

Estadística Descriptiva I

Metodología de la Investigación en Enfermería

Cátedra de Bioestadística
Universidad de Extremadura

31 de enero de 2012

¿De qué trata?

Descripción conjunto concreto de datos (sin generalizar)

- Clasificación \rightsquigarrow **Tablas de frecuencia**
- Representación \rightsquigarrow **Gráficos**
- Resumen \rightsquigarrow **Valores típicos**

Tipos de variables

Según el programa SPSS

- Cualitativas (factores)
 - Nominales: Grupo sanguíneo, sexo
 - Ordinales: Grado enfermedad, encuesta satisfacción
- Cuantitativas o de escala : Temperatura, estatura, glucemia, n^o hijos...

Gráficos

Tipos de gráficos

- Variables cualitativas \rightsquigarrow Diagrama sectores o de barras
- Variables cuantitativas:
 - Discretas \longleftrightarrow Diagrama barras
 - Continuas \longleftrightarrow Histograma

*Nota: En este capítulo se estudiará una única variable.

Variable cualitativa (Grupo sanguíneo)

- Tablas de frecuencias
- Gráficos (sectores, barras)
- Valores típicos ¿?

Tabla de frecuencias

Grupo	f_i	\hat{p}_i
O	2892	0,458
A	2625	0,416
B	570	0,090
AB	226	0,036
Total	6313	1

Diagrama de sectores

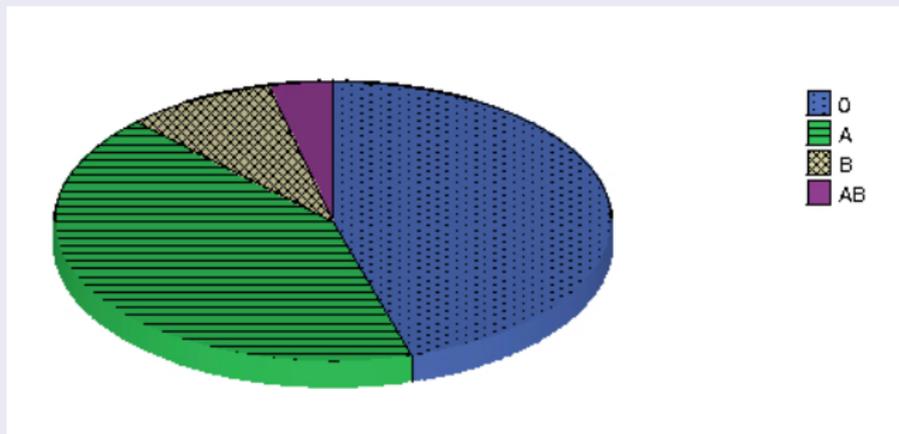


Figura: Grupo sanguíneo

Otra opción: diagrama de barras

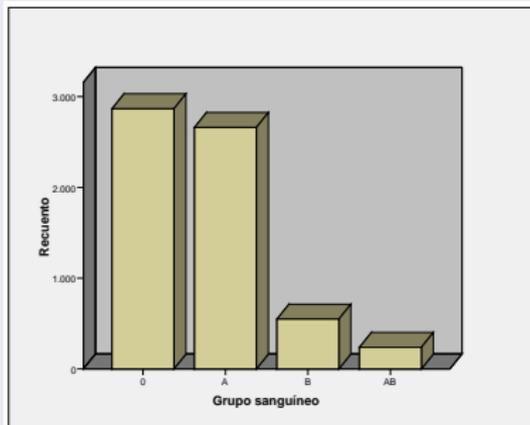


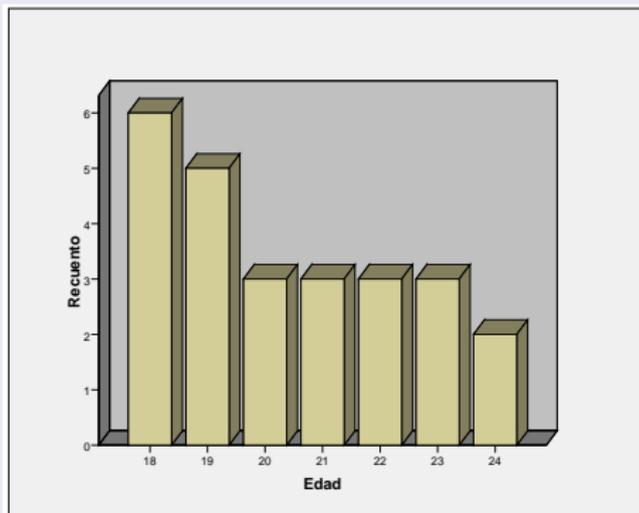
Figura: Grupo sanguíneo

Variable cuantitativa discreta

Edad alumnos titulación universitaria: tabla frecuencias

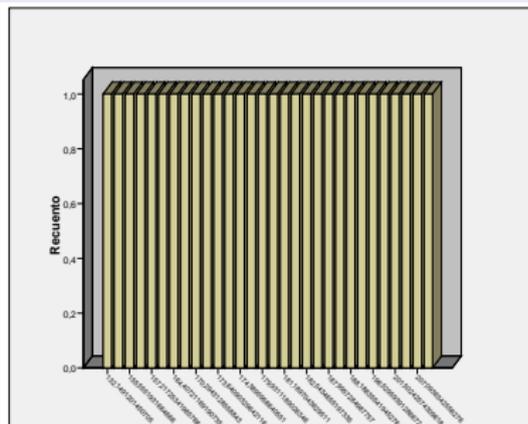
Edad	f_i	\hat{p}_i
18	6	0.24
19	5	0.20
20	3	0.12
21	3	0.12
22	3	0.12
23	3	0.12
24	2	0.08
Total	25	1

Gráfico: Diagrama de barras



Variable cuantitativa continua

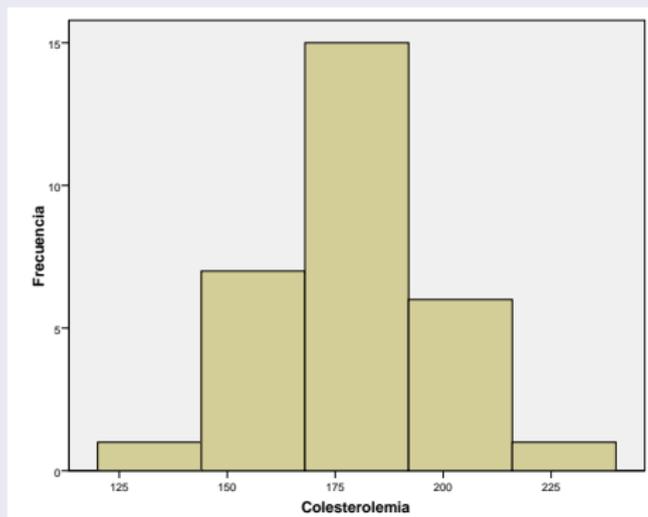
Colesterolemia n=30 Diagrama de barras



Agrupar datos por intervalos

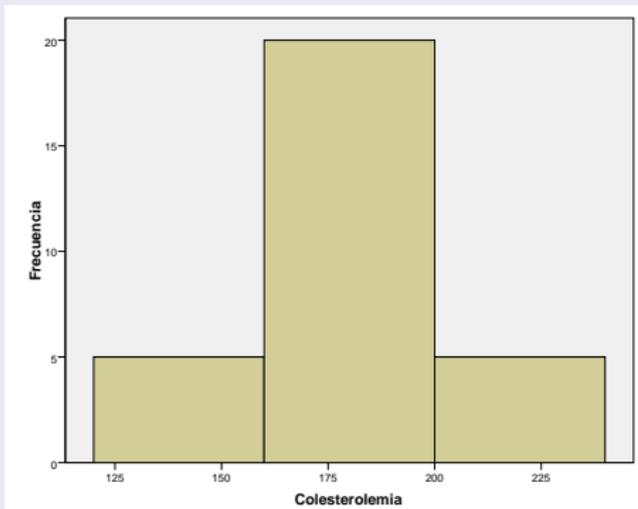
Histograma

Agrupamos por intervalos $k=5$



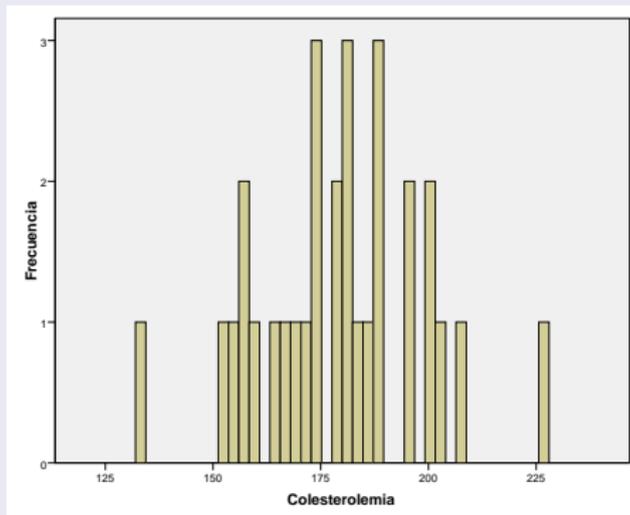
¿Cuántos intervalos?

Glucemia $k=3$



¿Cuántos intervalos?

Glucemia $k=50$



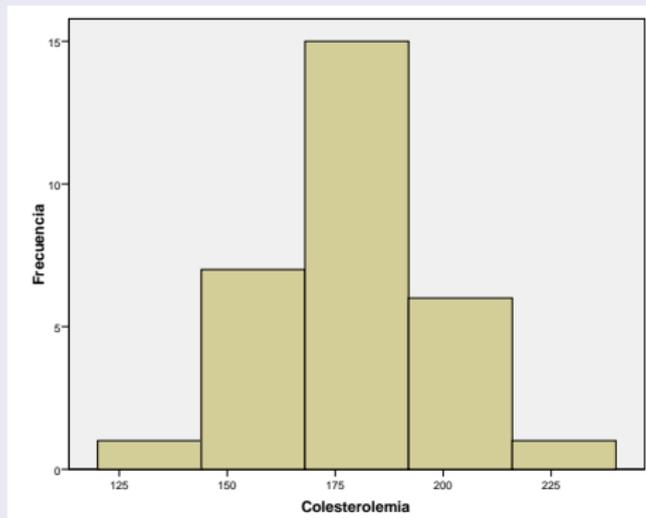
¿Cuántos intervalos? Ley de Sturges

número de datos → número de intervalos

Datos	Intervalos
16-31	5
32-64	6
65-125	7

Histograma

Glucemia $k=5-6$



Otros gráficos para variables continuas

Tallo-hoja

Lectura de cadmio

Frecuencia	Tallo-Hoja
1	2 . 0
6	3 . 007999
9	4 . 022445679
11	5 . 1122345678
7	6 . 0111226
1	7 . 0

Unidad: 0.01

Diagrama de caja o box-plot

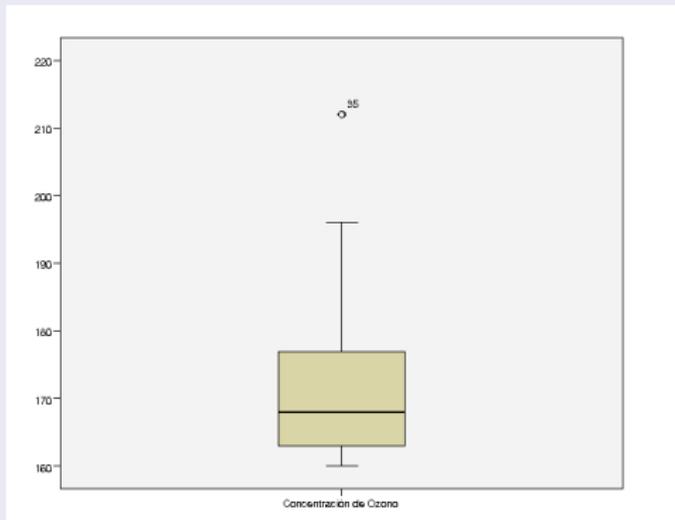


Figura: Concentración ozono

Valores típicos

Resumir numéricamente la información

- **Medidas de centralización**
- Medidas de posición
- **Medidas de dispersión**
- Medidas de forma

Medidas de centralización

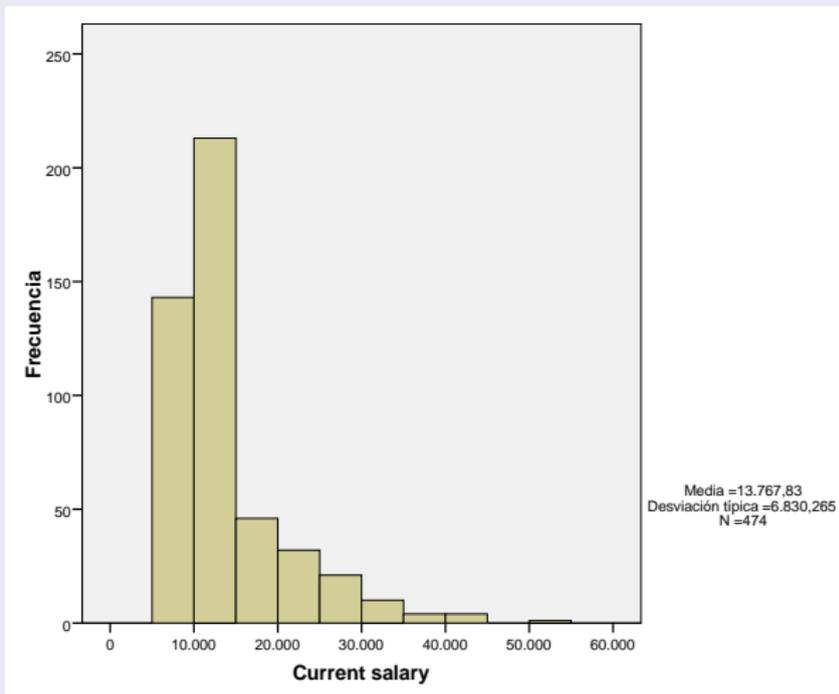
Centro de los datos, número más representativo

- Media aritmética (centro de gravedad):

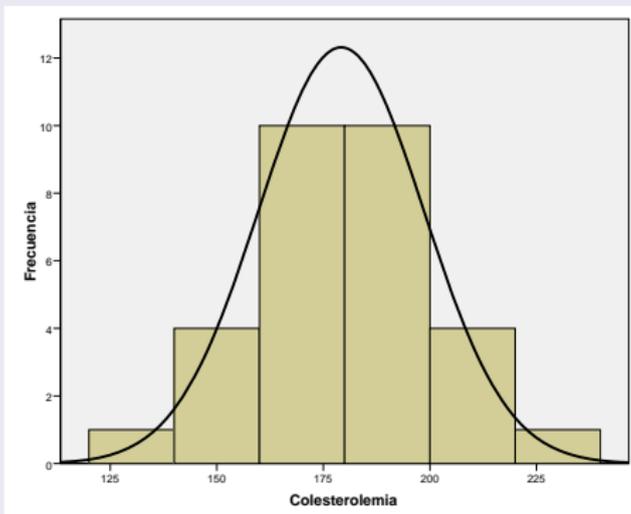
$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

- Mediana (orden): dato que queda en medio una vez ordenados de menor a mayor.

Otro caso de sesgo $n=474$



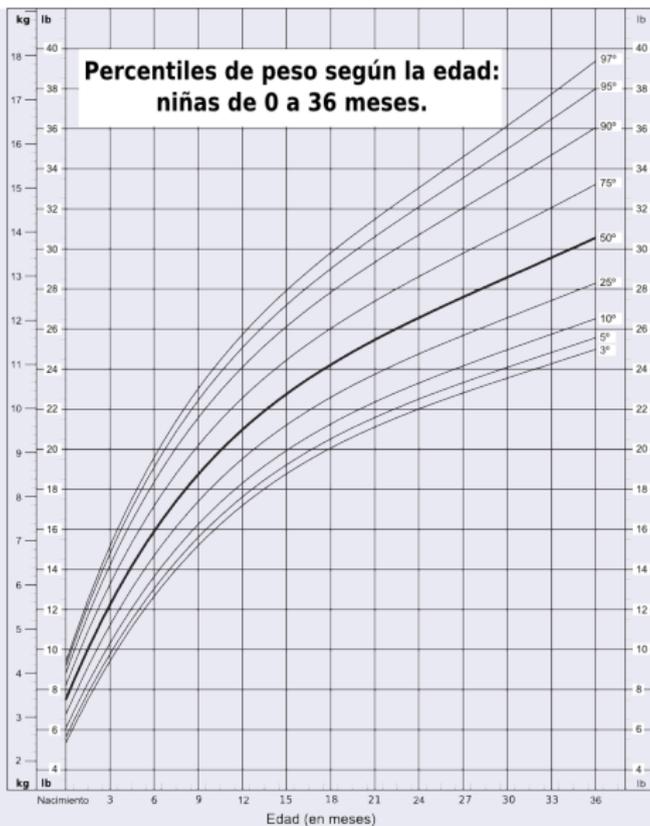
Distribución normal: media



Medidas de posición

Orden

- Mediana: 50 %
- Cuartiles: Q_1 , Q_2 y Q_3 : 25 %, 50 % y 75 %, resp.
- Deciles: D_1, D_2, \dots, D_9
- Percentiles [Ejemplo : $p_{50} = D_5 = Q_2 = \text{mediana}$]



Medidas de dispersión

Variabilidad de los datos

- En relación con la media: varianza \rightsquigarrow **desviación típica**

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$$

- En relación con el orden: **amplitud intercuartil**

$$R_I = Q_3 - Q_1$$

IMPORTANTE: Resumen de la información

Centralización + dispersión

[media + desviación típica]

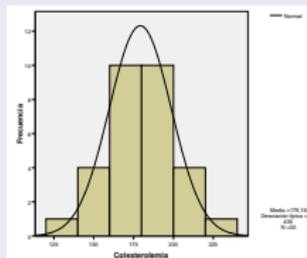
[mediana + amplitud intercuartil]

media-desviación típica

Resumen perfectamente la información si la distribución de los datos sigue un modelo normal

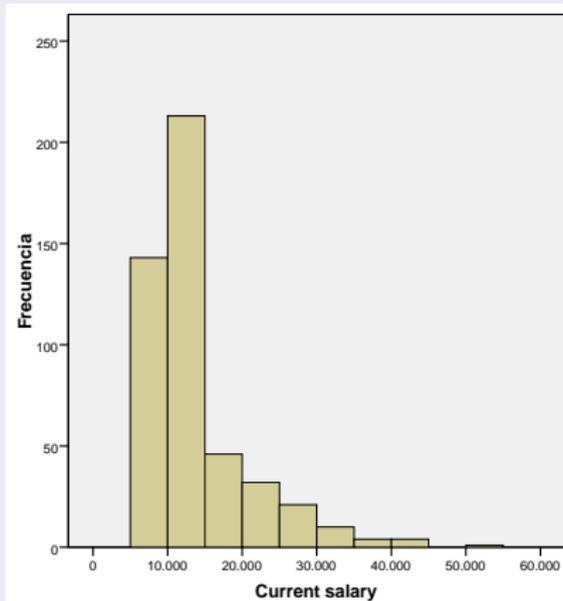
- $\bar{x} \pm s \rightsquigarrow 68\%$
- $\bar{x} \pm 2s \rightsquigarrow 95\%$
- $\bar{x} \pm 3s \rightsquigarrow 99\%$

Podemos construir la distribución partiendo exclusivamente de \bar{x} y s .



mediana-rango intercuartil

En aquellas situaciones en que la media sea poco representativa:
fuerte sesgo + valores extremos.



Medidas de forma

Cálculo complicado

- Coeficiente de asimetría
- Coeficiente de aplastamiento o curtosis.

Asimetría

