

Estadística Descriptiva

Bioestadística

Jesús Montanero Fernández

Cátedra de Bioestadística. UEx

3 de diciembre de 2010

¿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 SPSS

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

Problemas estadísticos: **relación**

Tipos de relaciones

- Cualitativa (factor) \leftrightarrow Cuantitativa (Comparación de grupos o tratamientos)
- Cuantitativa \leftrightarrow Cuantitativa (Regresión)
- Cualitativa (factor) \leftrightarrow Cuanlitativa (Tablas de contingencia)

Gráficos

Tipos de variables según gráfico

- Cualitativas \rightsquigarrow **Diagrama sectores**
- Cuantitativas discretas \rightsquigarrow **Diagrama barras**
- Cuantitativas continuas \rightsquigarrow **Histograma**

***Nota:** estudiaremos ahora una sola variable

Variable cualitativa: diagrama de sectores

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

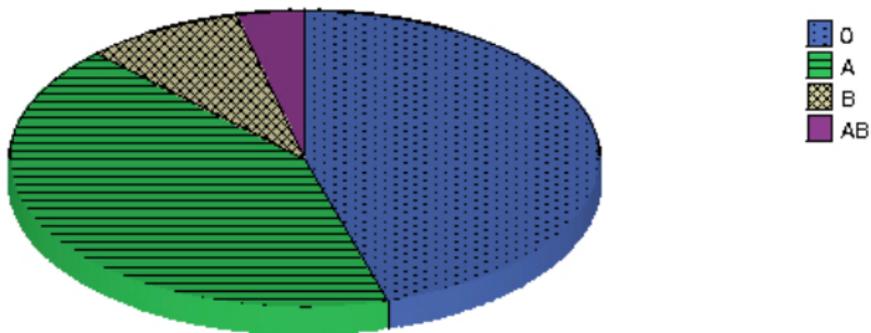
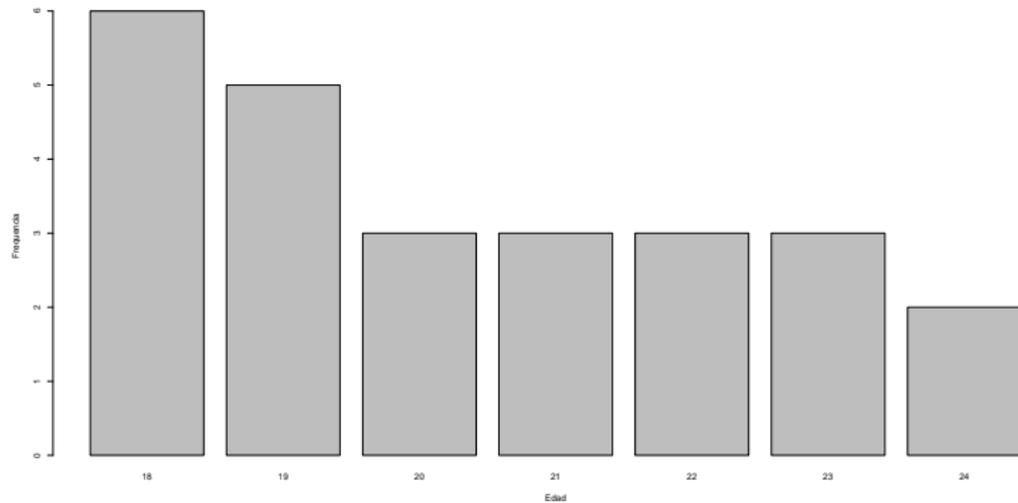


Figura: Grupo sanguíneo

Variable cuantitativa discreta: diagrama de barras

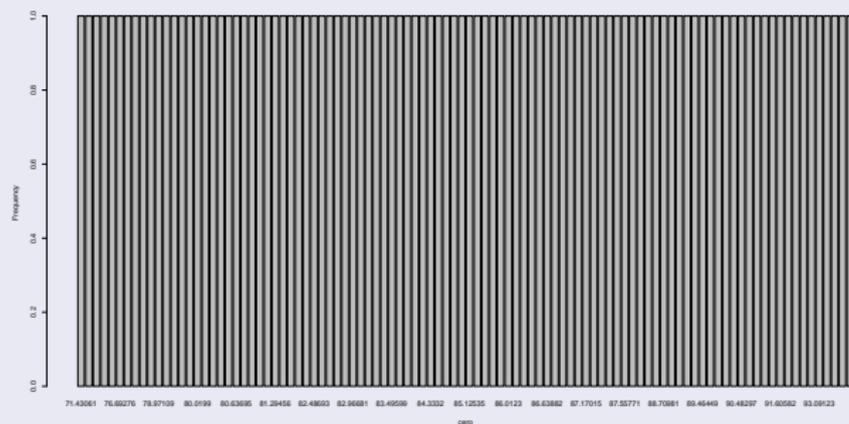
Edad alumnos titulación universitaria

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



Variable cuantitativa continua

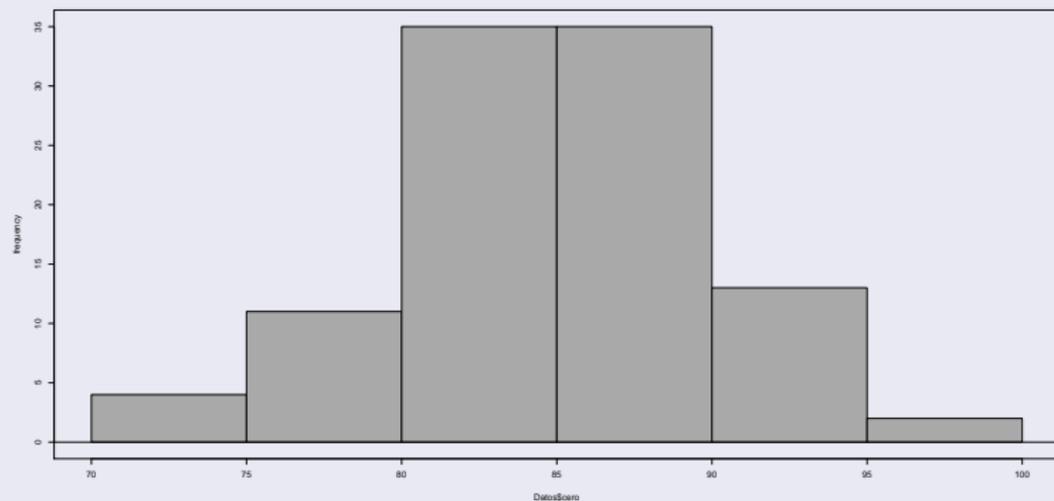
Glucemia n=100: Diagrama de barras



Agrupar datos por intervalos

Histograma

Glucemia k=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

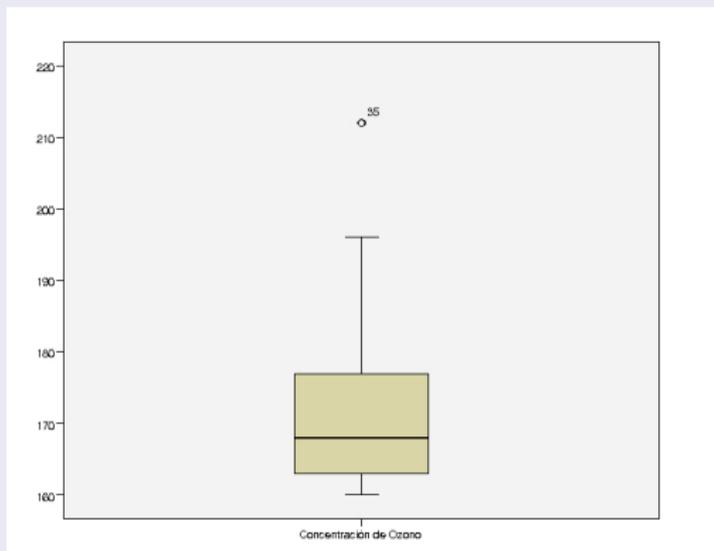


Figura: Concentración ozono

Valores típicos

Resumir numéricamente la información

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

Medidas de centralización

Centro de los datos, número más representativo

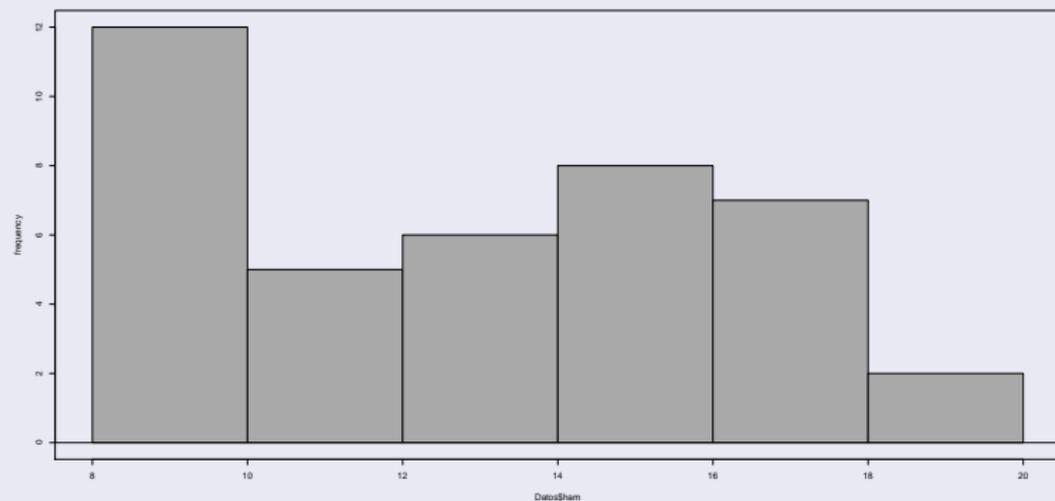
- Media aritmética (centro de gravedad):

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

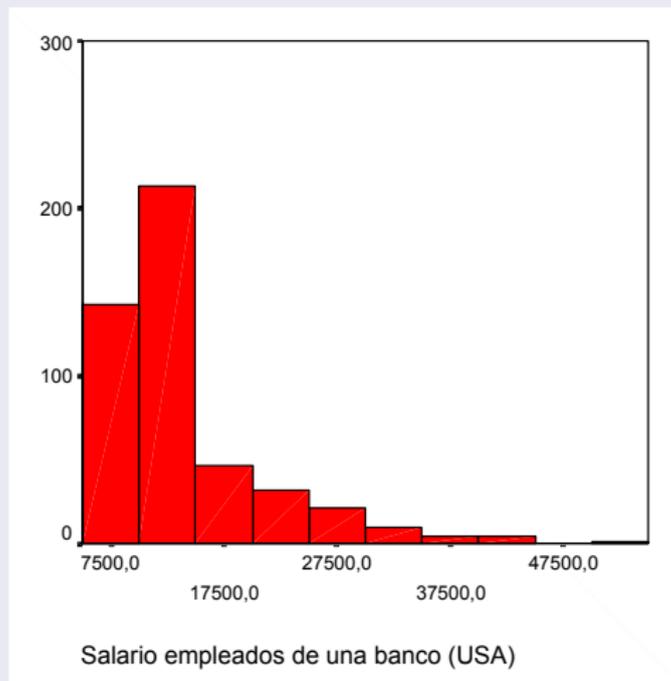
- Mediana (orden): dato que queda en medio una vez ordenados de menor a mayor.

La mediana es robusta: no se ve influenciada por la presencia de valores extremos

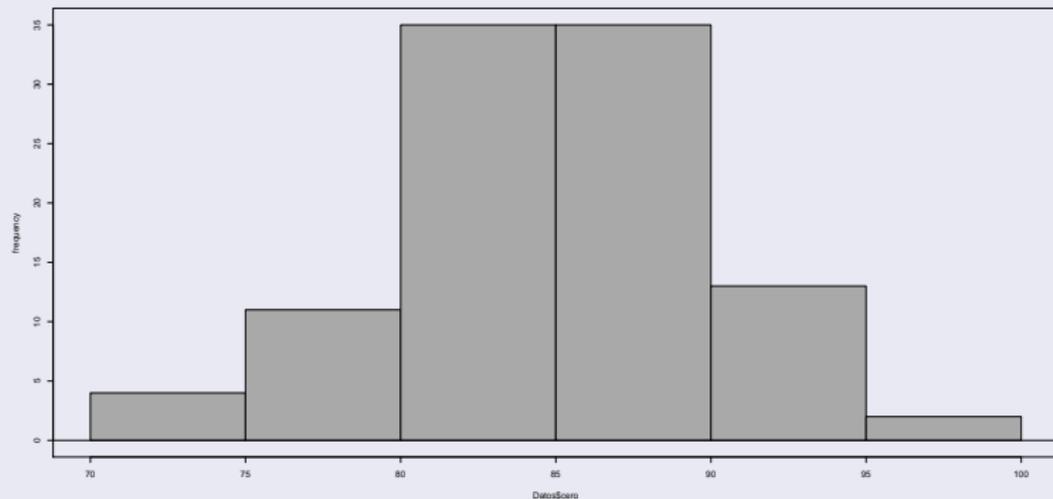
Ansiedad n=20



Otro caso de sesgo $n=350$



Distribución normal: media



Medidas de dispersión

Variabilidad de los datos

- Varianza \rightsquigarrow desviación típica
- Amplitud (o rango) intercuartil

Resumen de la información

Centralización-dispersión

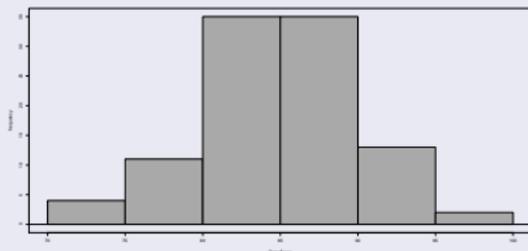
¿media-desv. típica o mediana-rango inter.?

media-desviación típica

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

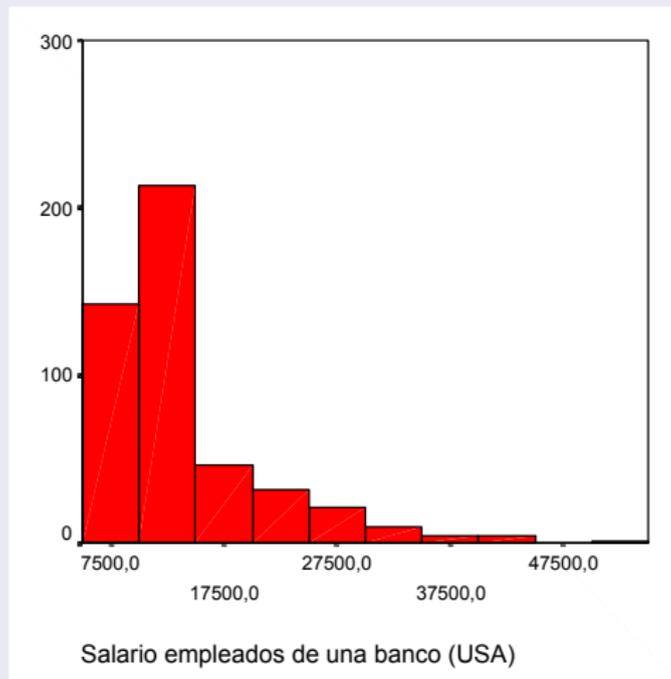
- $\bar{x} \pm s \rightsquigarrow 68\%$
- $\bar{x} \pm 2s \rightsquigarrow 95\%$
- $\bar{x} \pm 3s \rightsquigarrow 99\%$

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.



mediana-rango intercuartil

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

