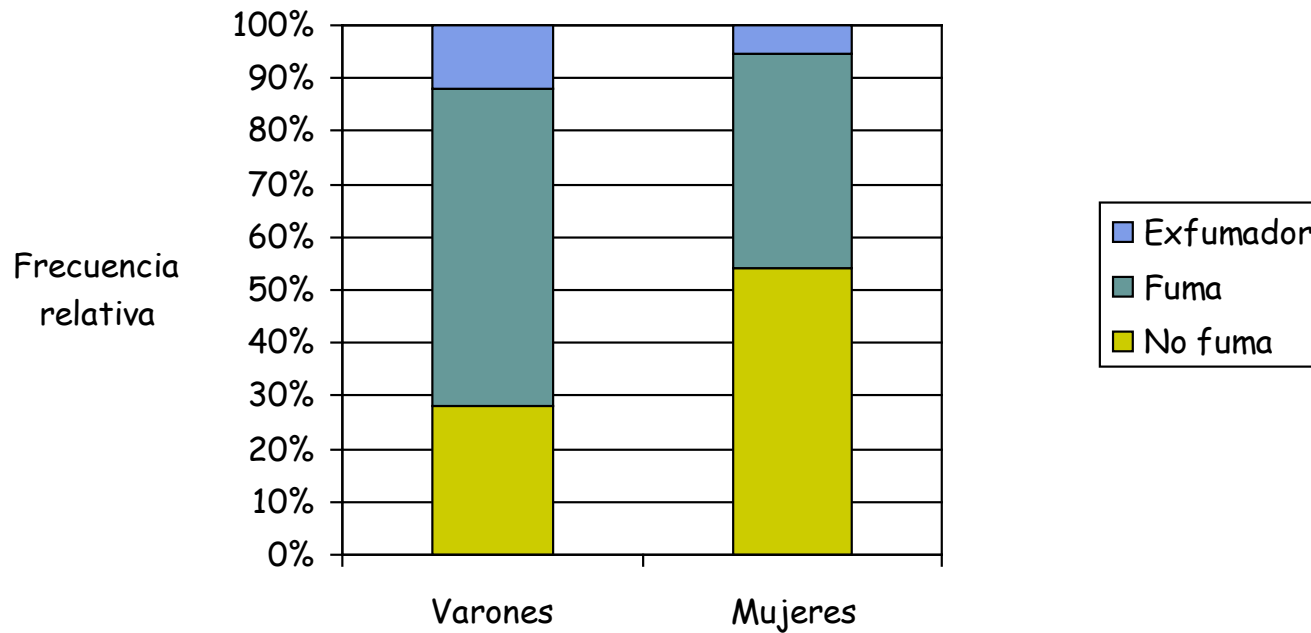
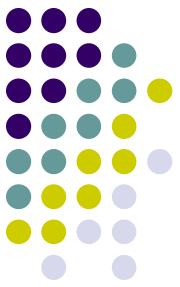
Diagramas de barras agrupado

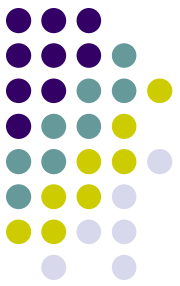


# Diagrama de barras agrupado



# Diagramas de barras agrupado





## 2 ó más variables: “cuali y cuanti”

Analizar→Comparar medias→Medias

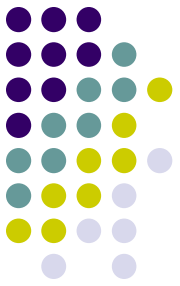
Salarios de los trabajadores de una empresa de EEUU por sexo

### Informe

Salario actual

Sexo	Media	Desv. típ.	Mínimo	Máximo
Hombre	\$41,441.78	\$19,499.214	\$19,650	\$135,000
Mujer	\$26,031.92	\$7,558.021	\$15,750	\$58,125
Total	\$34,419.57	\$17,075.661	\$15,750	\$135,000





# 2 ó más variables: “cuali y cuanti”

Analizar-→Informes-→Informes de estadísticos en filas

	Salario
Sexo	actual

Hombre

Media \$41,442

Mínimo \$19,650

Máximo \$135,000

Desv.Típ \$19,499

Mujer

Media \$26,032

Mínimo \$15,750

Máximo \$58,125

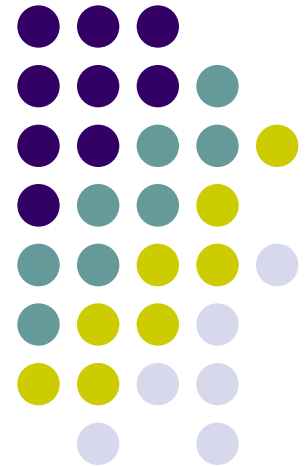
Desv.Típ \$7,558

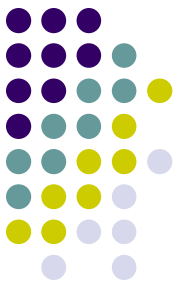
Salarios de los trabajadores de una empresa de EEUU por sexo

# Gráficos variables cualitativa\*cuantitativa

---

Diagrama de líneas





# Diagrama de líneas

**Medias de niveles de hierro para las distintas categorías de cabello5**

