

Bioestadística y uso de software científico



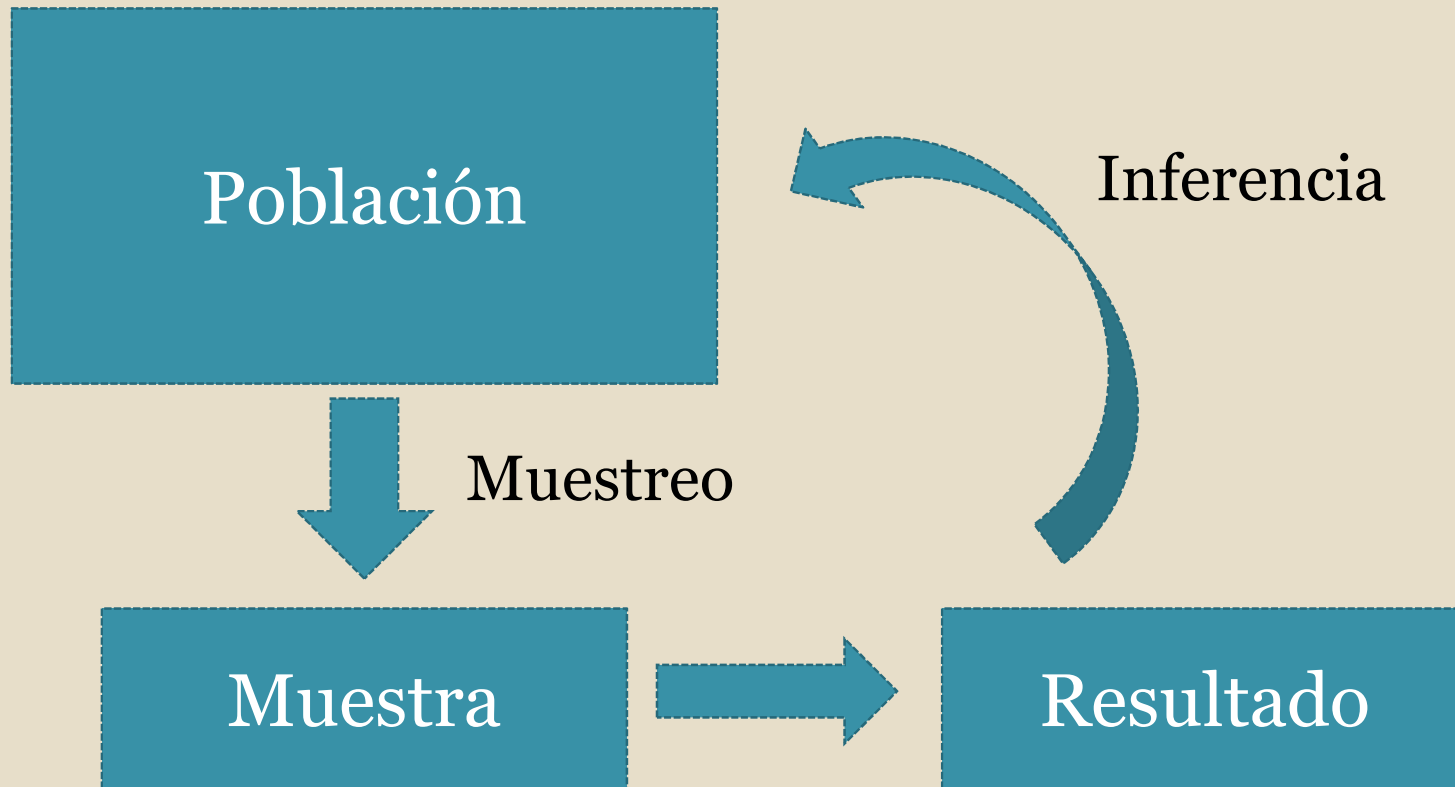
TEMA 1

INTRODUCCIÓN. ESTADÍSTICA DESCRIPTIVA

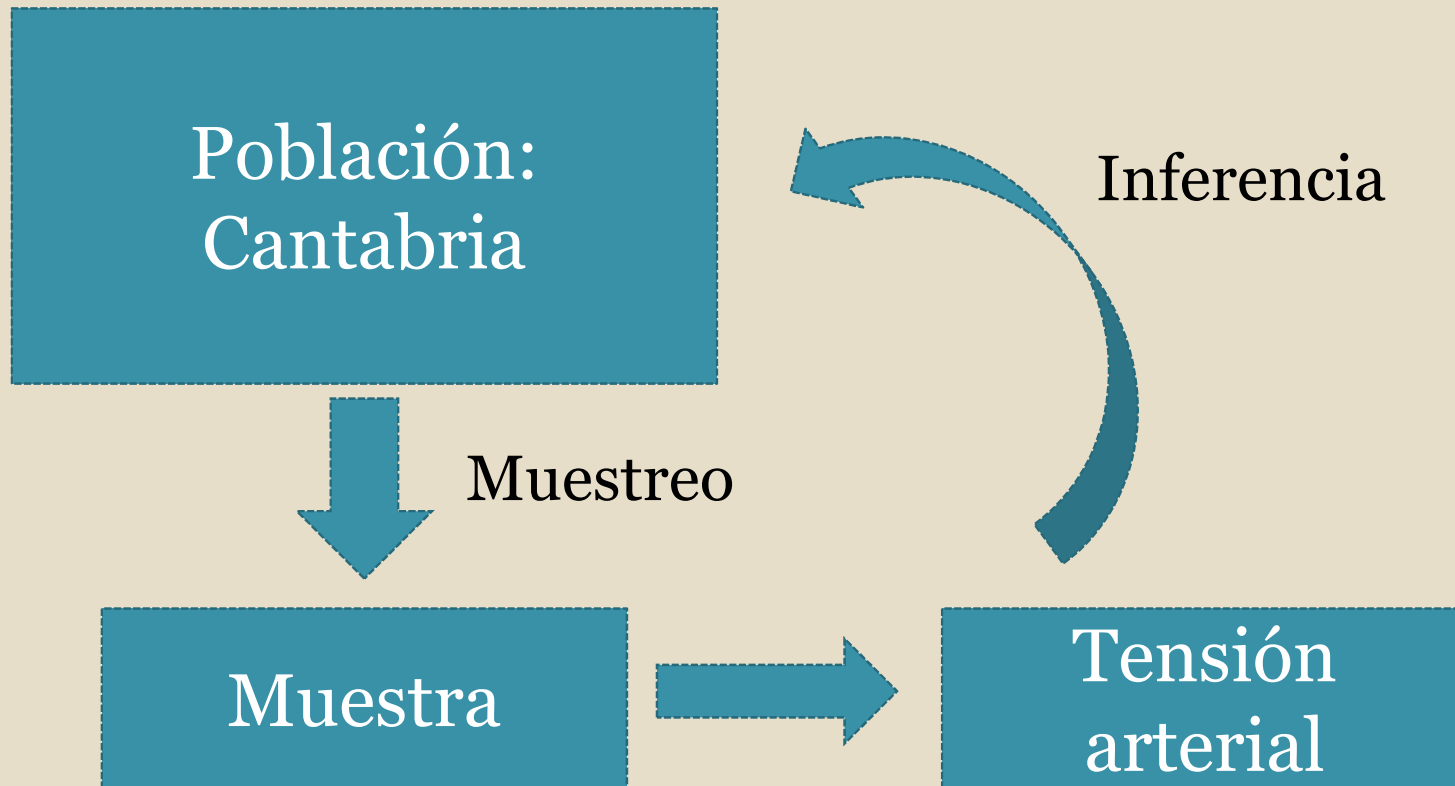


- Población y muestra
- Tipos de variable
- Representación gráfica
- Medidas que resumen una muestra o una población
 - Medidas de tendencia central
 - Medidas de dispersión
 - Medidas de posición
 - Medidas de forma

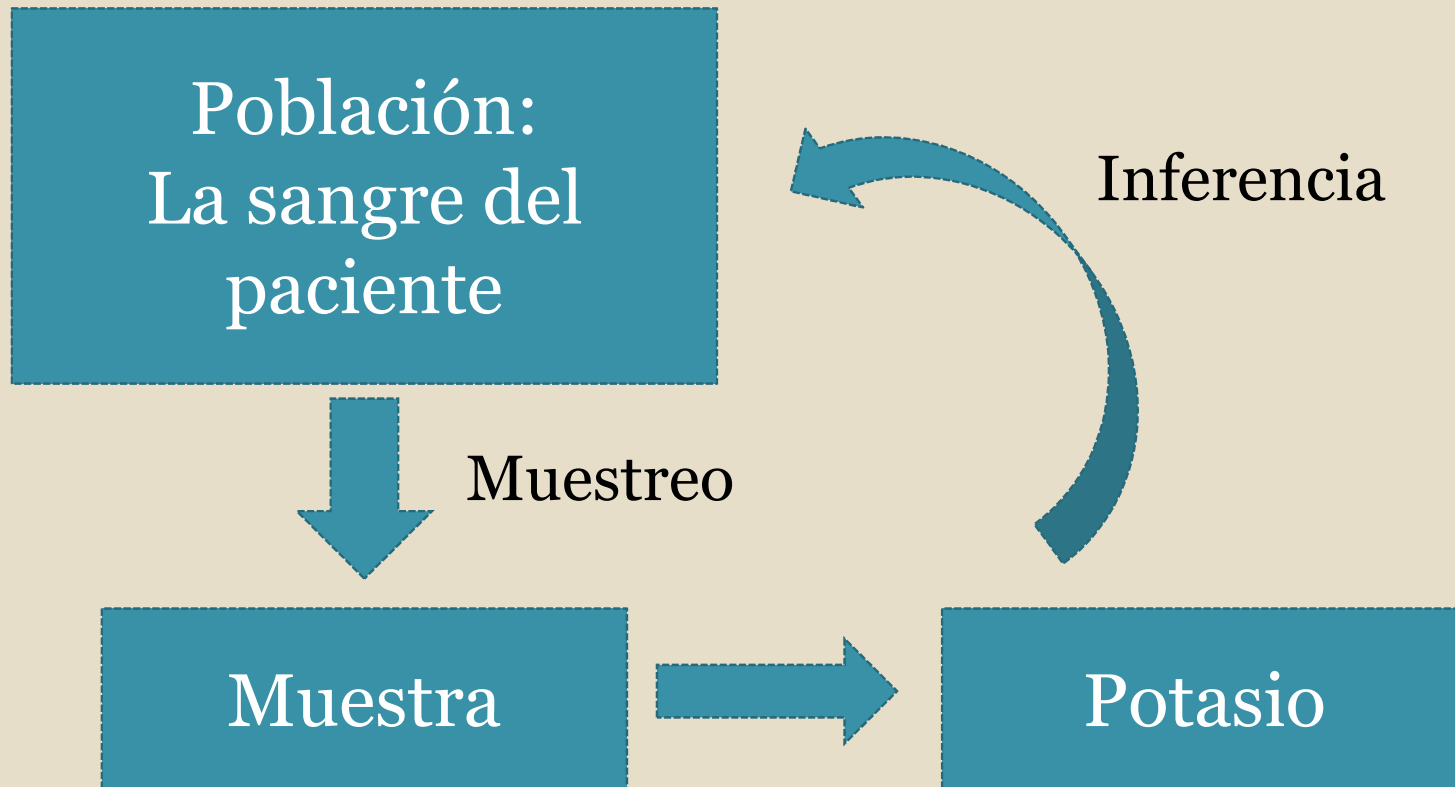
Descripción e inferencia



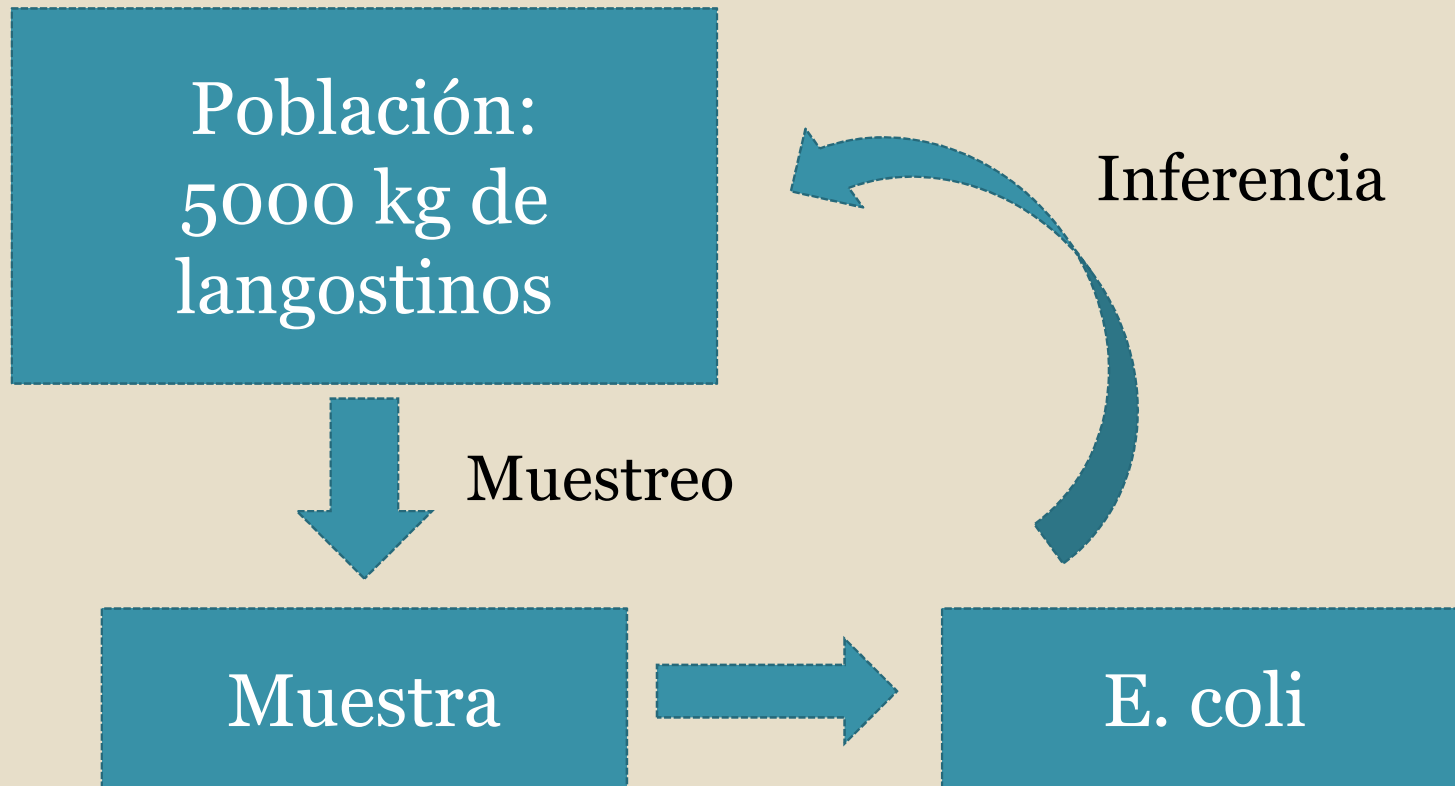
Descripción e inferencia



Descripción e inferencia



Descripción e inferencia



Inferencia



- Una población:
 - ¿Cuál es la frecuencia de cáncer en Cantabria?
 - ¿Qué tensión arterial tiene el paciente X?
- Dos poblaciones:
 - ¿Es más frecuente el cáncer en varones que en mujeres?
 - ¿Es más eficaz el tratamiento A que el B?

Tipos de variable



- Variables cualitativas (categóricas)
 - Binarias (dicotómicas)
 - Categóricas no ordenadas (nominales)
 - Categóricas ordenadas (ordinales)
- Variables cuantitativas
 - Discretas (habitualmente son el resultado de contar)
 - Continuas (habitualmente son el resultado de medir)

Tipos de variable: Binaria



- Sexo: Varón o mujer
- Enfermedad: Enfermo / No enfermo
- Una enfermedad concreta: Infarto / No infarto
- El resultado de un análisis: $VSG > 20$ / $VSG \leq 20$
- Asistencia al curso: Asiste / No asiste
- Resultado del curso: Aprueba / No aprueba
- Equipo de fútbol: At. Madrid / No At. Madrid

Tipos de variable: Nominal



- Estado civil: Soltero / Casado / Viudo / Divorciado / Separado
- Tabaquismo: Fumador / Ex fumador / No fumador
- Enfermedades: Infarto / Diabetes / Alzheimer /...
- Equipo de futbol: At. Madrid / Real Madrid / Barcelona / Depor / ...

Tipos de variable: Ordinal



- Enfermedad: IAM / Angina / Nada
- Criterios de Ramsom
- Índice ASA
- Clase funcional NYHA
- Tabaquismo: No / 1-5 / 6-10 / 11-20 / >20
- Edad: 0-14 / 15-24 / 25-44 / 45-64 / 65-74 / ≥ 75
- Gravedad: Leve / Moderado / Grave / Crítico
- Tipo de cirugía: Limpia / Limpia-contaminada / Contaminada / Sucia

Tipos de variable: Cuantitativas



- **Discretas:**
 - Edad (años cumplidos): 0, 1, 2,...
 - Frecuencia cardíaca
 - Número de hijos
- **Continuas**
 - Edad: 20.3
 - Altura
 - Peso
 - Temperatura

Tipos de variable



- Cualitativa
 - Binomial y nominal: CLASIFICAR
 - Ordinal: JERARQUIZAR
- Cuantitativa
 - Discreta: CONTAR
 - Continua: MEDIR

Representación gráfica

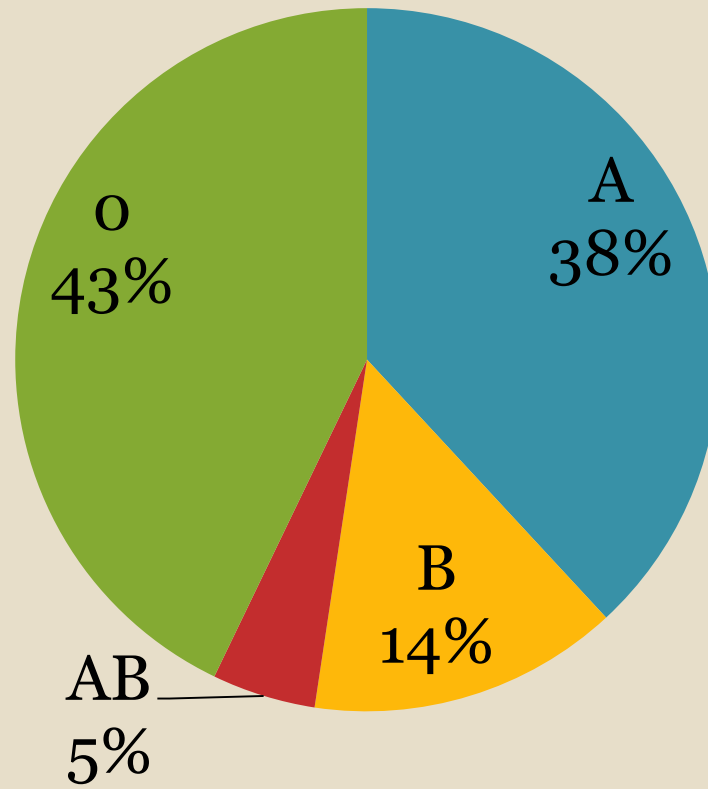


- Una variable
 - Sectores (“tarta”)
 - Diagrama de barras
 - Histograma
 - Tallo y hojas
 - Polígono de frecuencias
 - Caja
- Dos variables
 - Dispersión
 - Líneas

Sectores



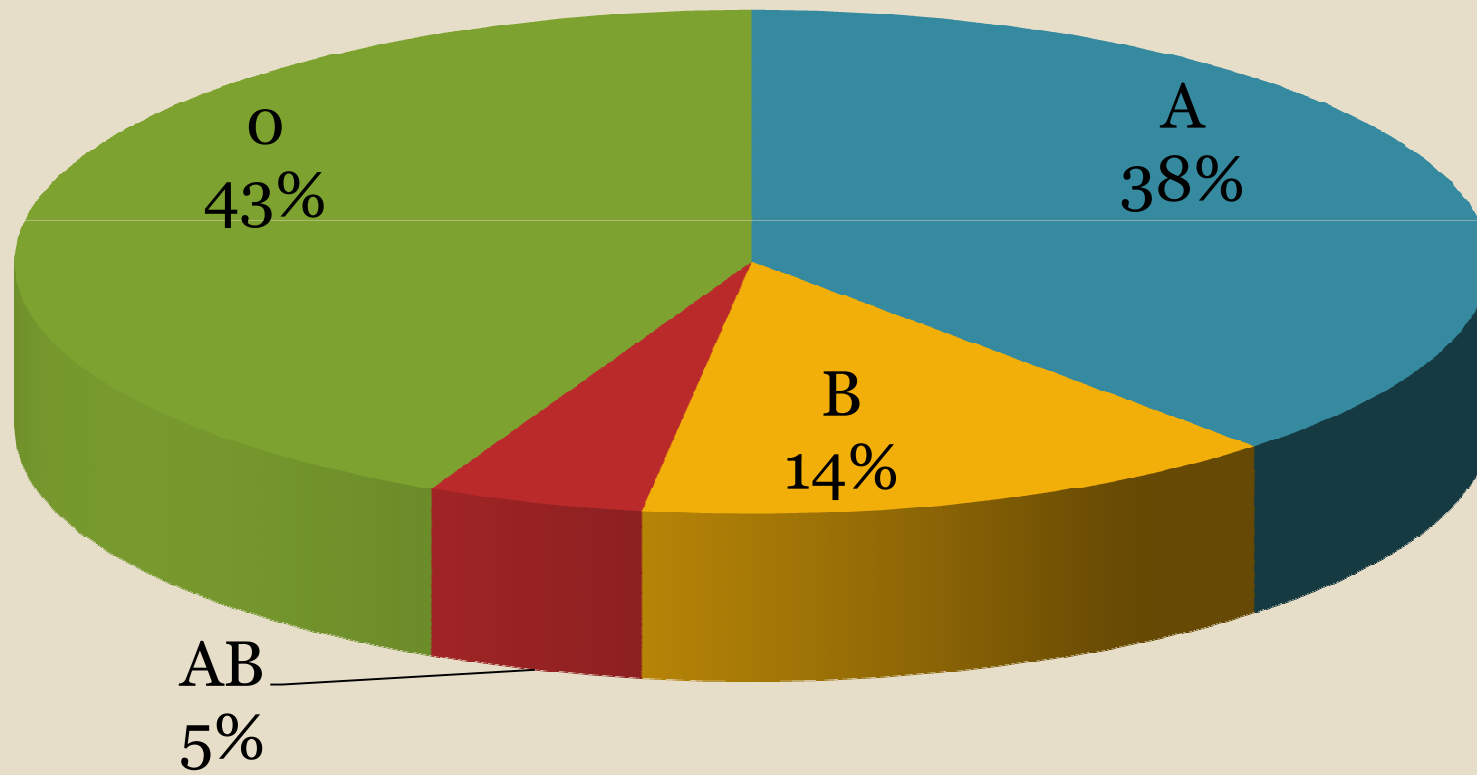
Grupo sanguíneo



Sectores



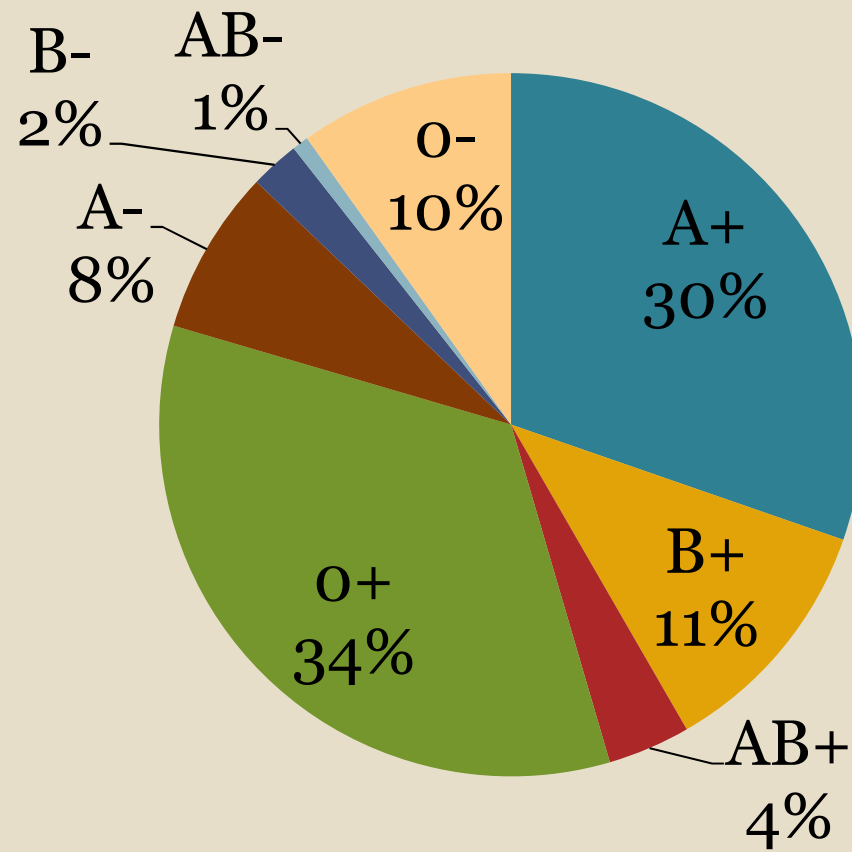
Grupo sanguíneo



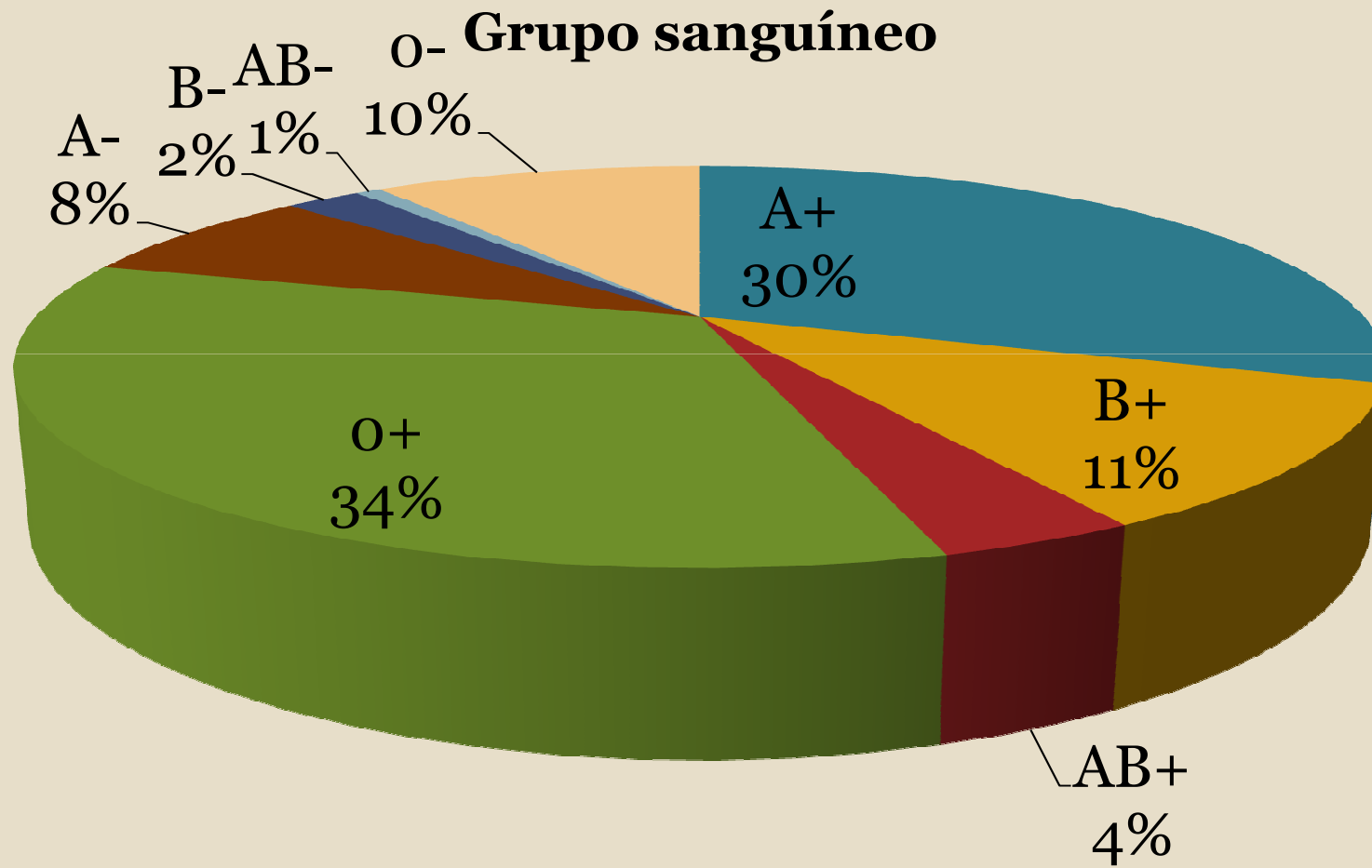
Sectores



Grupo sanguíneo



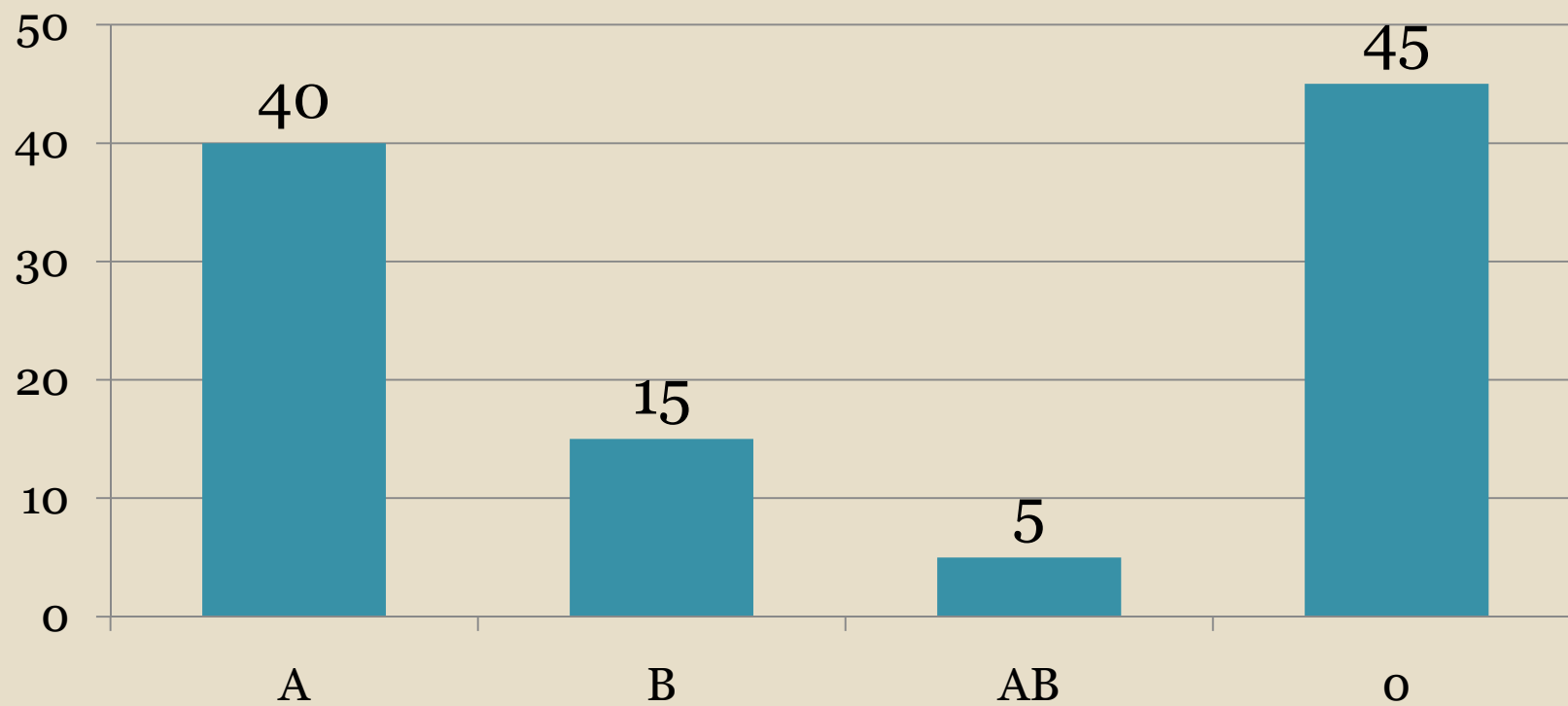
Sectores



barras



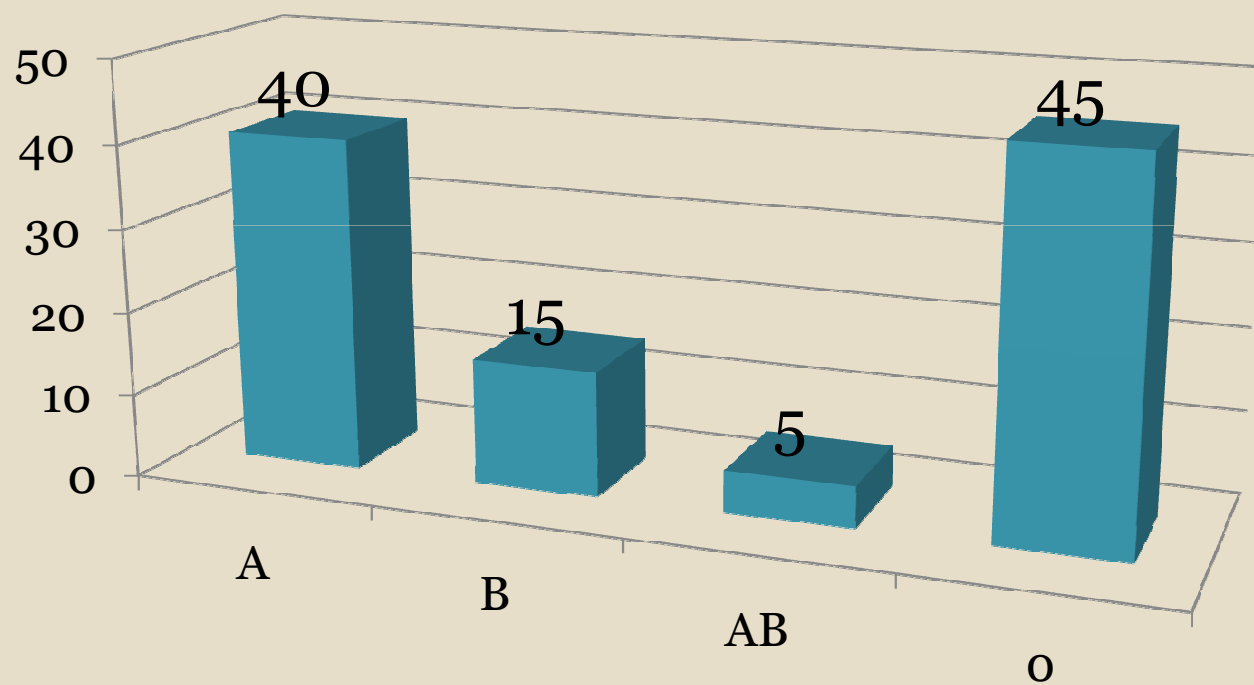
Grupo sanguíneo



barras



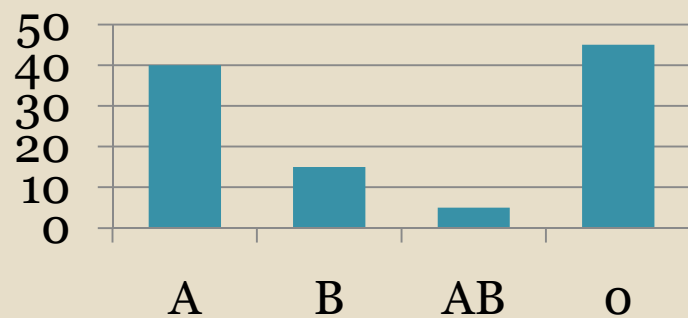
Grupo sanguíneo



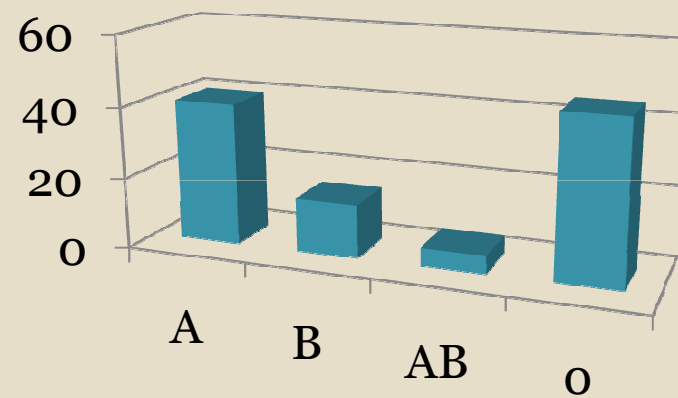
barras



Grupo sanguíneo



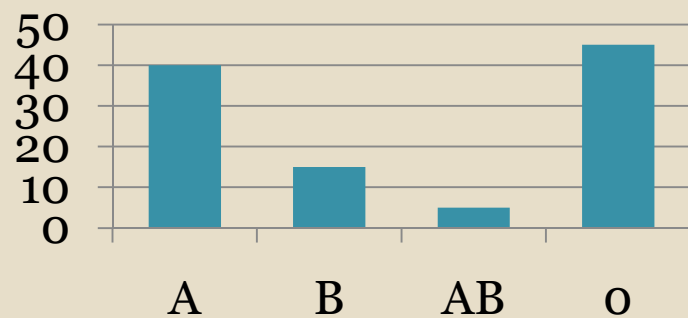
Grupo sanguíneo



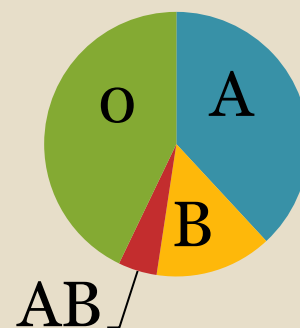
barras



Grupo sanguíneo



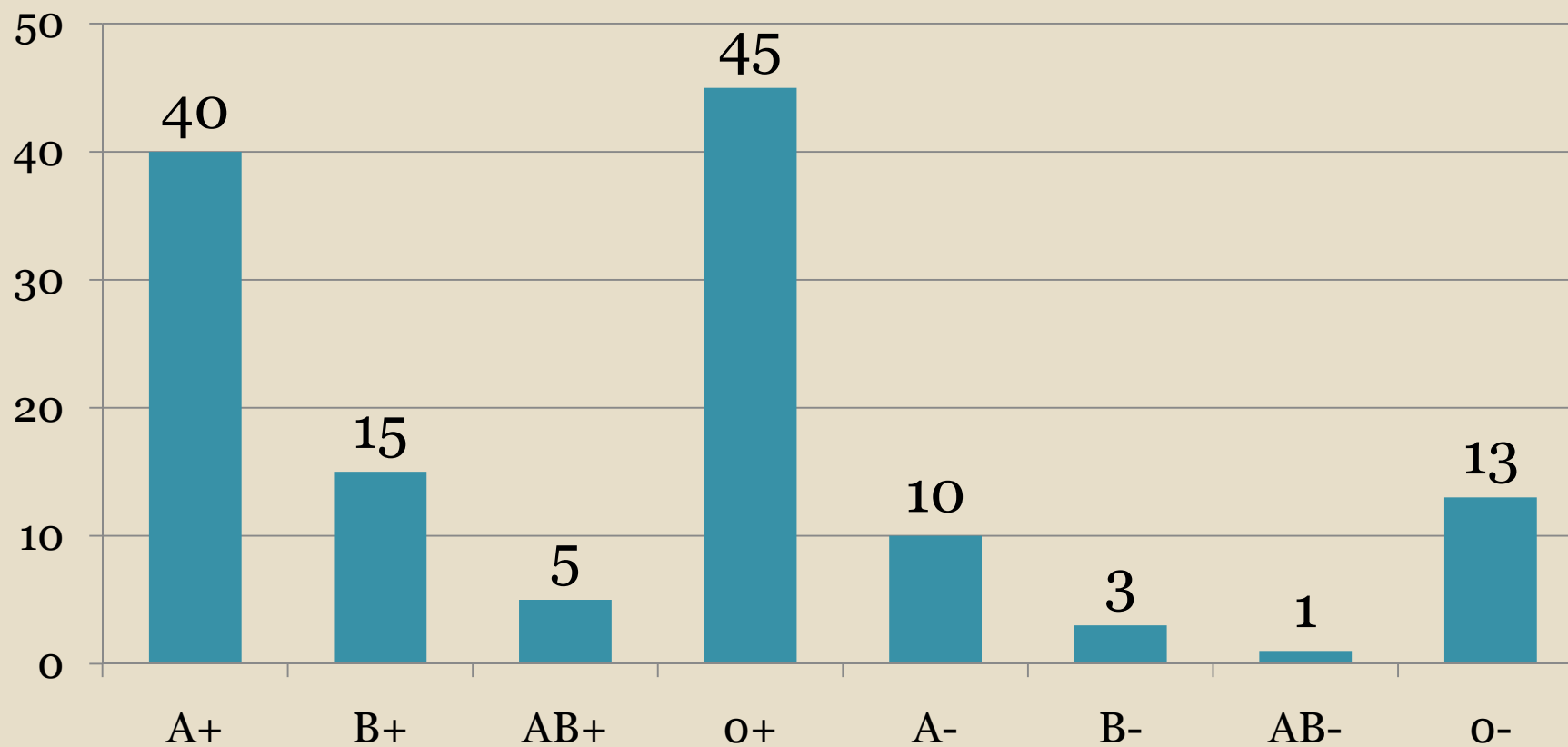
Grupo sanguíneo



barras



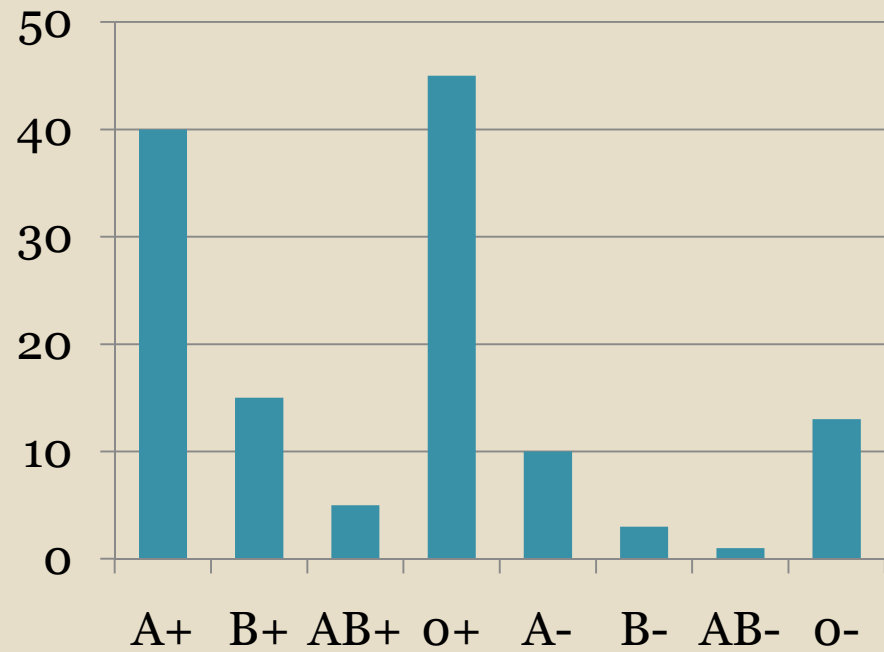
Grupo sanguíneo



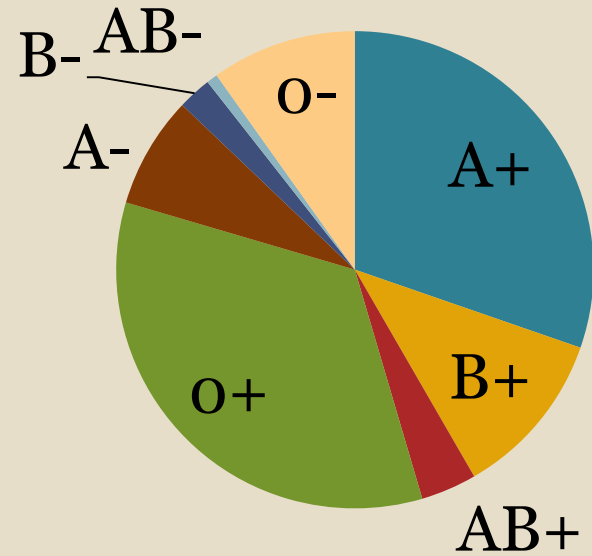
barras



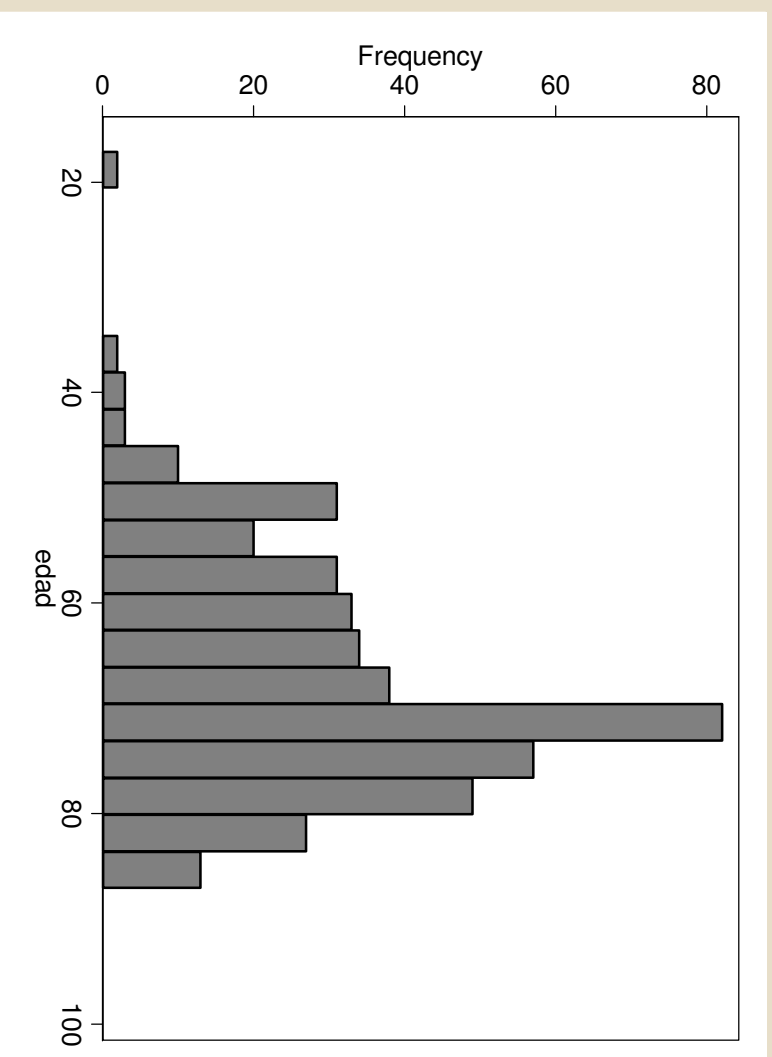
Grupo sanguíneo



Grupo sanguíneo



histograma



Tallo y hojas



7* | 0000000000000000000000111111111111111111111
7t | 22222222222222222222222222222222222233333333333333333333
7f | 4444444444444444444444555555555555555555555
7s | 66666666666666667777777777777777
7. | 88888888888888888888999999999999999999999
8* | 000000001111111111
8t | 2222223333333333
8f | 4444445555
8s | 6677

Diagrama de caja (“Box and whiskers”)

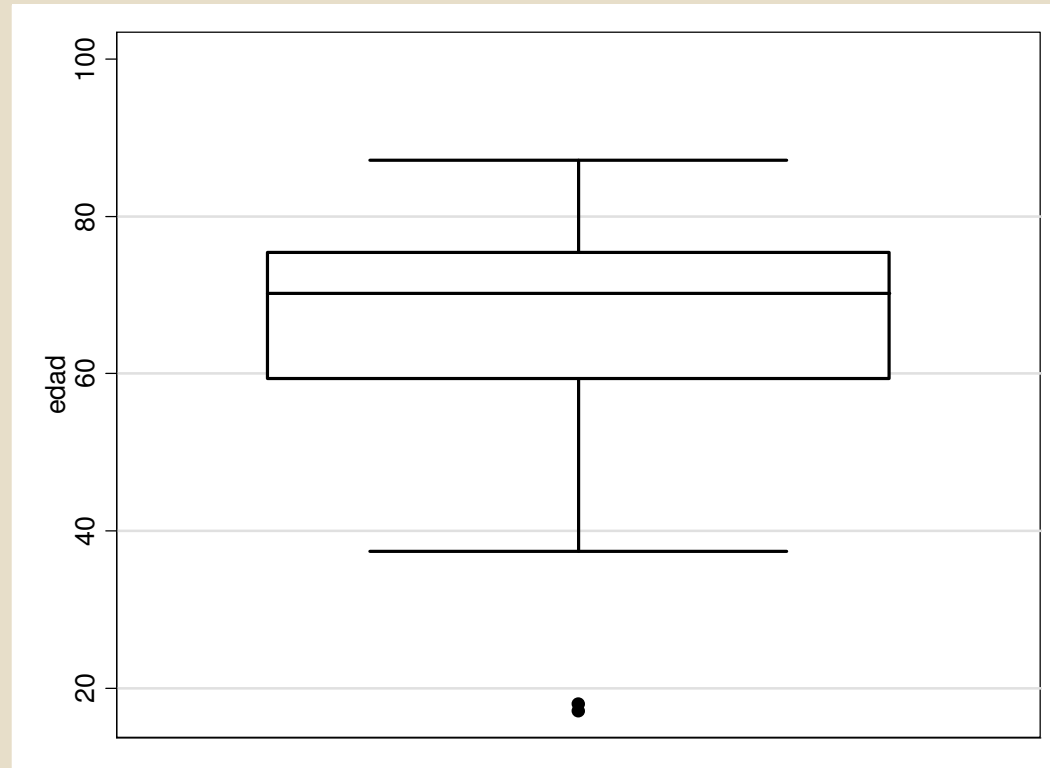
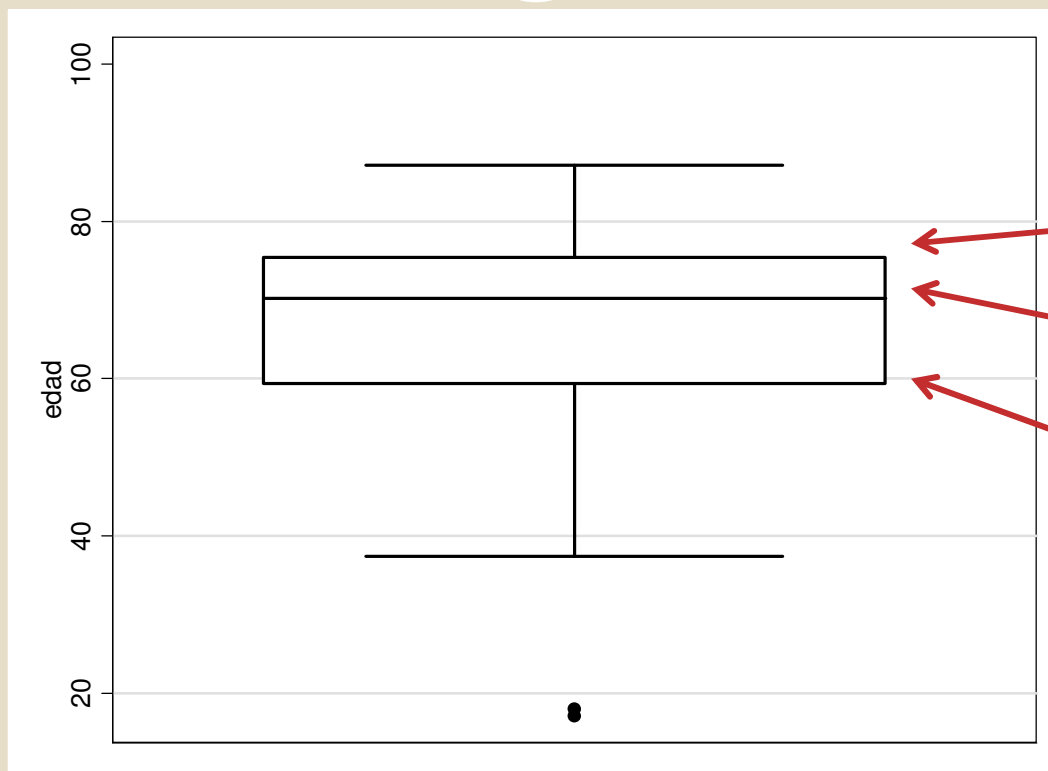


Diagrama de caja (“Box and whiskers”)

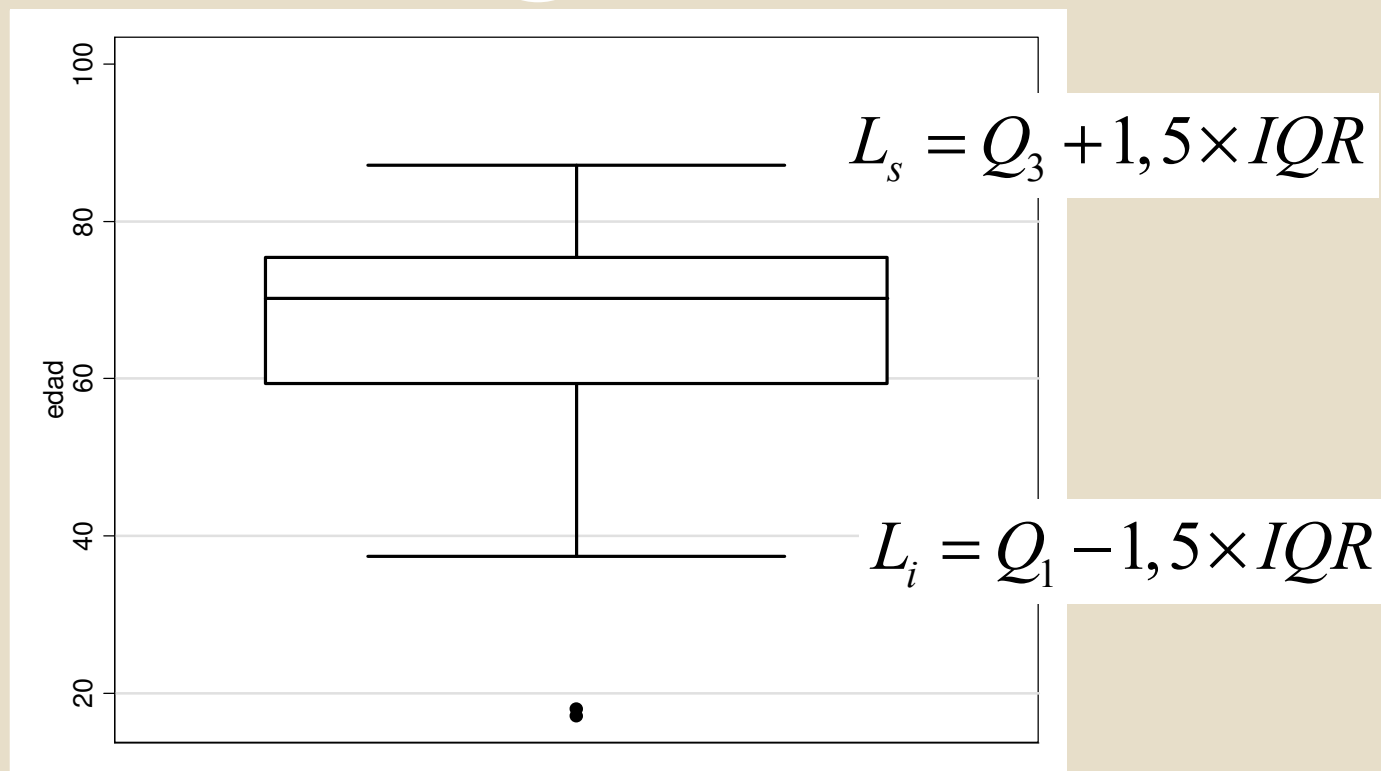


Cuartil 3

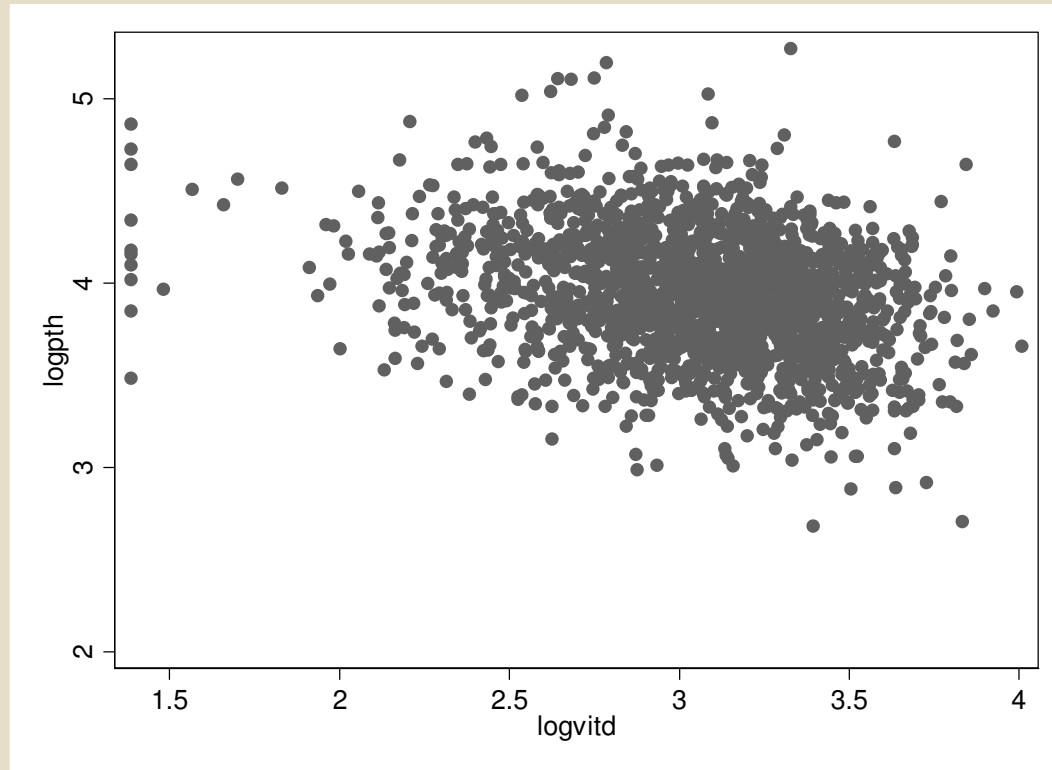
Mediana

Cuartil 1

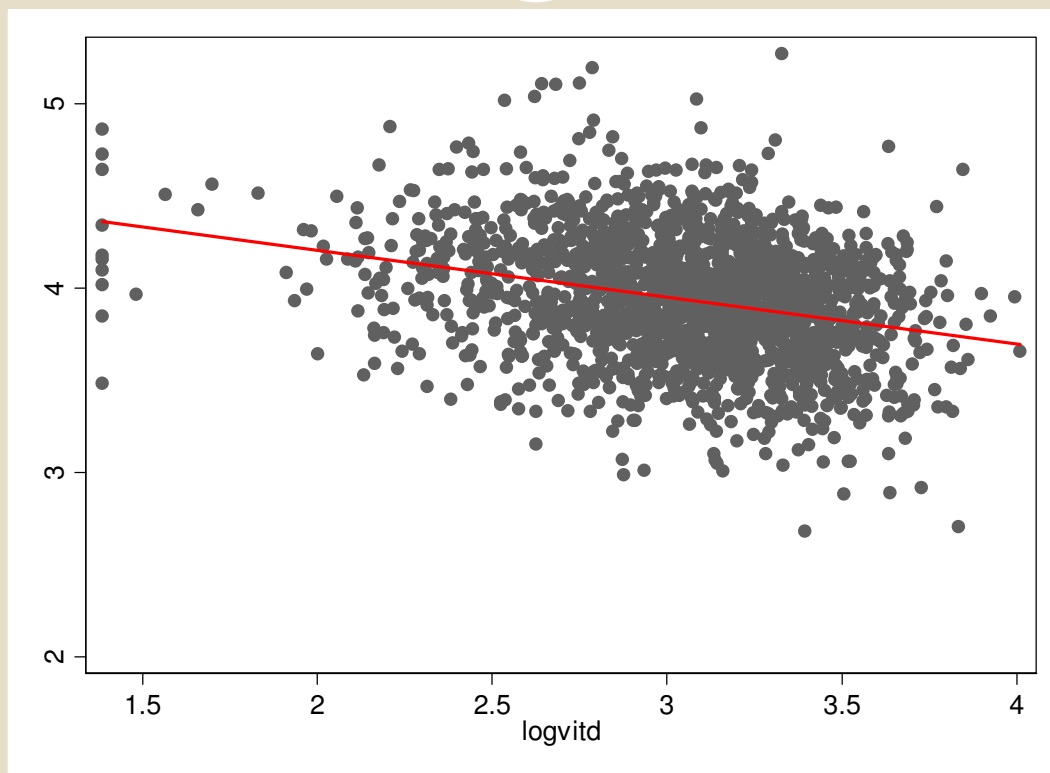
Diagrama de caja (“Box and whiskers”)



dispersión



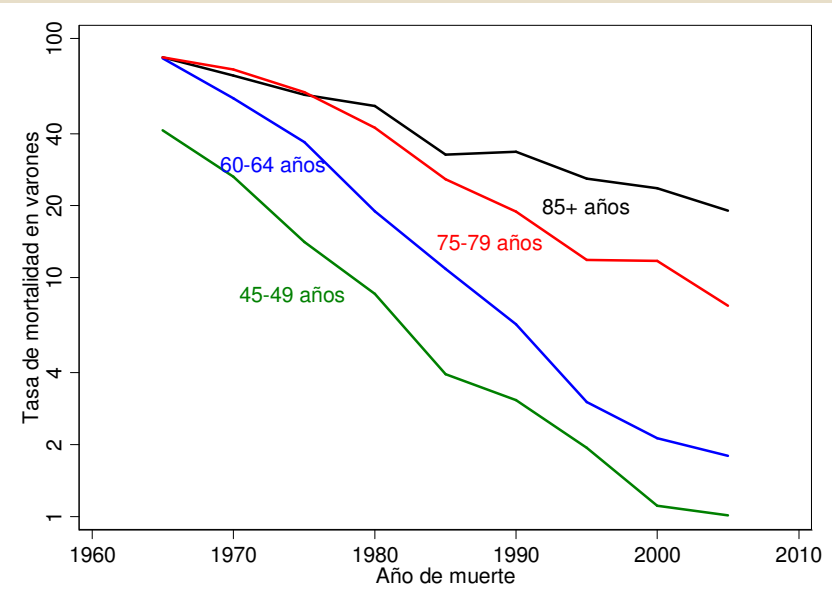
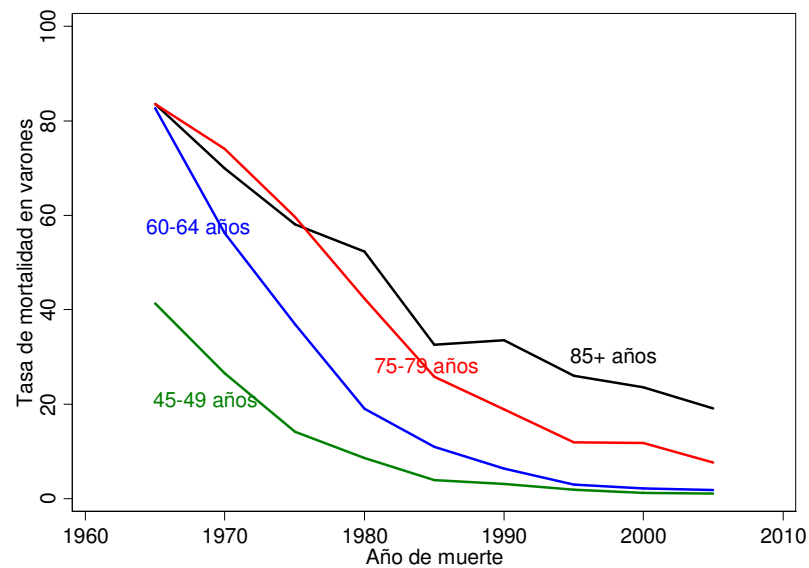
Dispersión+línea de tendencia



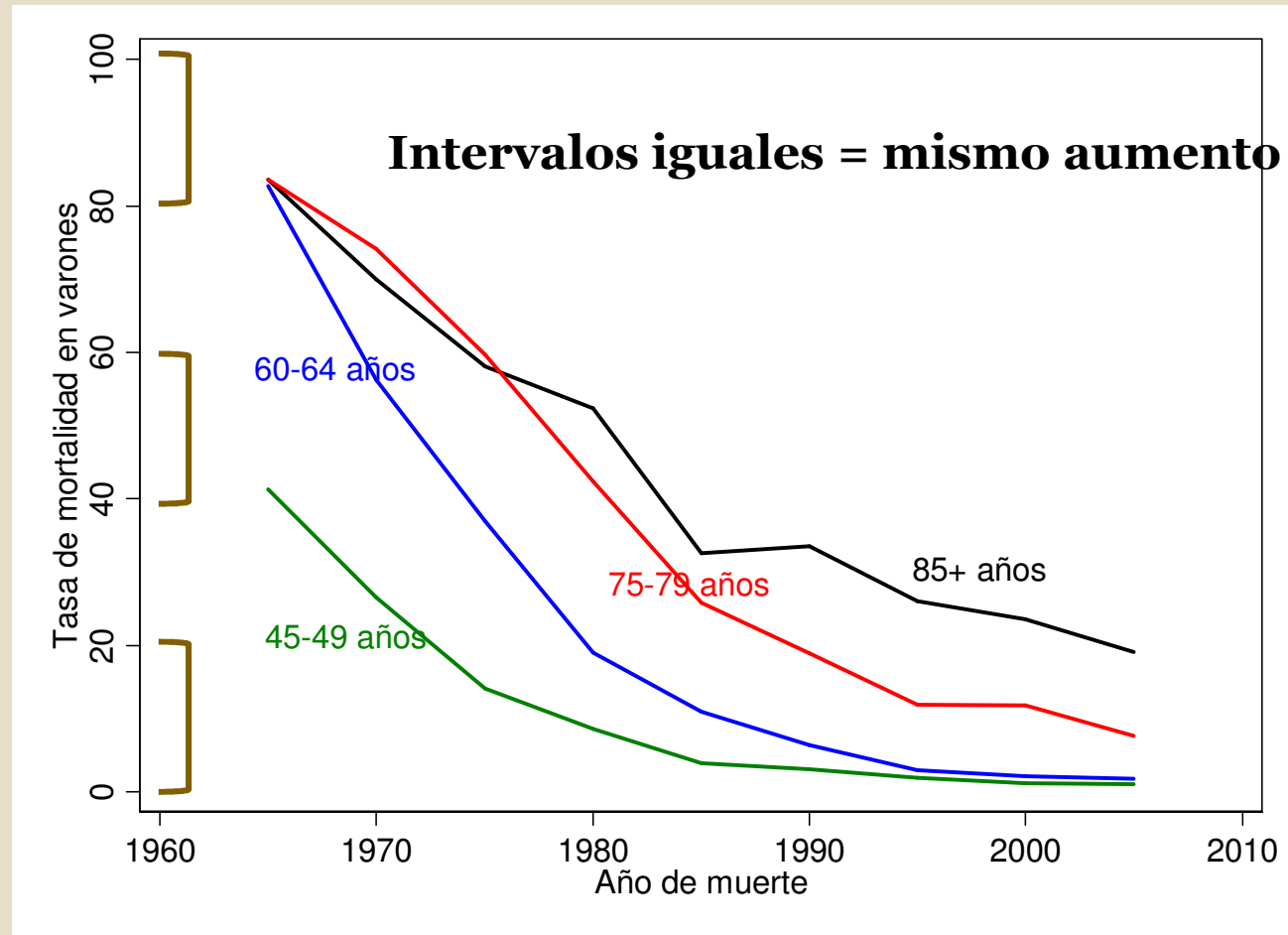
Líneas



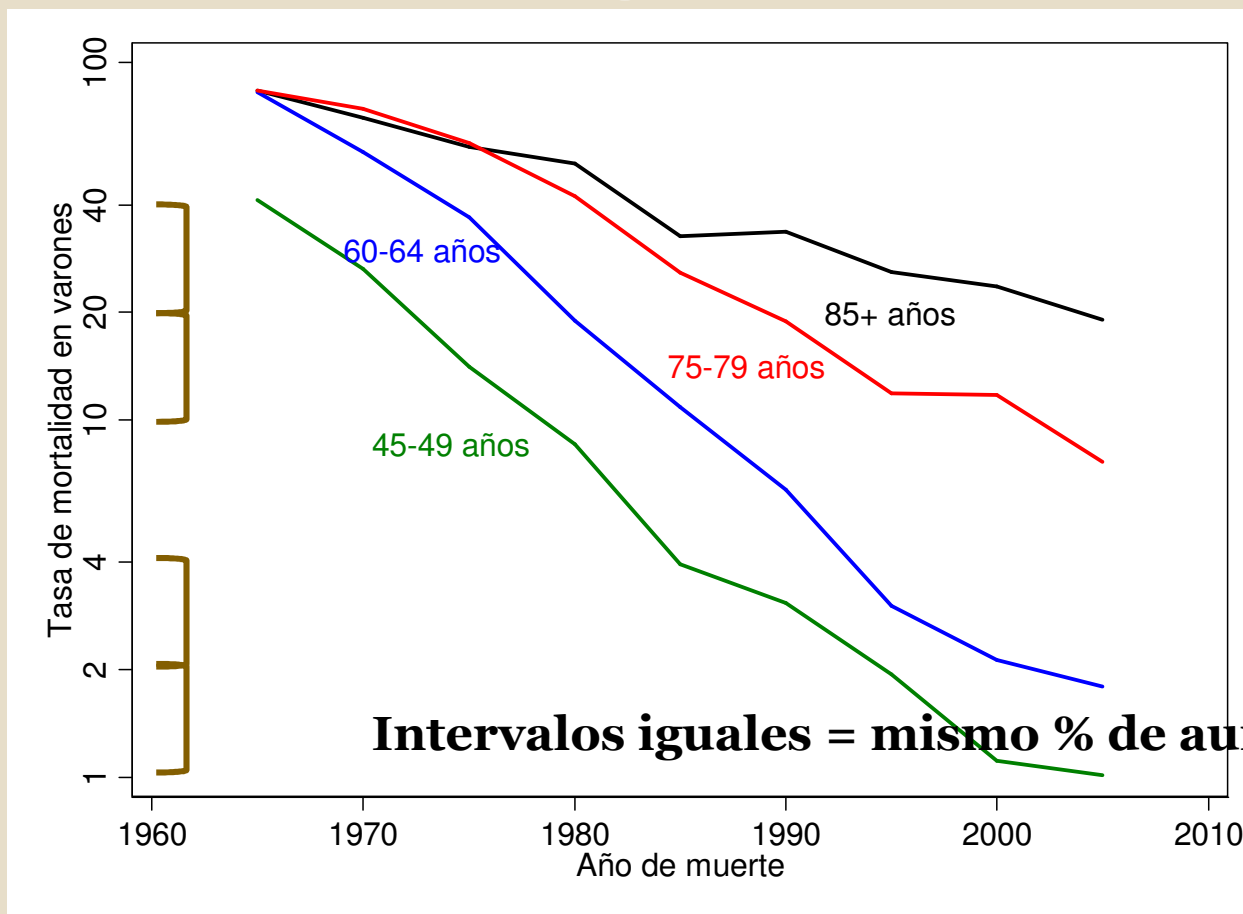
Mortalidad por tuberculosis en España



Escala aritmética

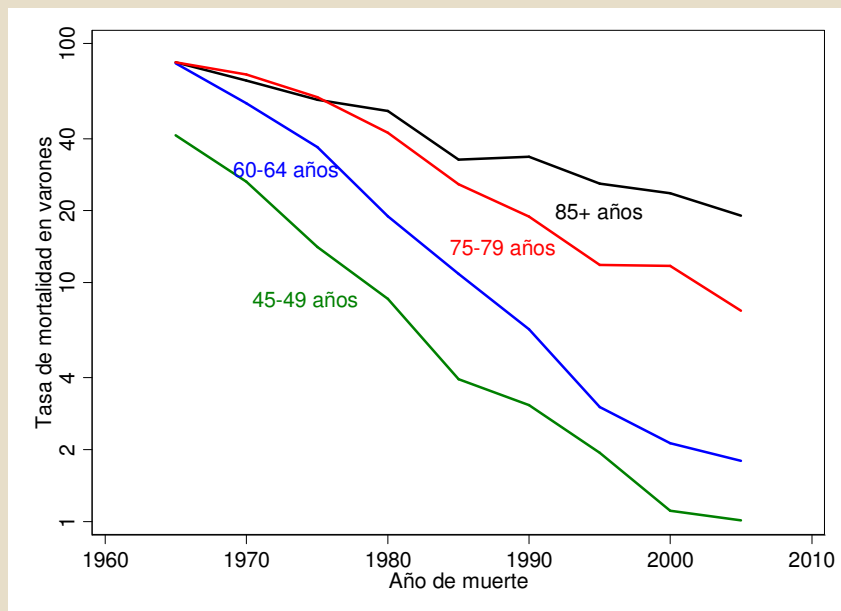
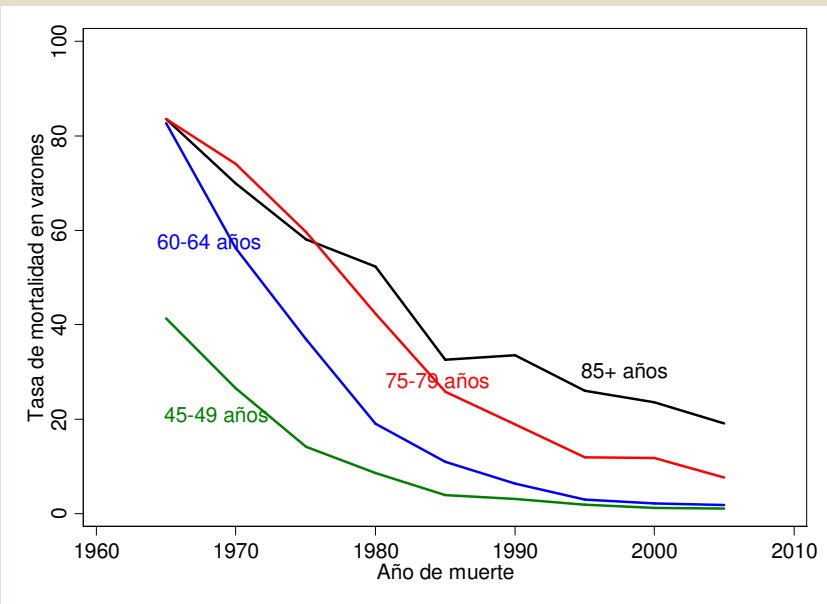


Escala logarítmica





Mortalidad por tuberculosis en España





Escala aritmética

- Más fácil de entender
- Líneas paralelas = los dos grupos aumentan o disminuyen en la misma cantidad

Escala logarítmica

- Ver “con lupa” algunas zonas del gráfico
- Líneas paralelas = los dos grupos aumentan o disminuyen en el mismo porcentaje

Diagrama de sectores



- Sólo para variables binarias o nominales
- Transmiten muy poca información

Diagrama de barras



- Variables categóricas
 - Binarias
 - Nominales
 - Ordinales
- Variables cuantitativas discretas

Histograma y diagrama de cajas



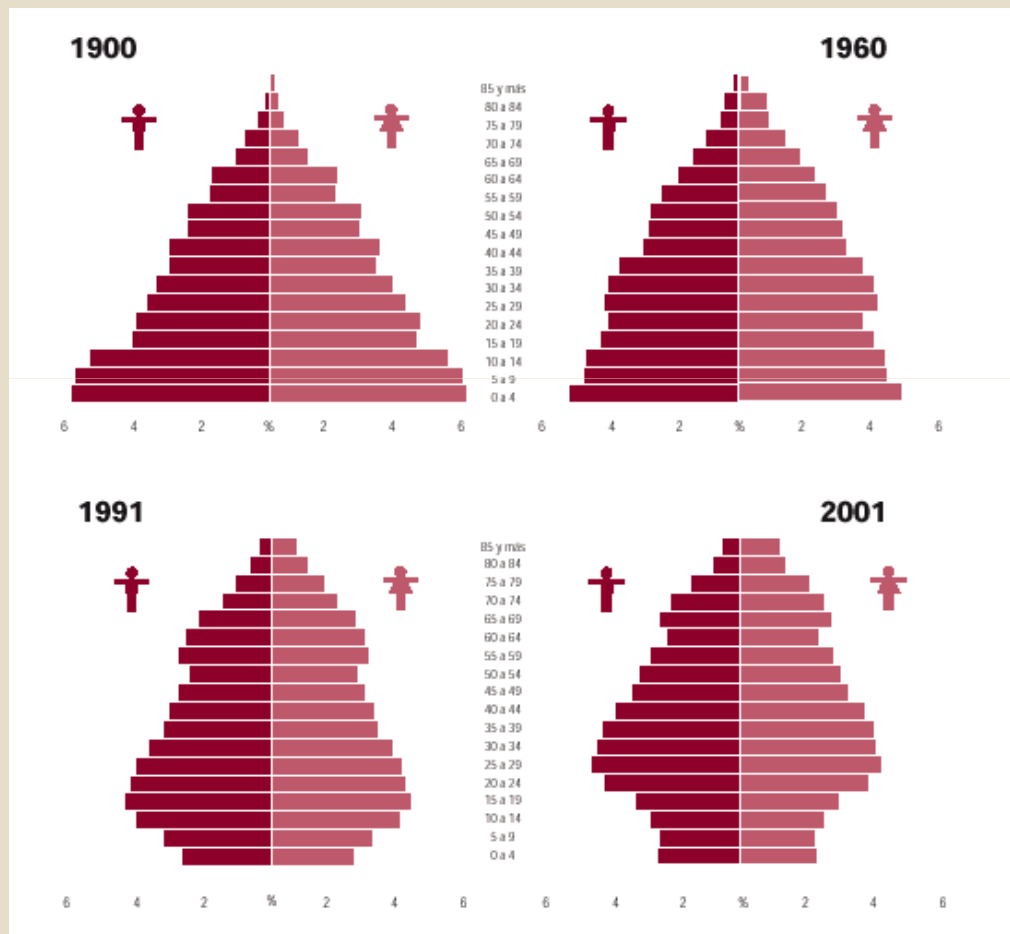
- Variables cuantitativas
 - Discretas
 - Continuas

Dos variables



- Relacionar dos variables categóricas
 - Diagrama de barras
- Relacionar una variable categórica y una cuantitativa
 - Diagrama de barras
 - Histograma
 - Diagrama de caja
- Relacionar dos variables cuantitativas
 - Diagrama de dispersión
 - Diagrama de líneas

Población española



Mortalidad por cáncer en España

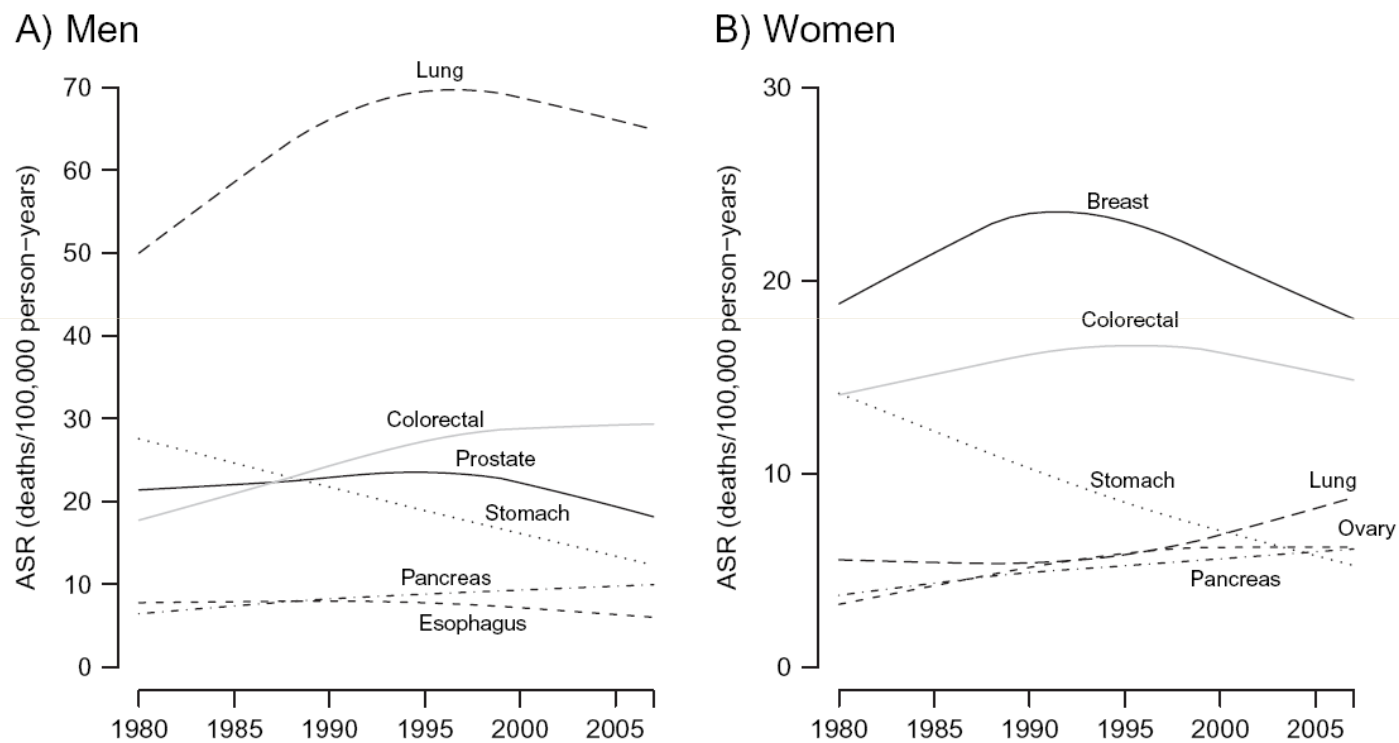
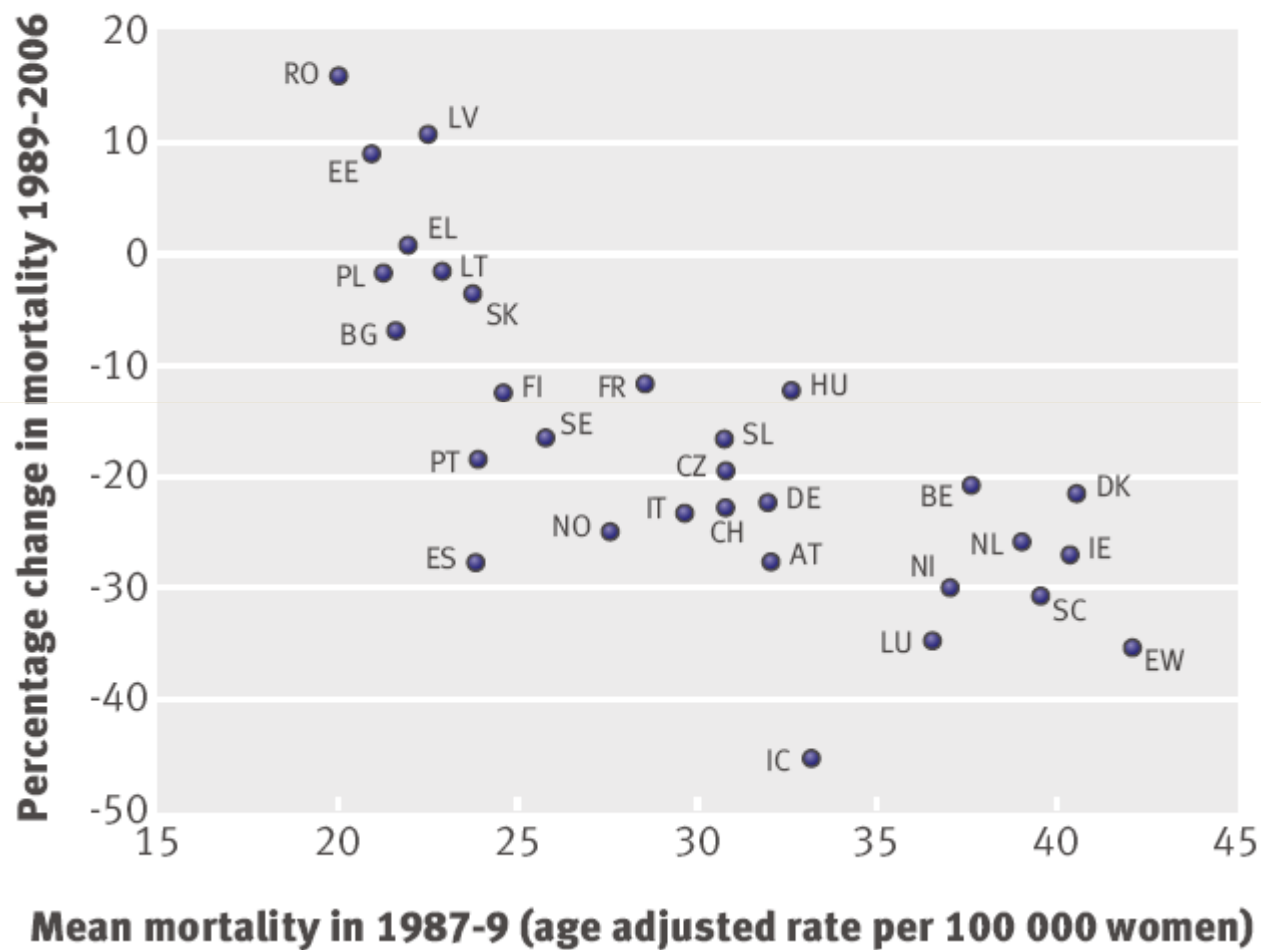


Figure 1. Mortality trends among men and women for selected types of cancer: Spain, 1980–2007.

Fuente: Cabanes et al, Ann Oncol 2010

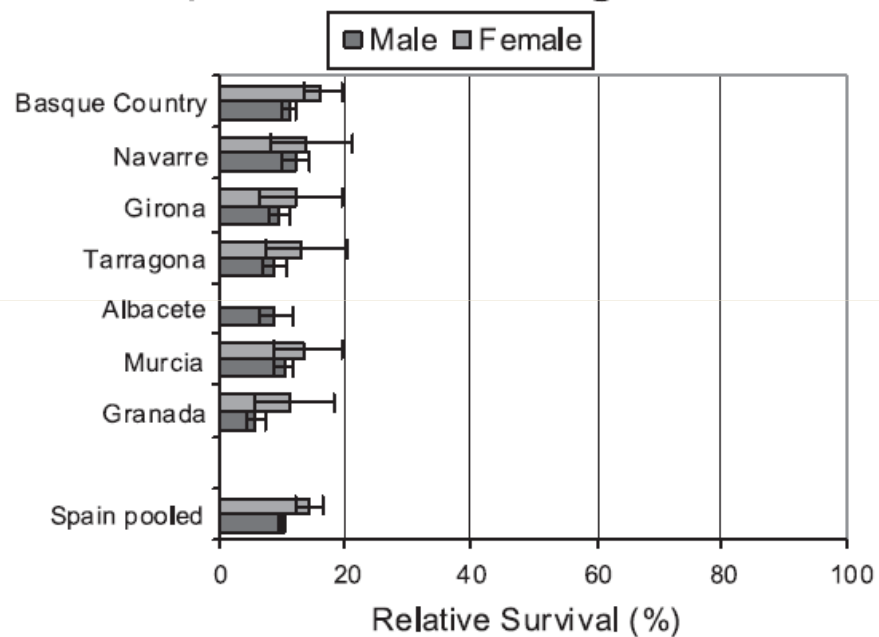
Cáncer de mama, Europa



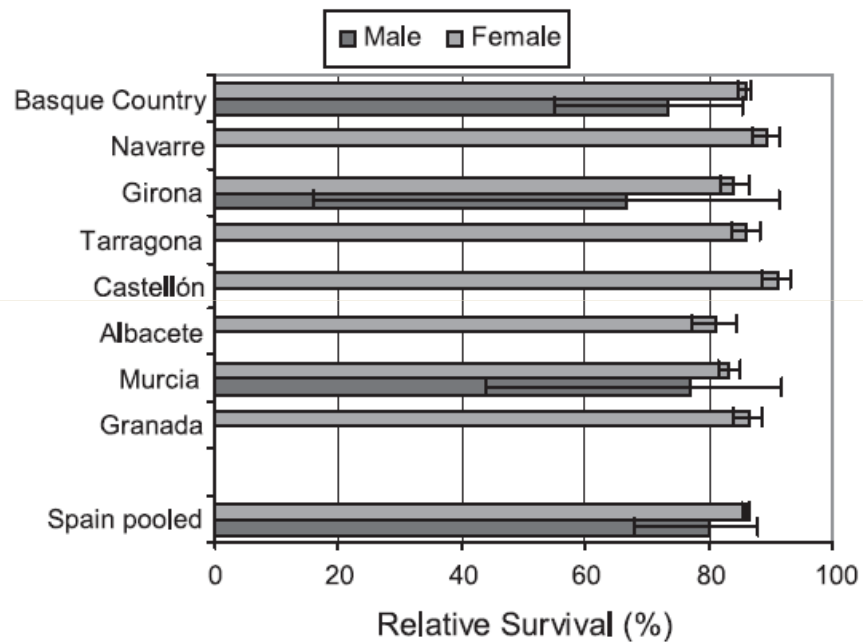
Fuente: Autier et al. BMJ 2010



Trachea, bronchus and lung

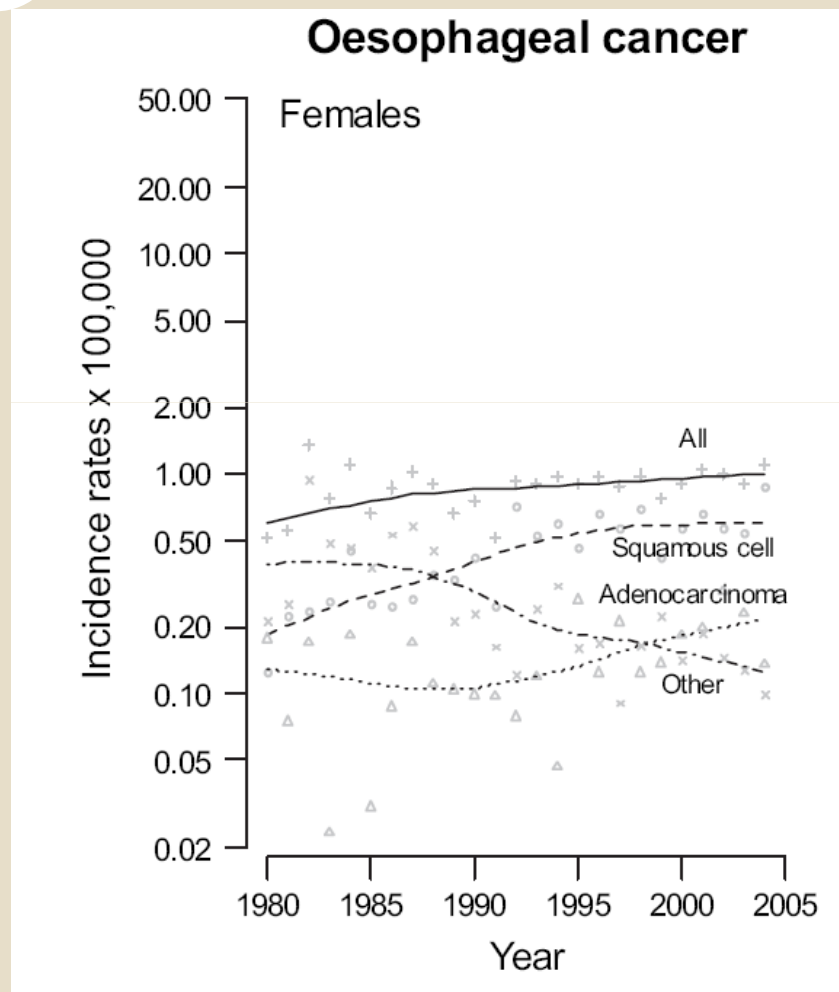
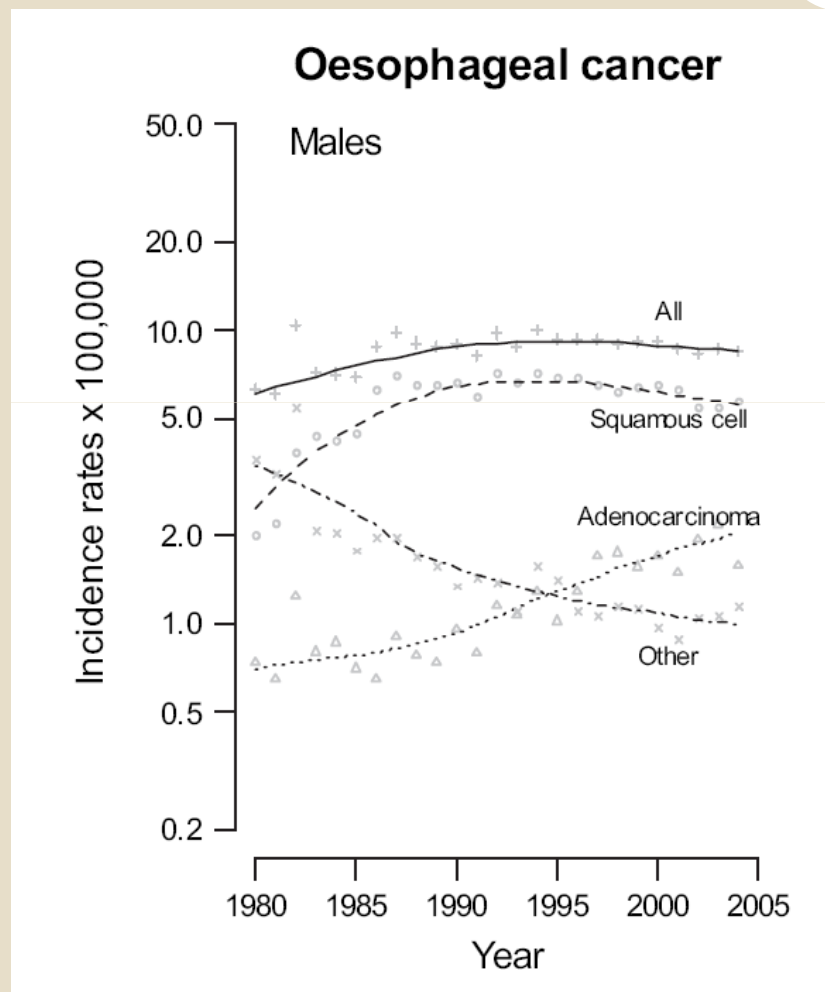


Breast

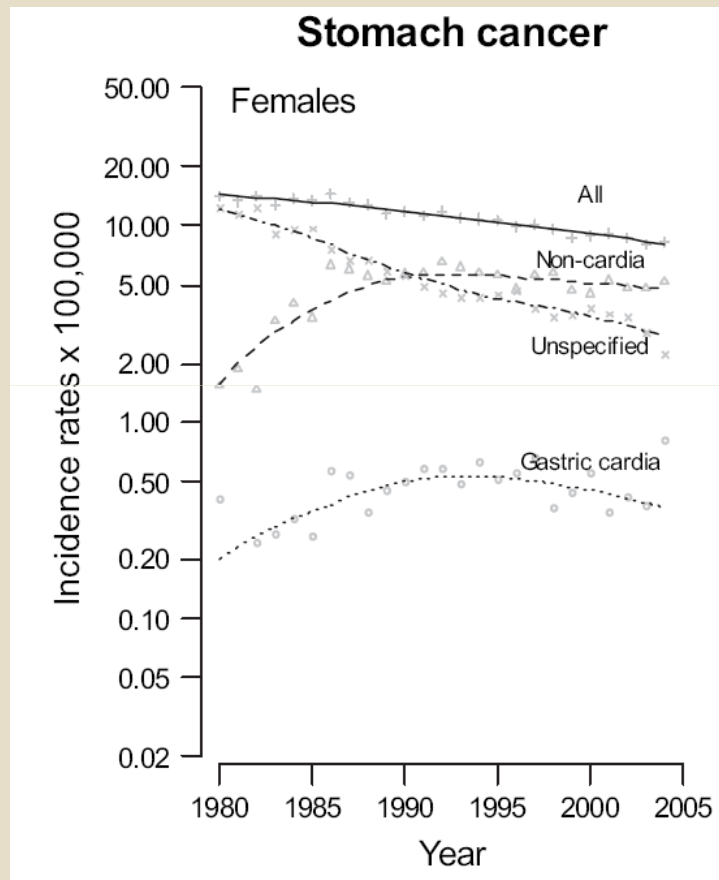
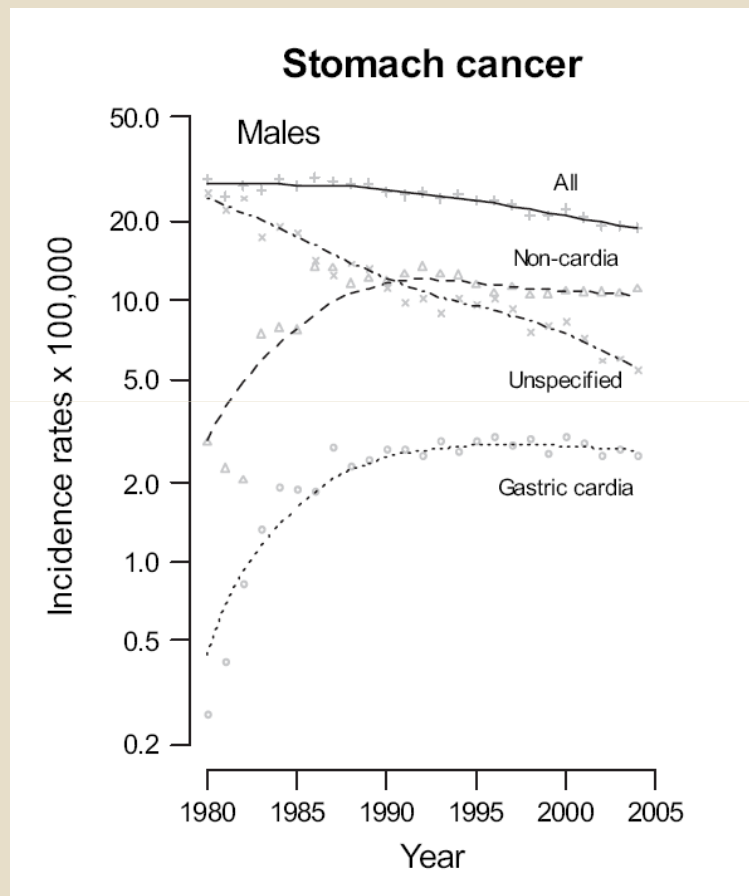


Fuente: Chirlaque et al. Ann Oncol 2010

Cáncer de esófago en España

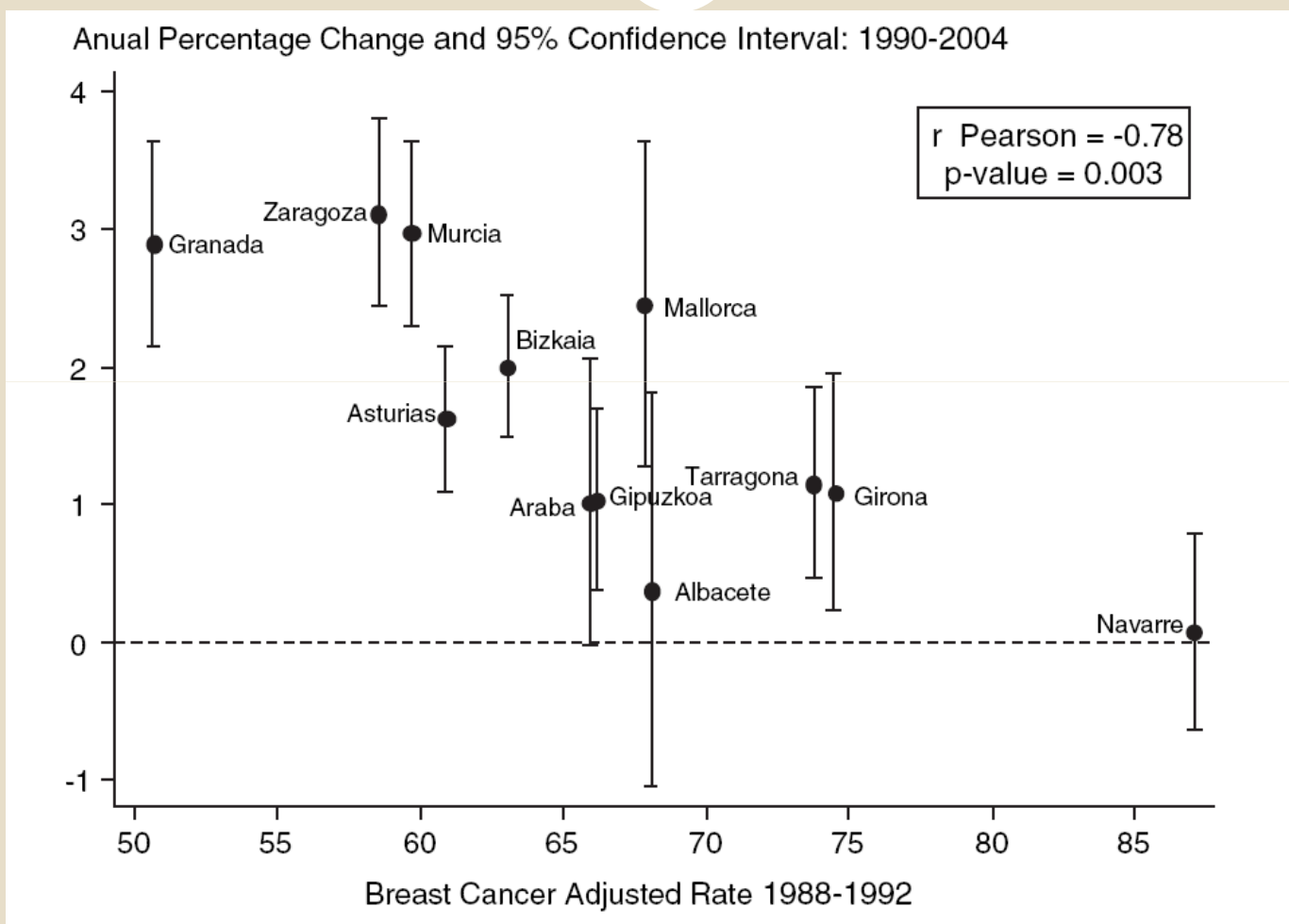


Cáncer de estómago en España



Fuente: Aragonés et al. Ann Ocol 2010

Cáncer de mama en España

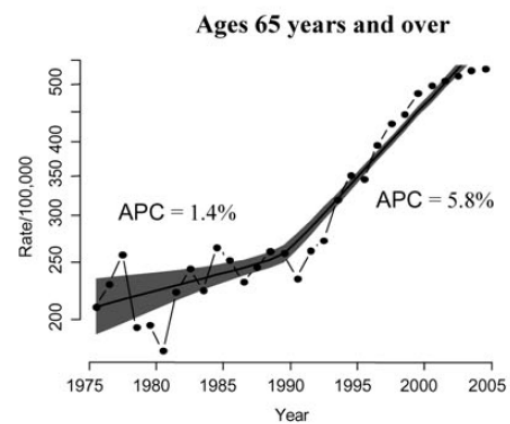
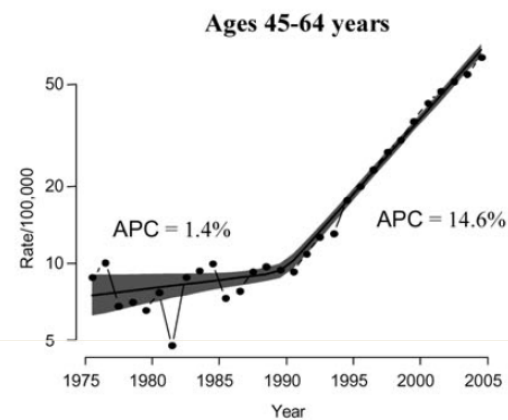
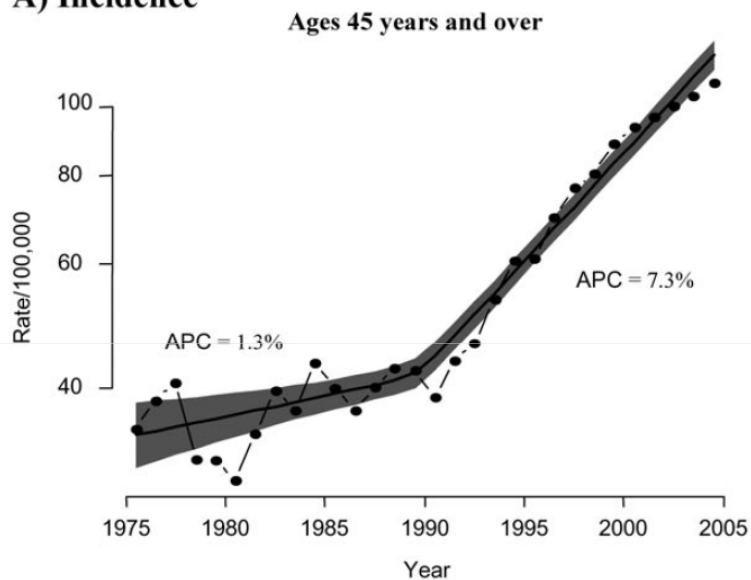


Fuente: Pollán et al. Ann Ocol 2010

Cáncer de próstata en España

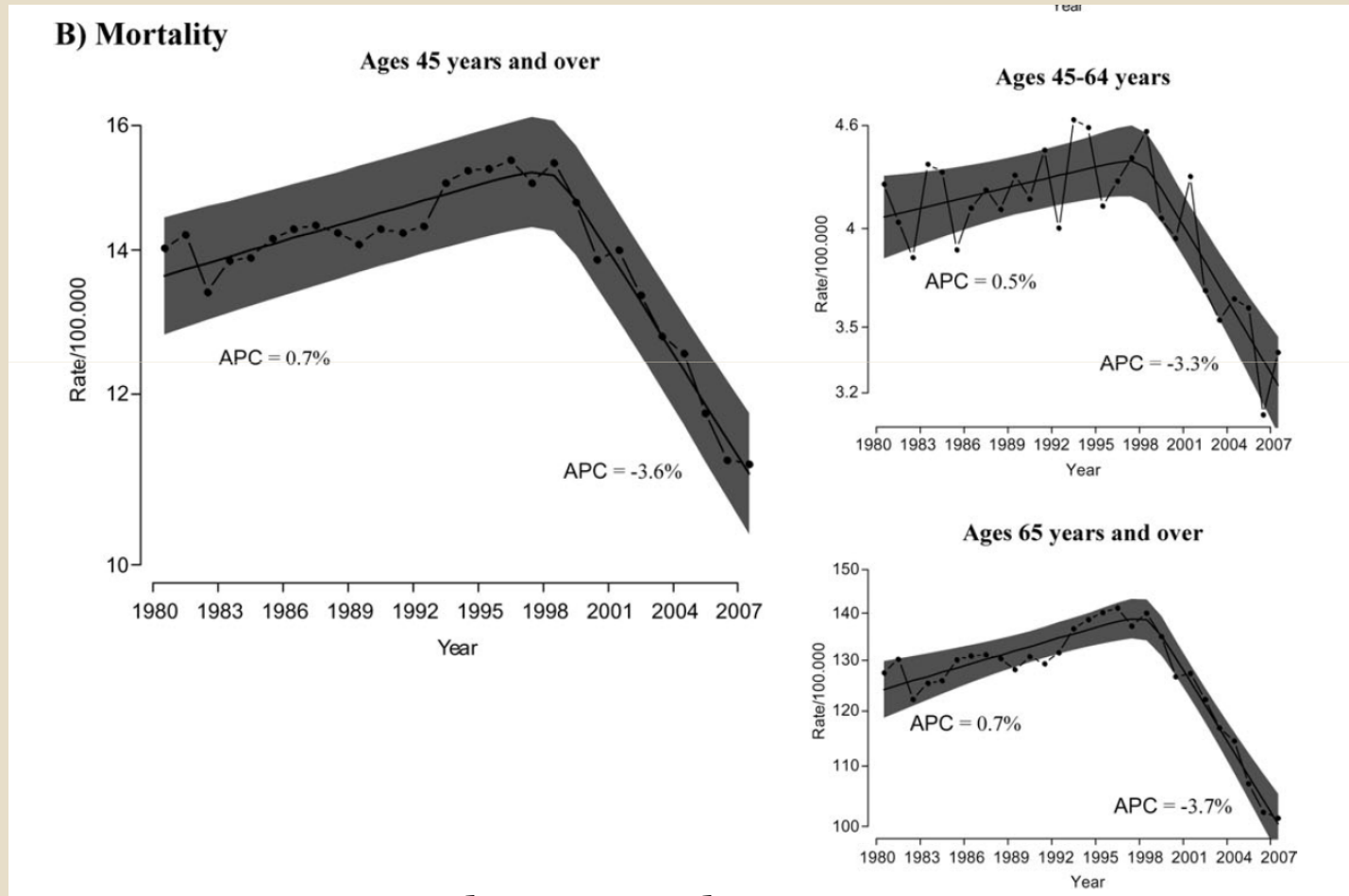


A) Incidence



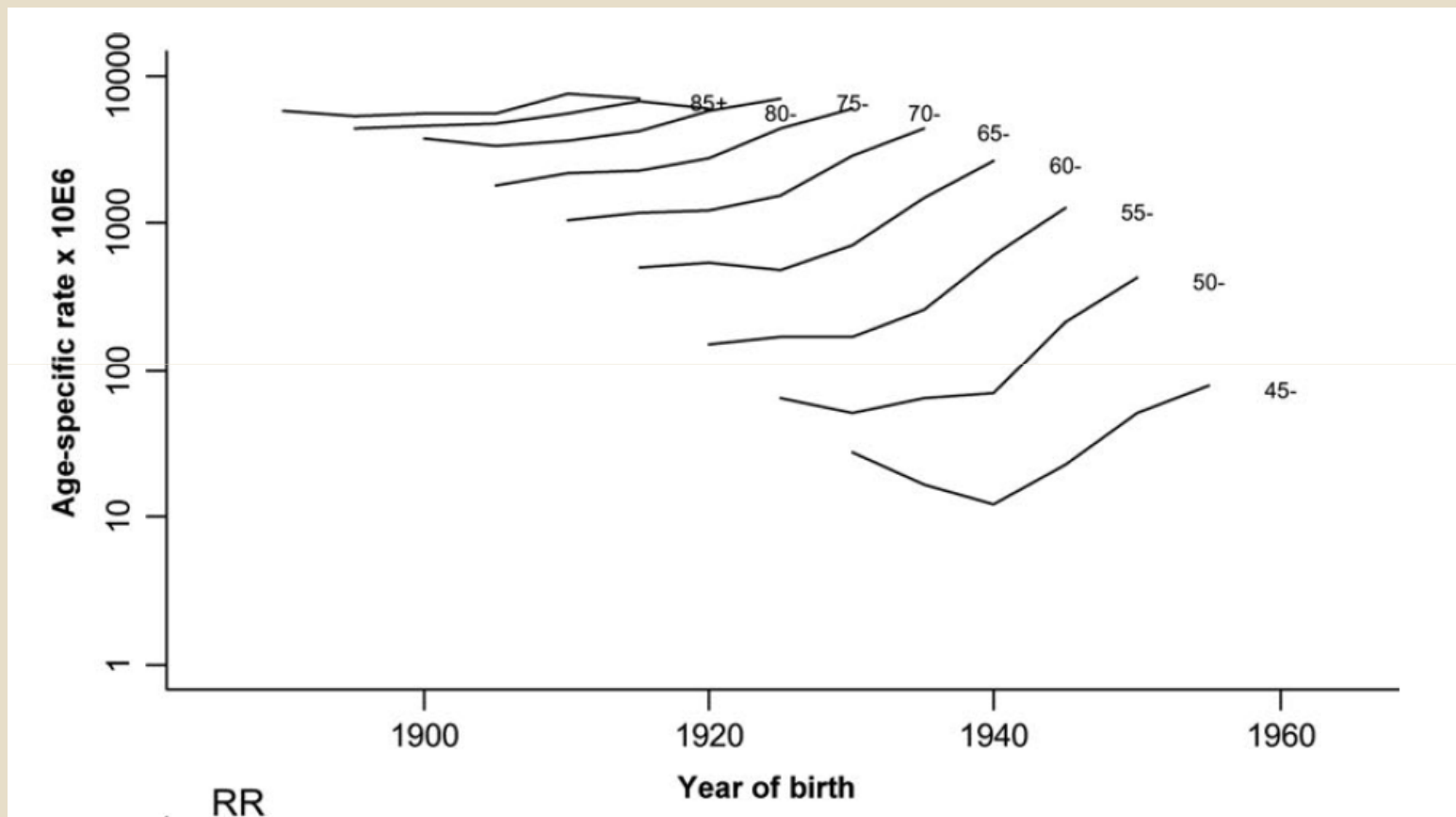
Fuente: Larrañaga et al. Ann Oncol 2010

Cáncer de próstata en España



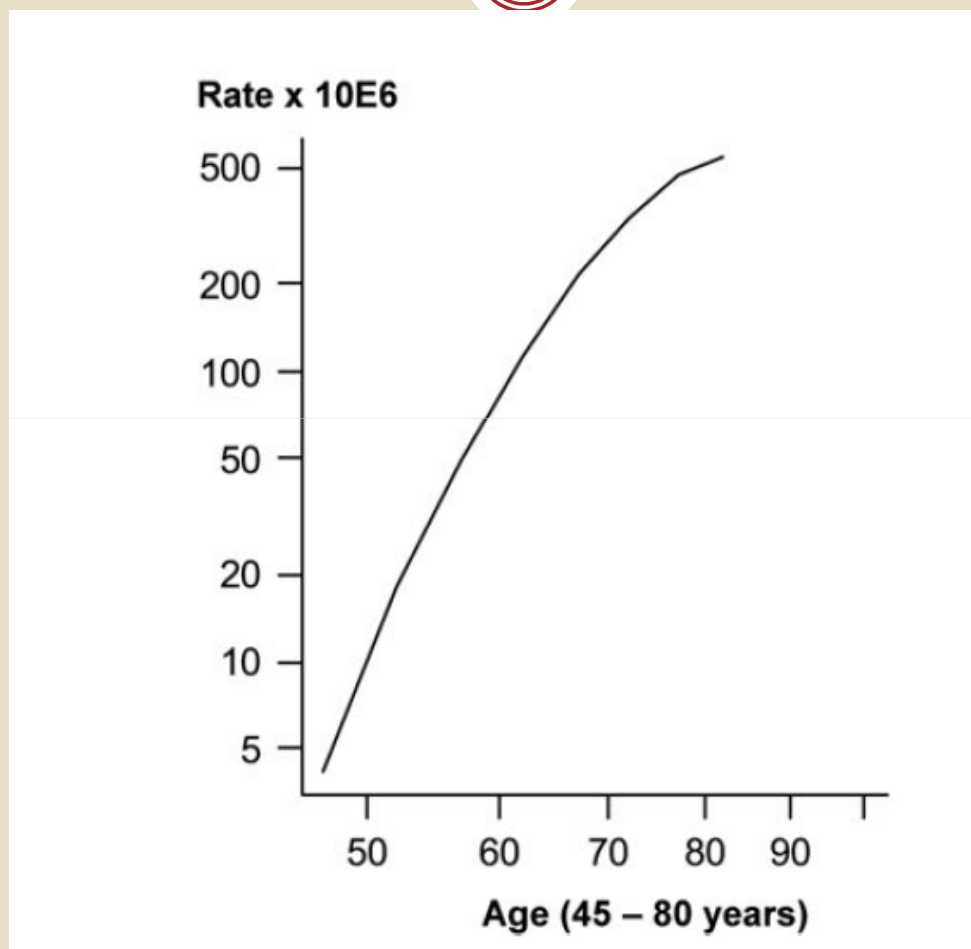
Fuente: Larrañaga et al. Ann Oncol 2010

Cáncer de próstata en España



Fuente: Larrañaga et al. Ann Oncol 2010

Cáncer de próstata en España



Fuente: Larrañaga et al. Ann Oncol 2010

Tabaquismo en España

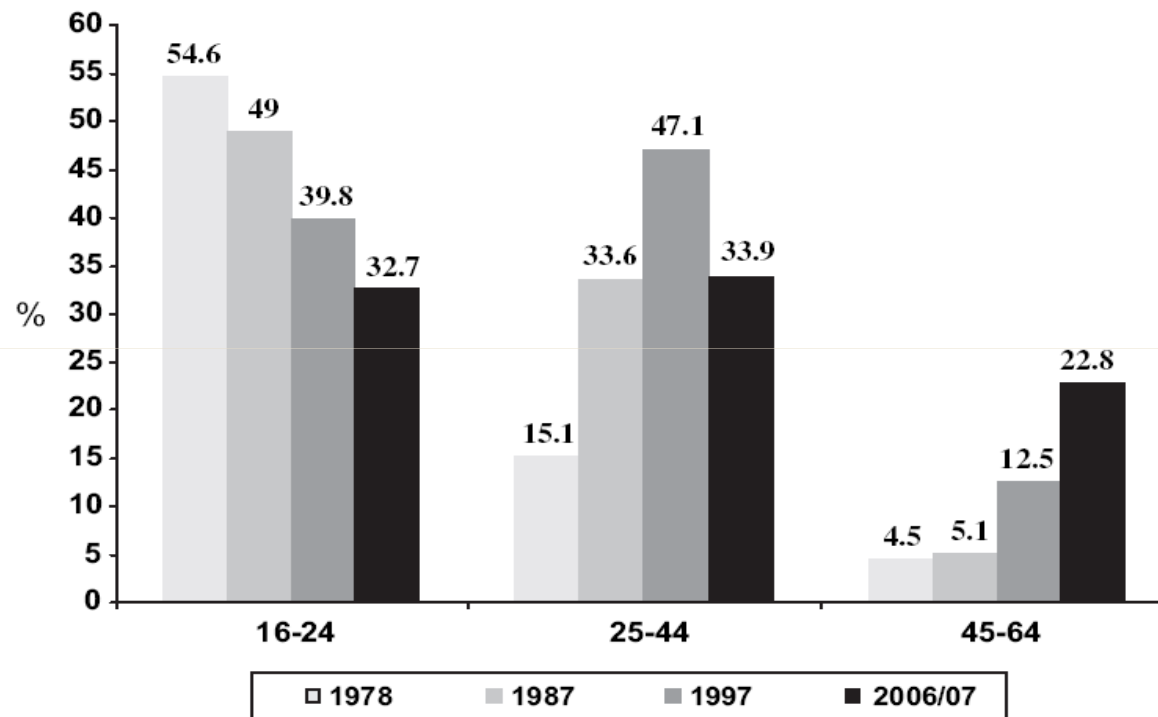


Figure 1. Prevalence (%) of female smokers, according to age: Spain, 1978–2006/07.

Fuente: Regidor et al. Ann Oncol 2010

Inactividad física en España

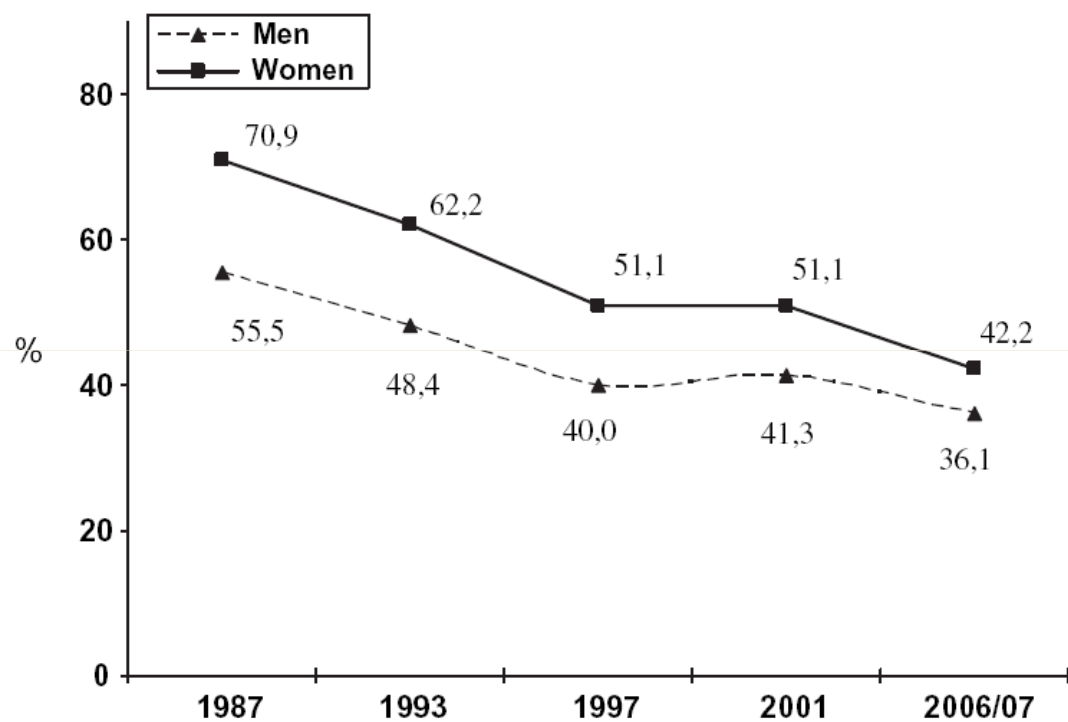
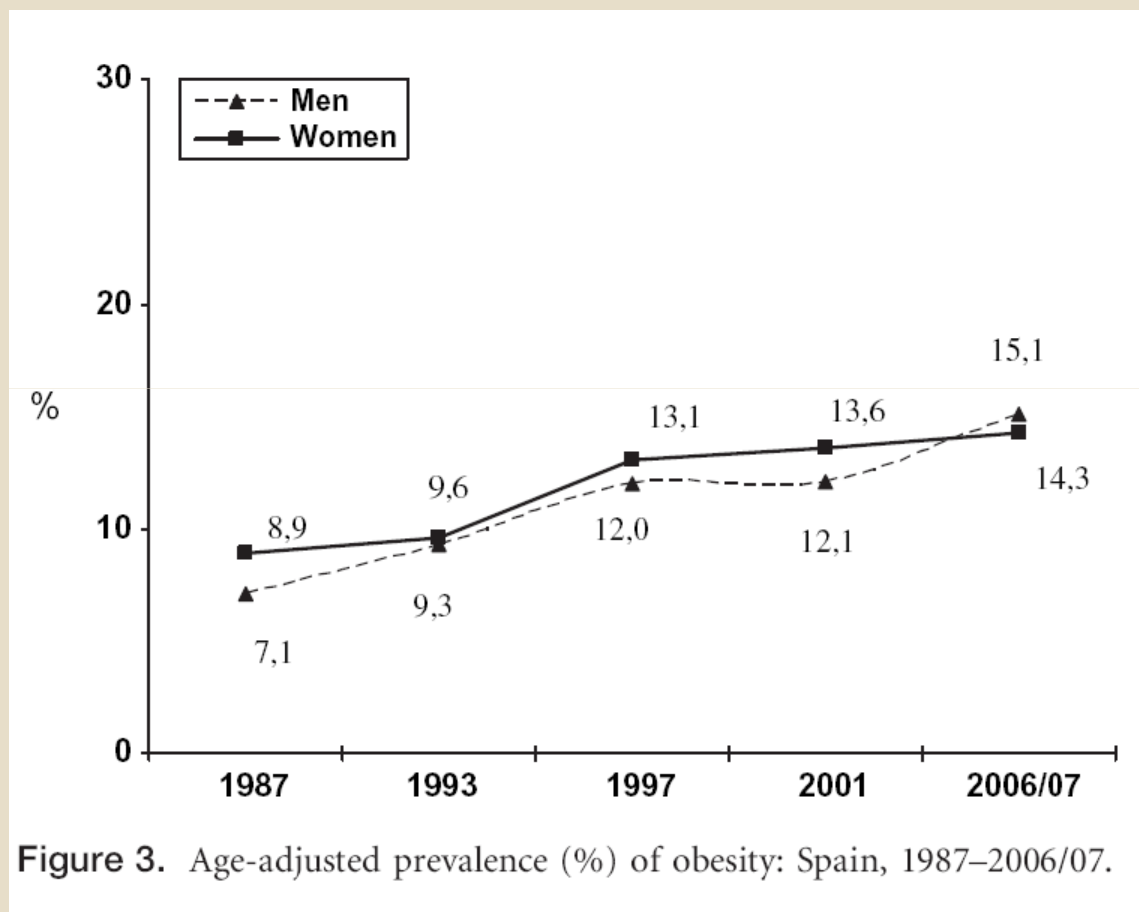


Figure 2. Age-adjusted prevalence (%) of physical inactivity: Spain, 1987–2006/07.

Fuente: Regidor et al. Ann Oncol 2010

Obesidad en España



Fuente: Regidor et al. Ann Oncol 2010

Bebedores de riesgo en España

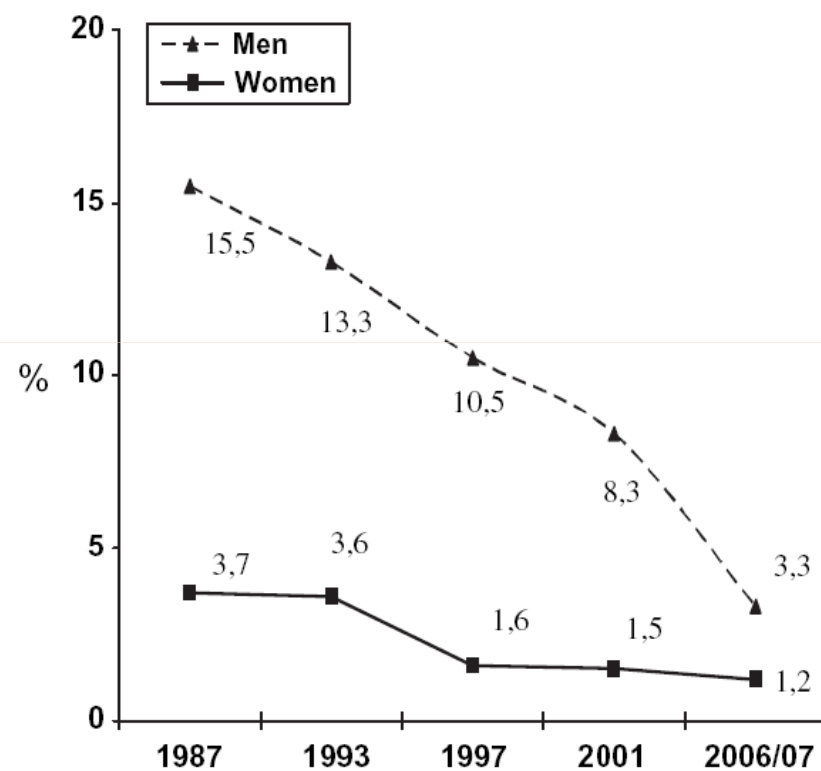
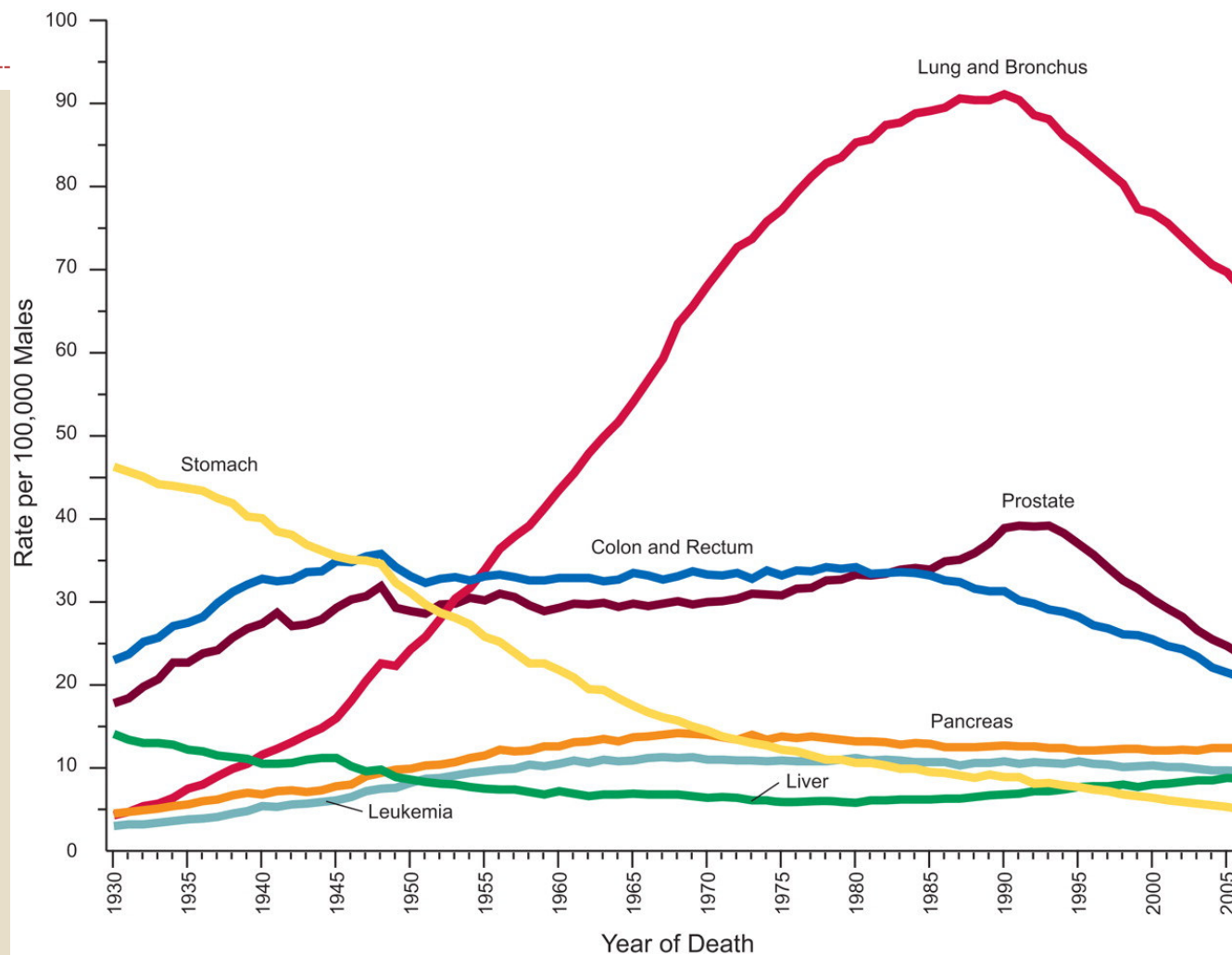


Figure 4. Age-adjusted prevalence (%) of risk drinkers: Spain, 1987–2006/07.

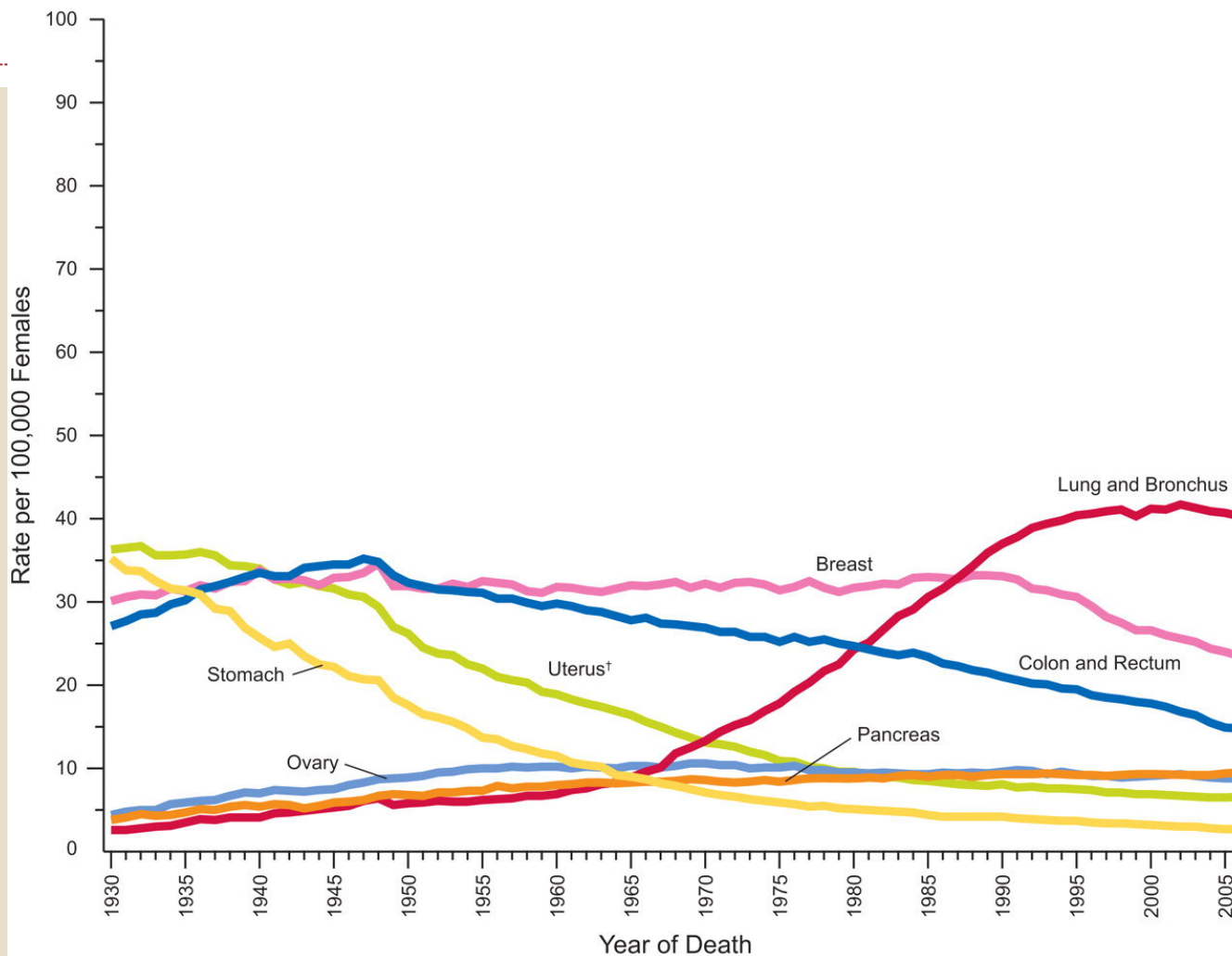
Fuente: Regidor et al. Ann Oncol 2010

FIGURE 4 Annual Age-Adjusted Cancer Death Rates* Among Males for Selected Cancers, United States, 1930 to 2006



From Jemal, A. et al.
CA Cancer J Clin 2010;0:caac.20073v1

FIGURE 5 Annual Age-Adjusted Cancer Death Rates* Among Females for Selected Cancers, United States, 1930 to 2006



From Jemal, A. et al.
CA Cancer J Clin 2010;0:caac.20073v1

FIGURA 4.1. MORTALIDAD GLOBAL POR CÁNCER EN LA UNIÓN EUROPEA 1998: TASAS AJUSTADAS POR LA POBLACIÓN ESTÁNDAR EUROPEA (POR 100.000)

HOMBRES



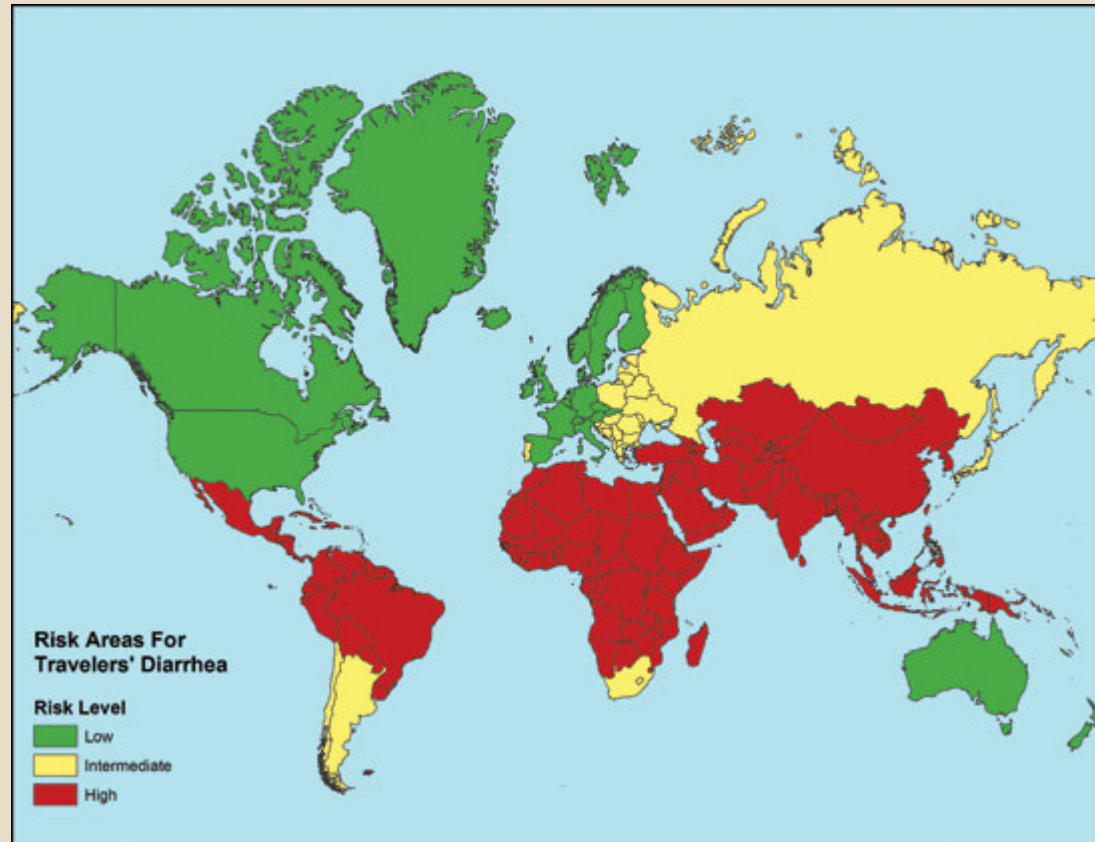
< 208,3 < 227,3 < 246,2 < 265,1 < 284,0

MUJERES

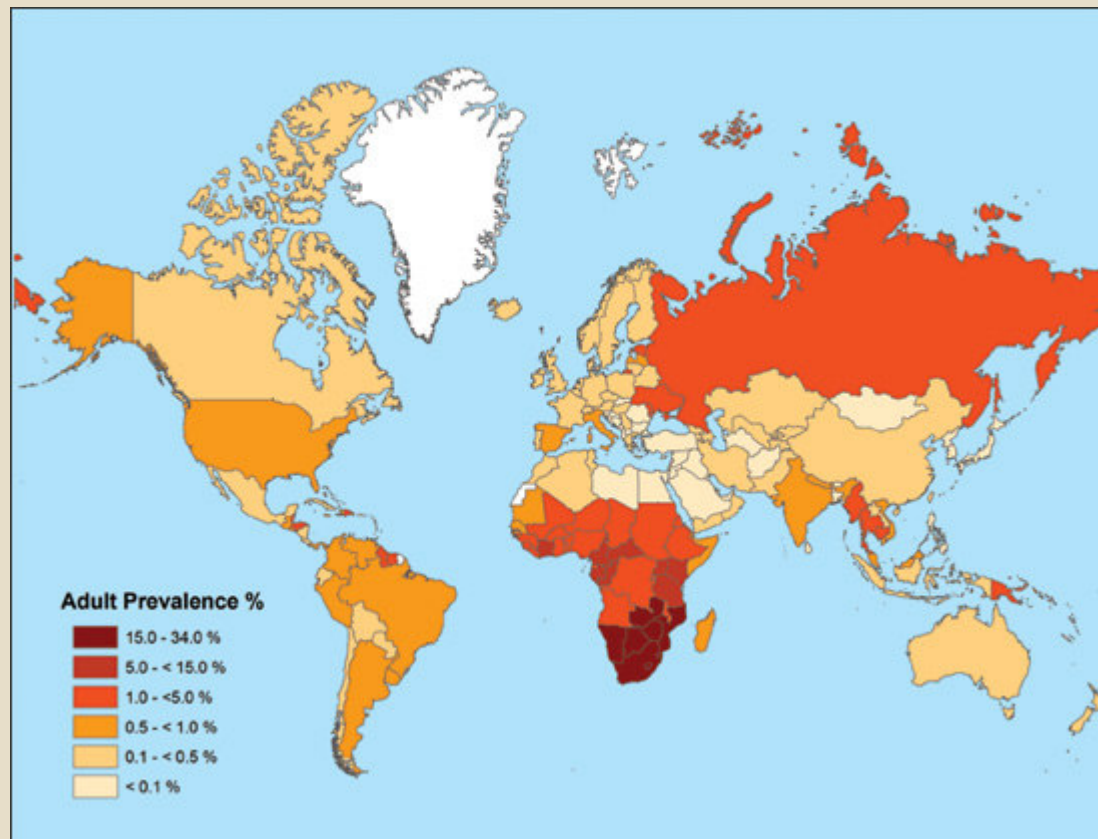


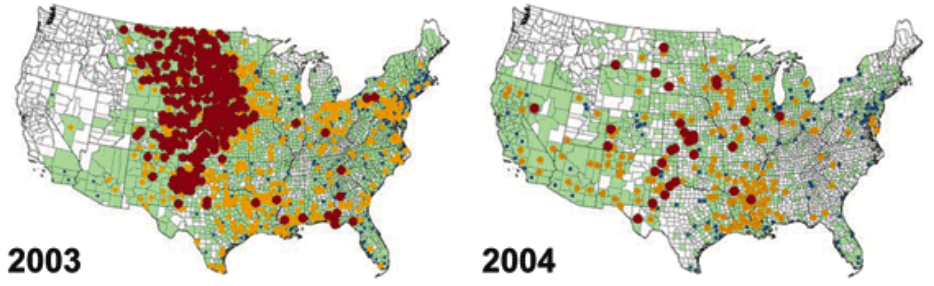
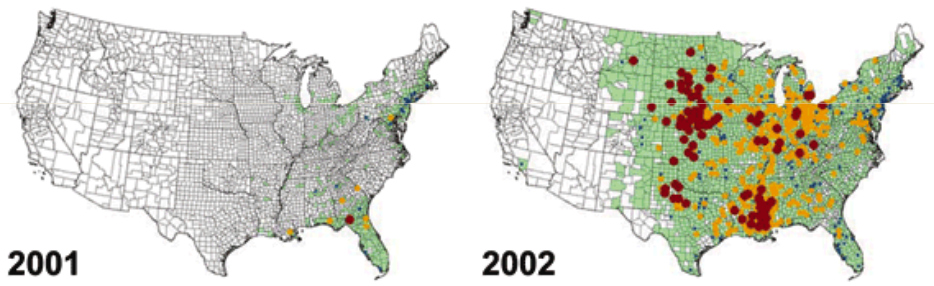
< 129,5 < 146,6 < 163,7 < 180,7 < 197,8

Fuente: EUCAN 1998 [Ferlay 1999].



Riesgo de infección por VIH





Incidence per million

● .01-9.99 ● 10-99.99 ● ≥ 100 ■ Any WNV Activity

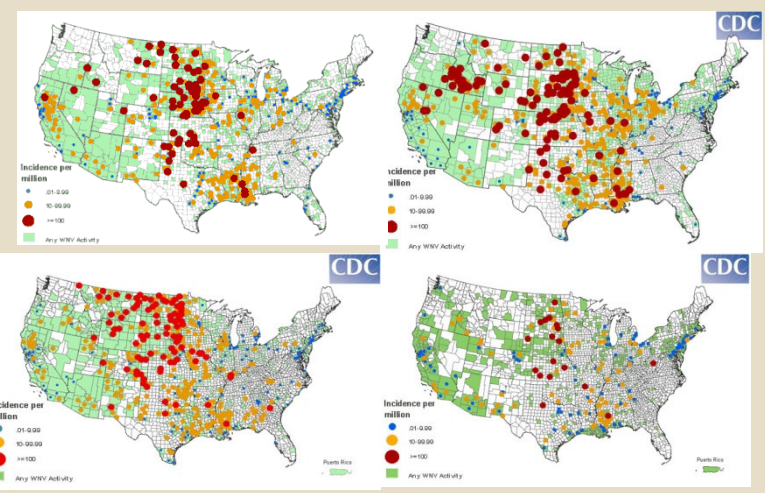




Figure 5.1 Probable cases of SARS worldwide, 7 August 2003

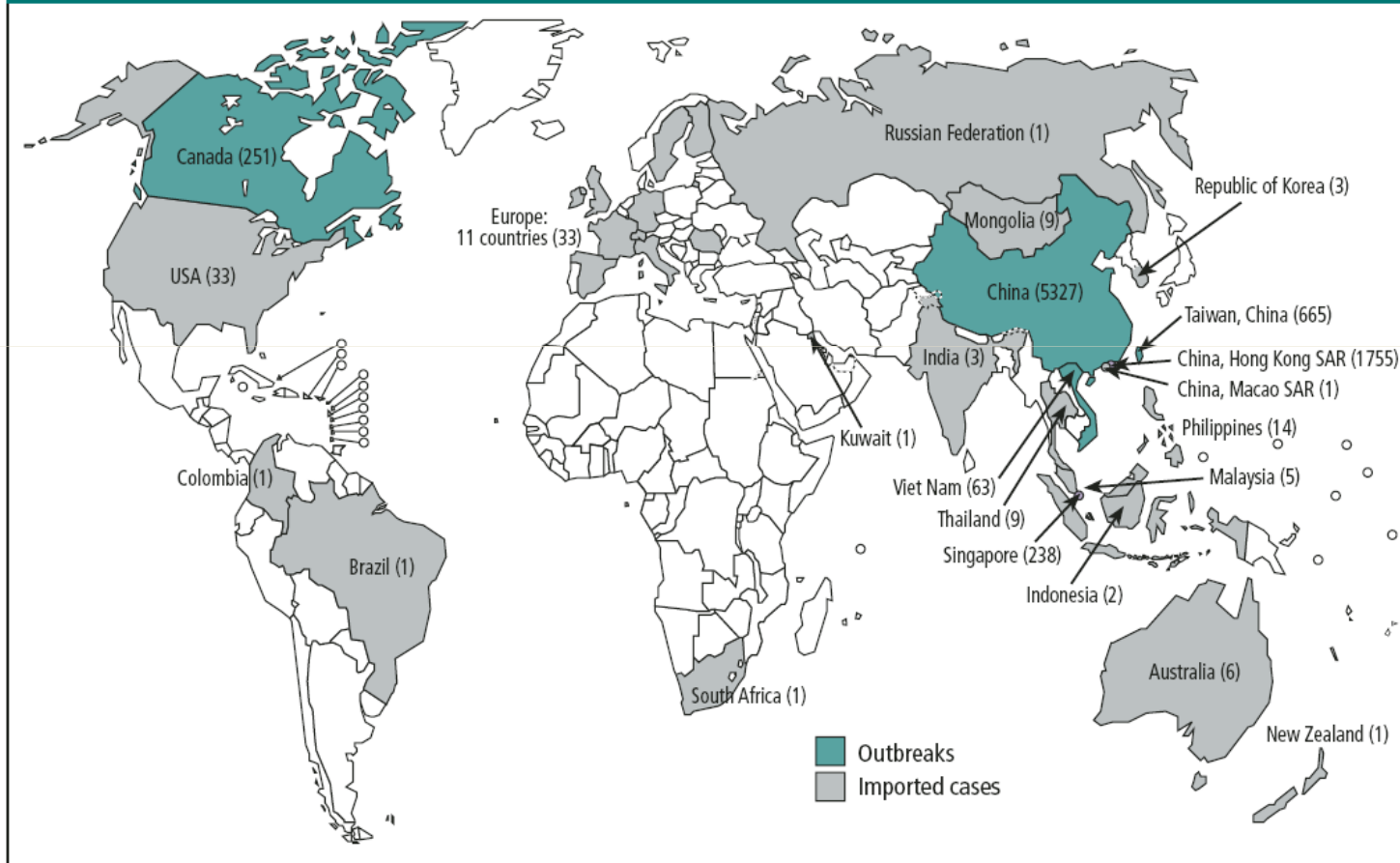
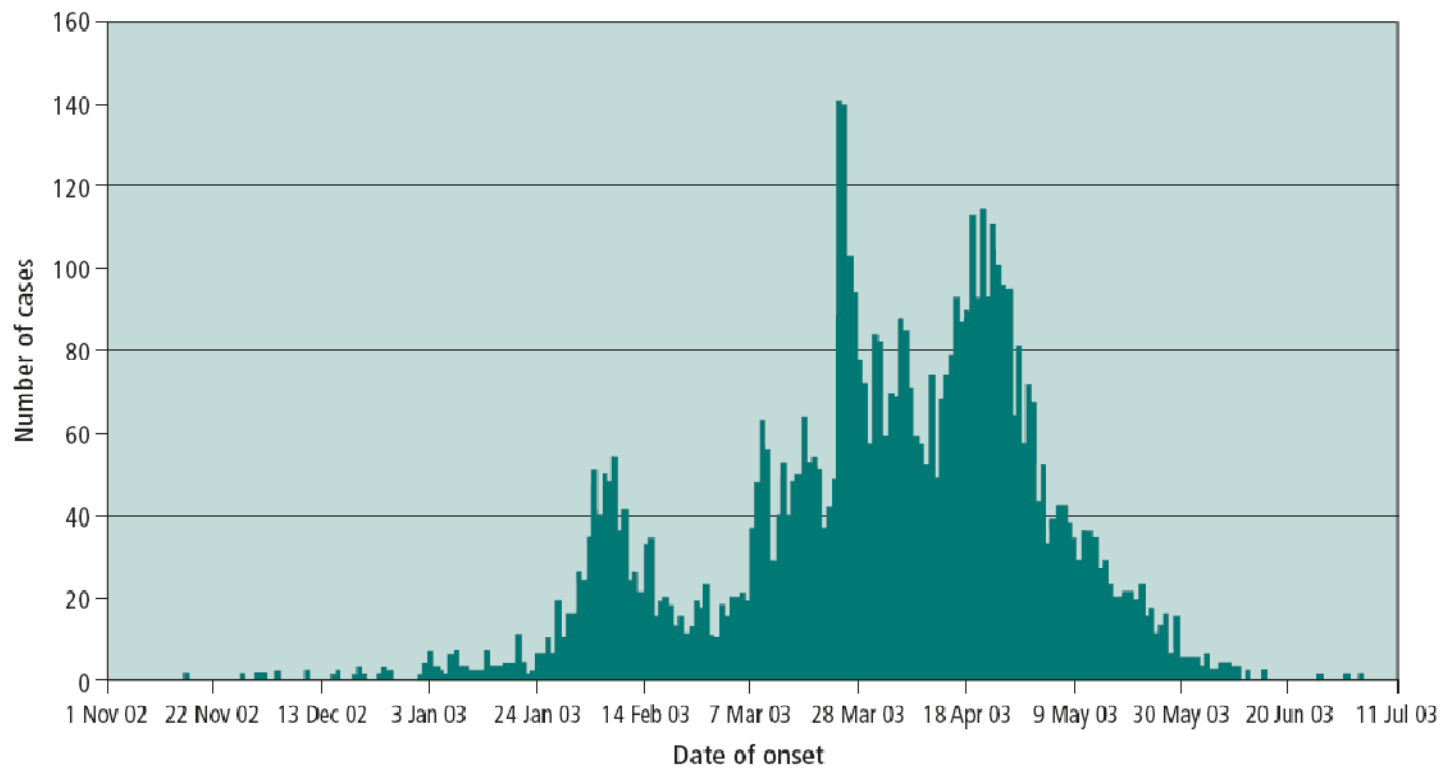


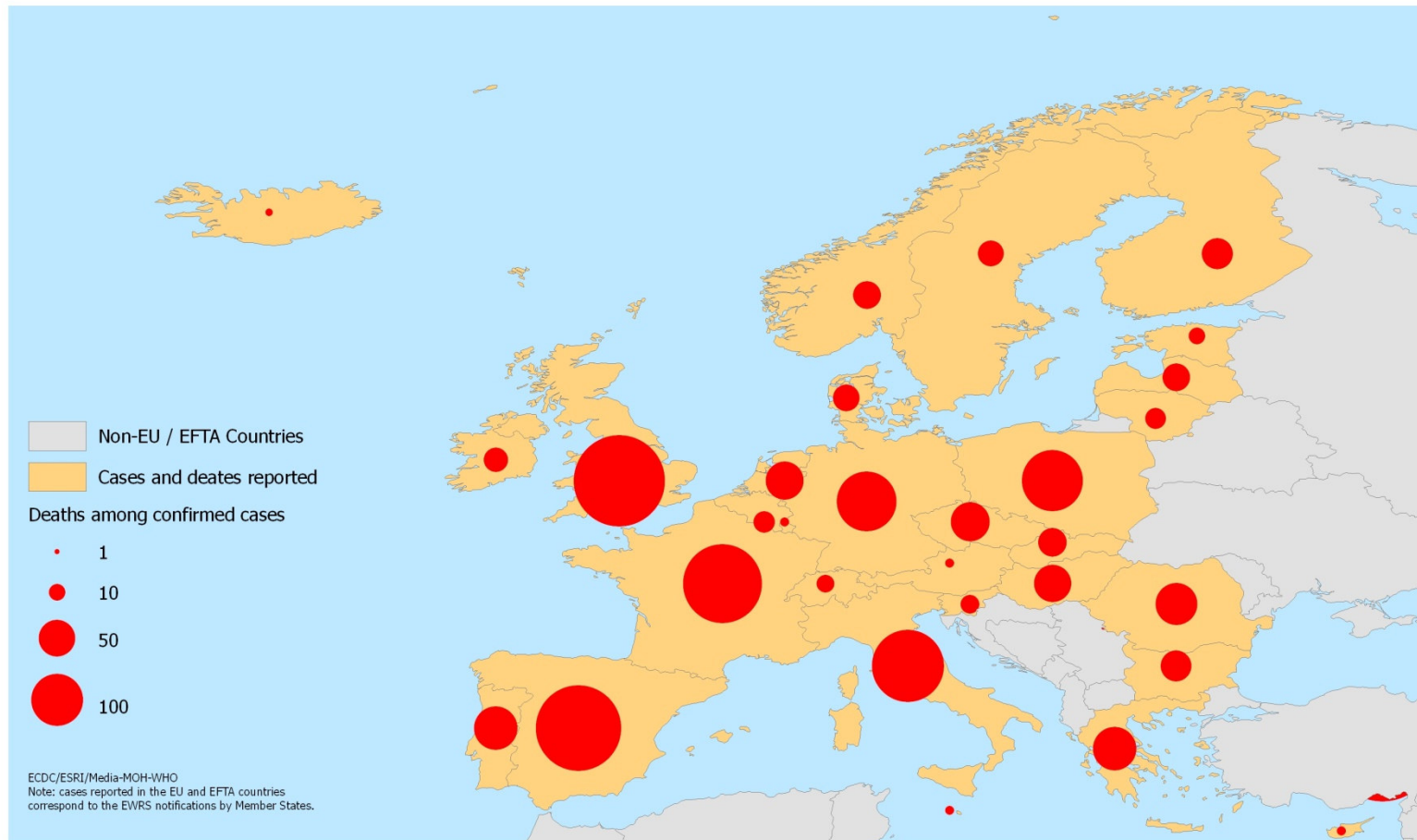


Figure 5.2 Probable cases of SARS worldwide,^a 1 November 2002–11 July 2003



^aThis graph does not include 2527 probable cases of SARS (2521 from Beijing, China), for whom no dates of onset are currently available.

Reported cumulative number of confirmed fatal cases of 2009 pandemic influenza A(H1N1) in EU and EFTA countries, as of 05 January 2010, 14.00 CEST



Número de muertes semanales en la UE

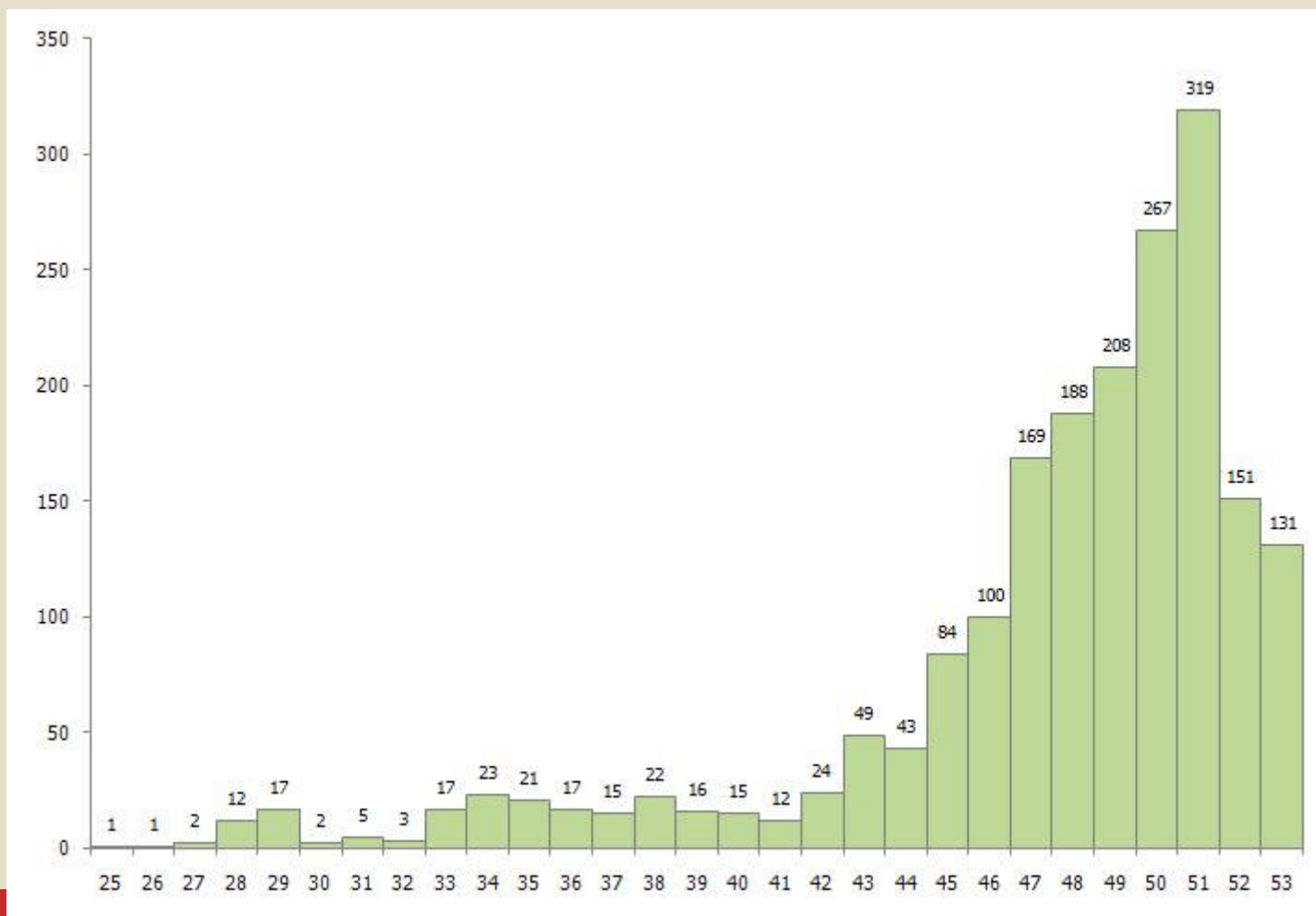
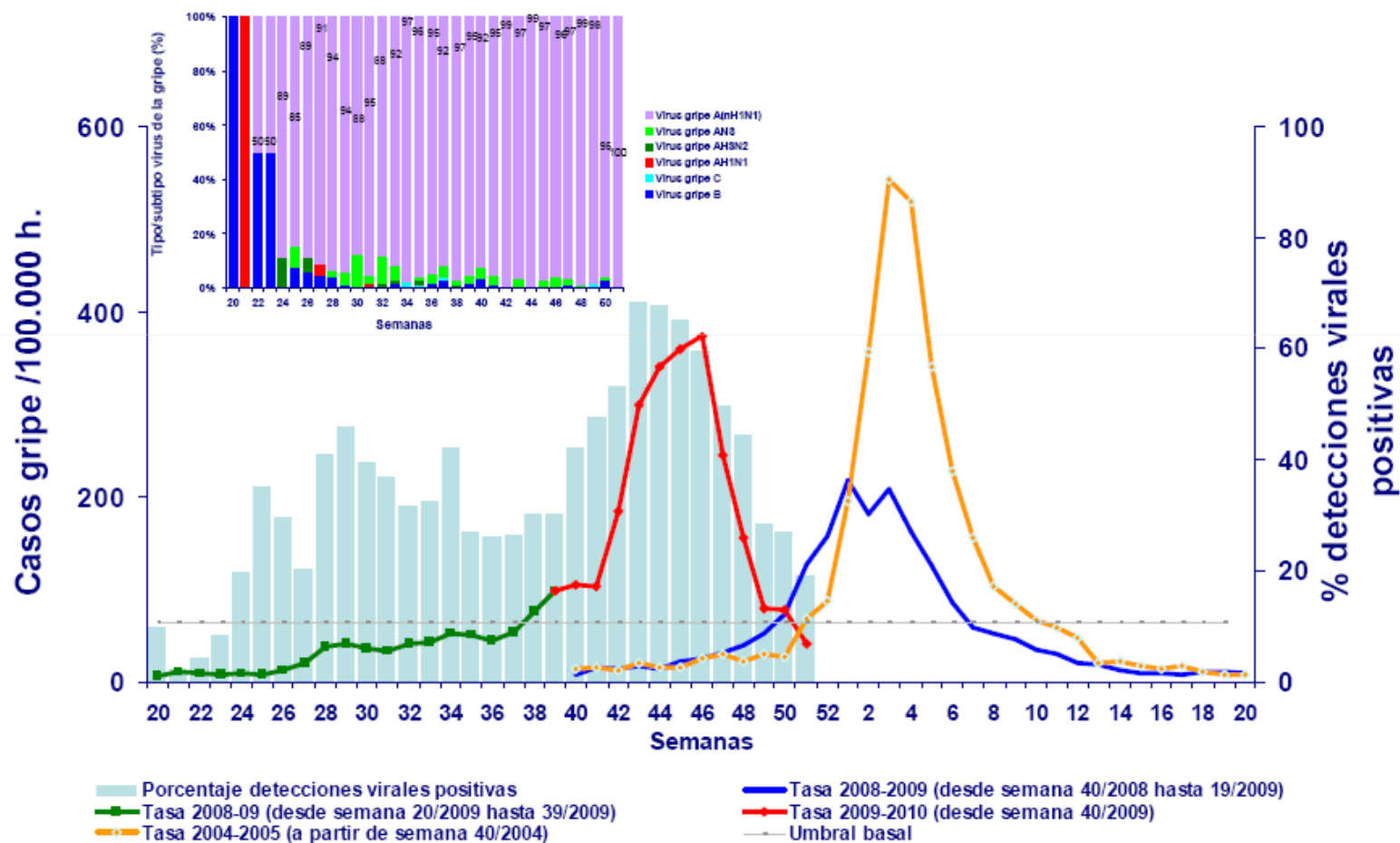
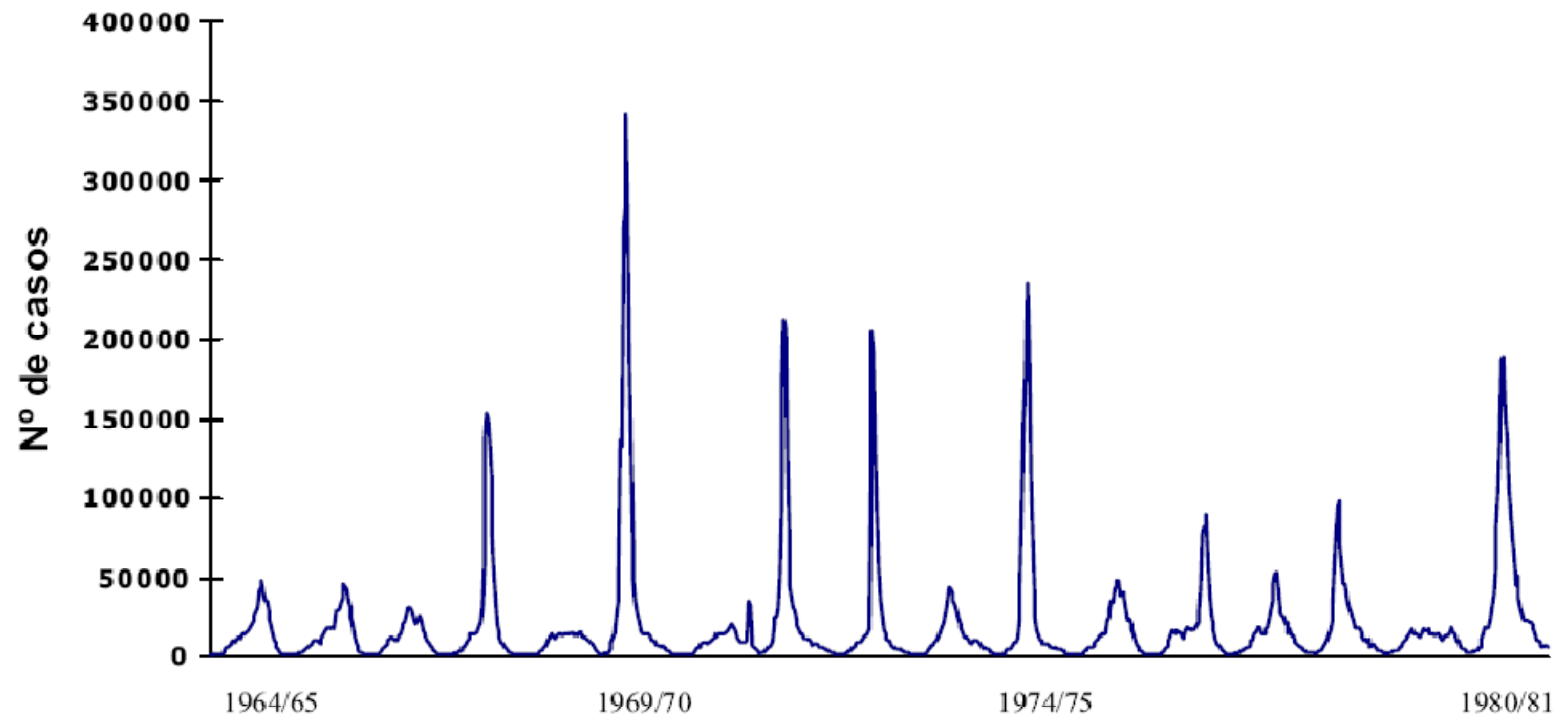


Figura 1. Tasa de incidencia semanal de gripe y porcentaje de detecciones virales positivas.
Semanas 20-51/2009. Sistemas centinela. España.

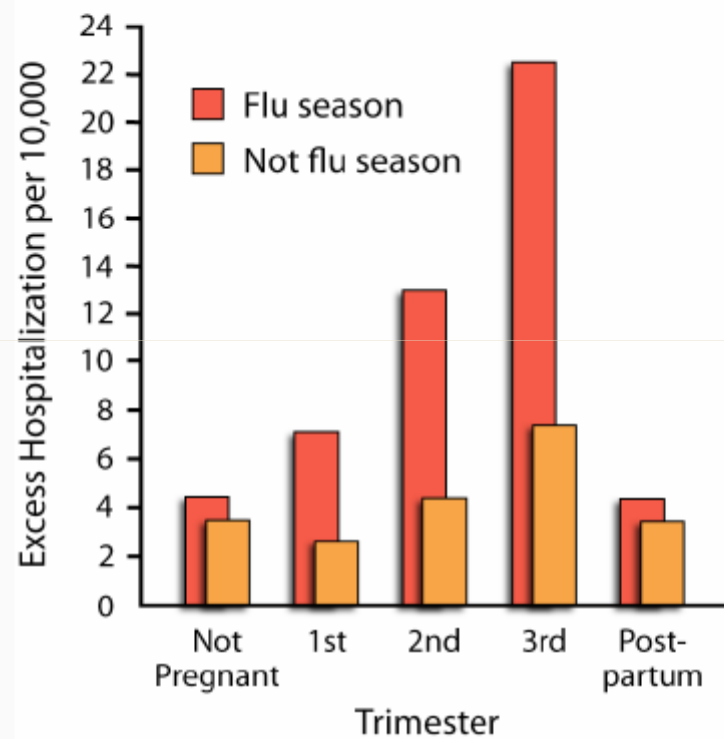


**Figura 6. Vigilancia de la Gripe. España.
Temporadas 1964-1981.
Casos de gripe declarados por semanas.**





Excess Hospitalization of Low Risk Women During Influenza and Non-Influenza Season



Adapted by CTLT from Neuzil, et al.



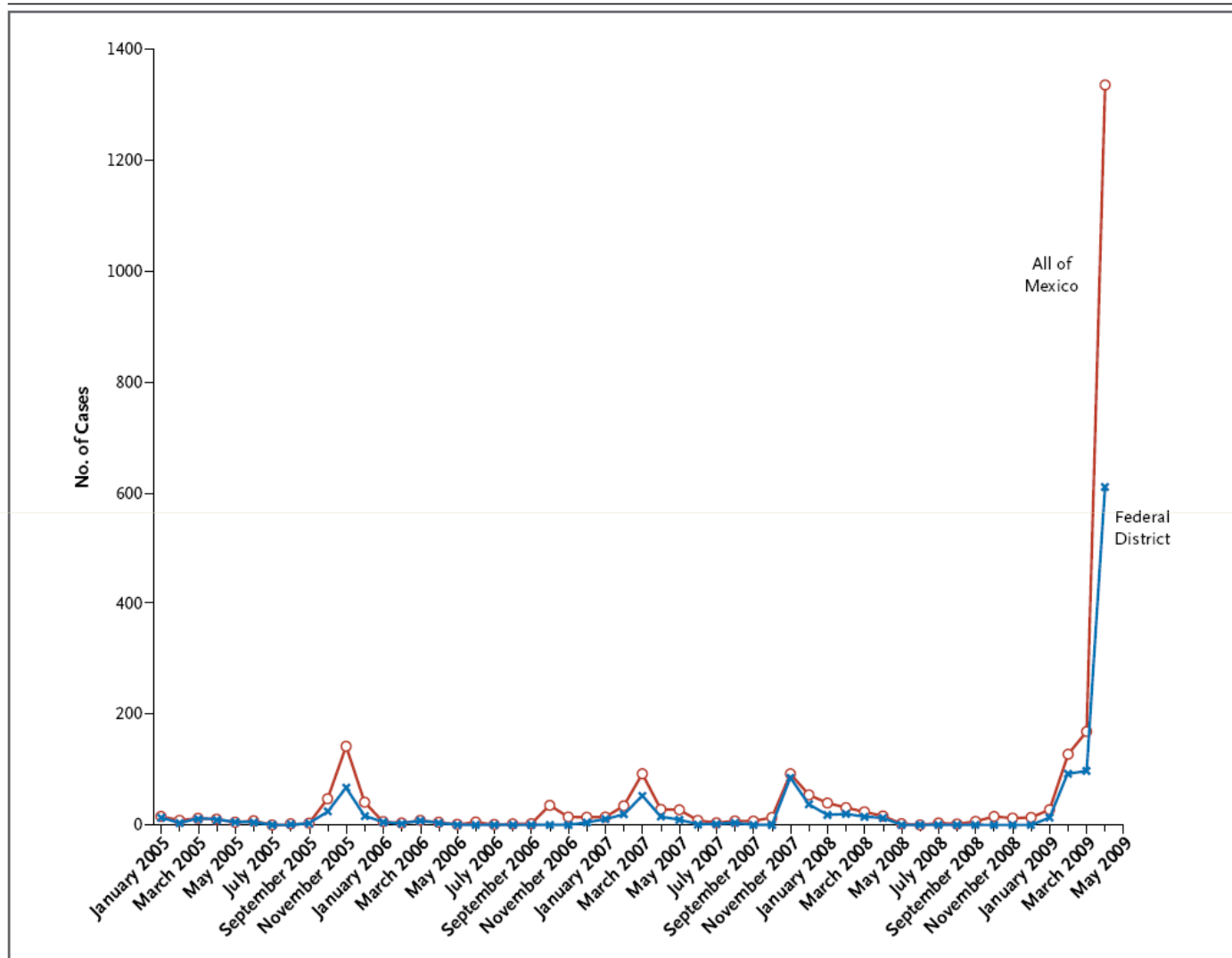
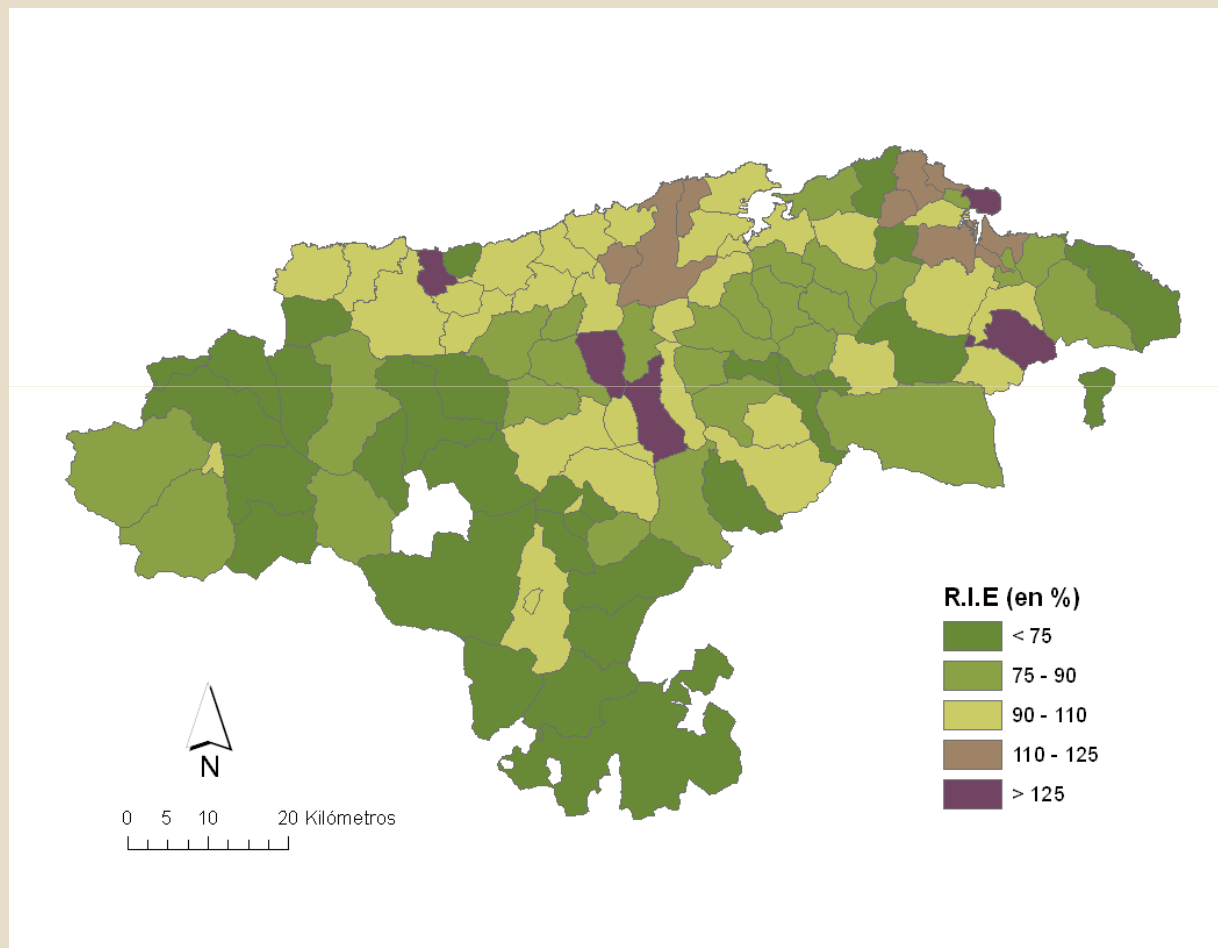


Figure 1. Time Series of Confirmed Cases of Influenza in the Federal District and in All of Mexico, According to Month (January 2005–April 2009).

Data are from the National Epidemiological Reference Laboratory.

Incidencia de cáncer en Cantabria



Medidas de una distribución



- Medidas de tendencia central
- Medidas de dispersión
- Medidas de posición
- Medidas de forma

Medidas de una distribución



- Medidas en la población: letras griegas
 - μ , σ^2 , σ
- Medidas en la muestra: letras latinas
 - m , s^2 , s

Medidas de tendencia central



- Media

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

- Media de 10, 15, 20, 35, 100

$$\mu = \frac{10 + 15 + 20 + 35 + 100}{5} = 36$$

Medidas de tendencia central



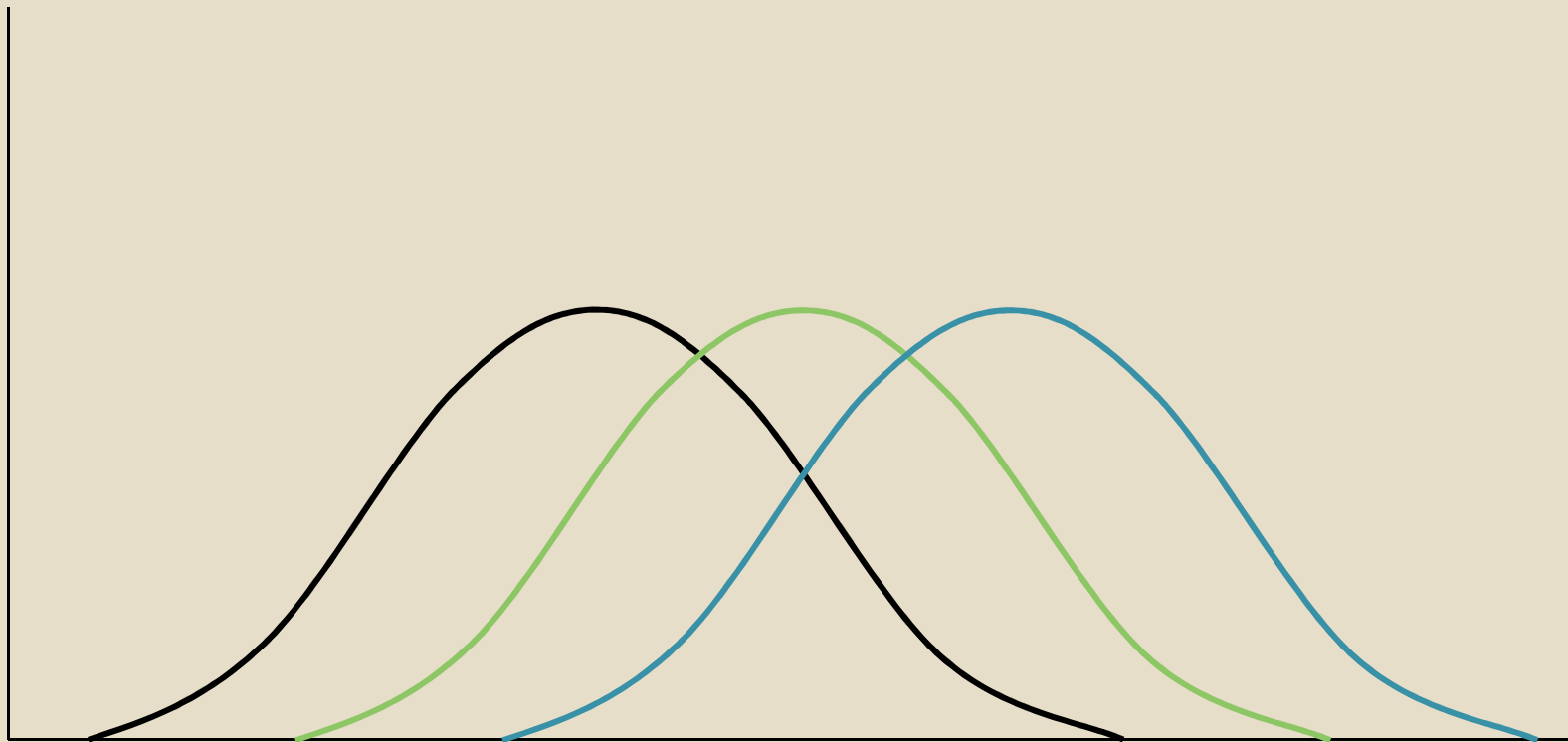
- Media de 10, 15, 20, 35, 100

$$\mu = \frac{10 + 15 + 20 + 35 + 100}{5} = 36$$

- Media de 10, 15, 20, 35 **Sensible a valores extremos**

$$\mu = \frac{10 + 15 + 20 + 35}{4} = 20$$

Distribuciones con diferente media



Medidas de tendencia central



- Mediana:
 - Valor central de una distribución

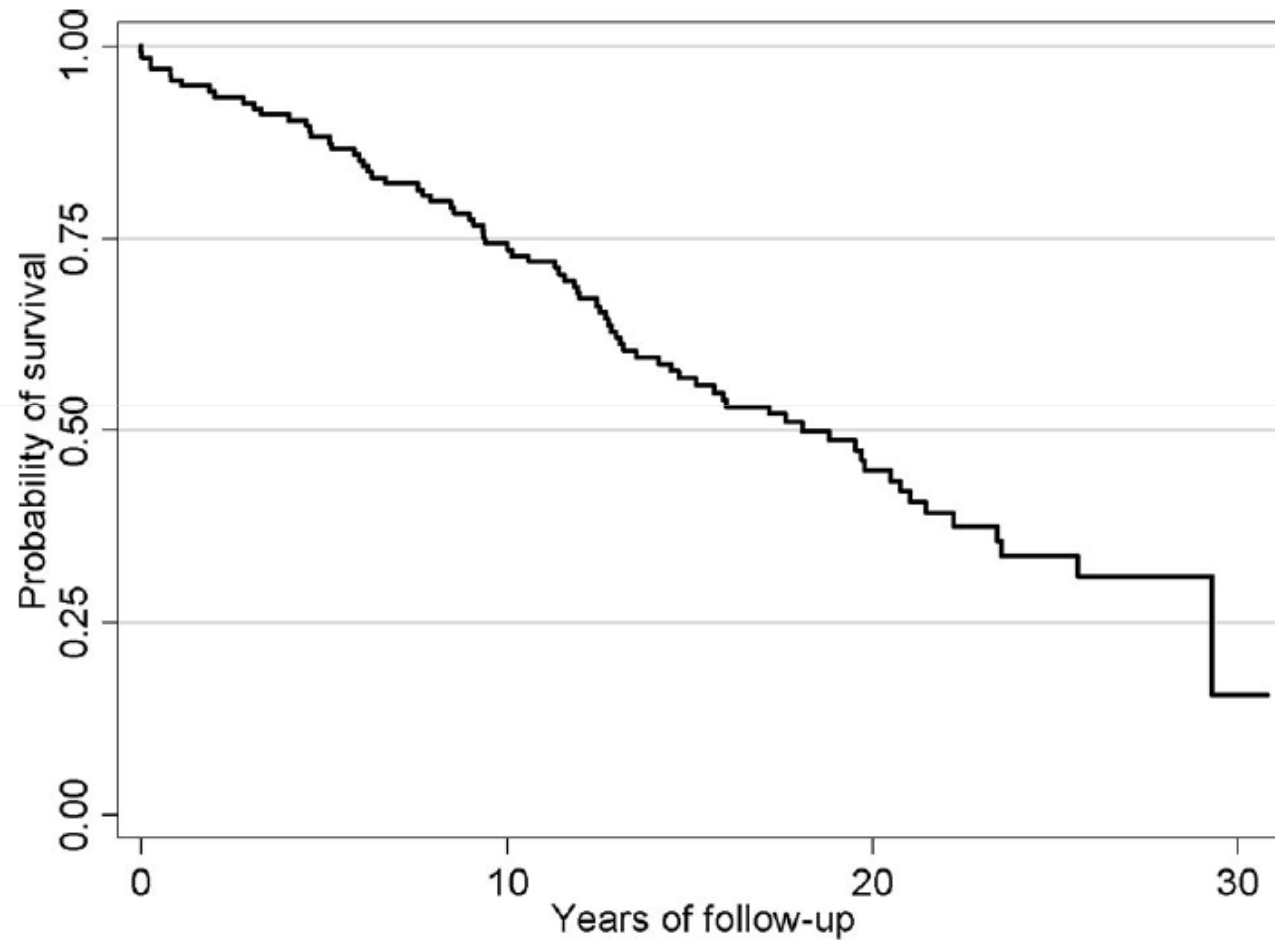
- Mediana de 10, 15, 20, 35, 100

- Mediana de 10, 15, 20, 35

17,5

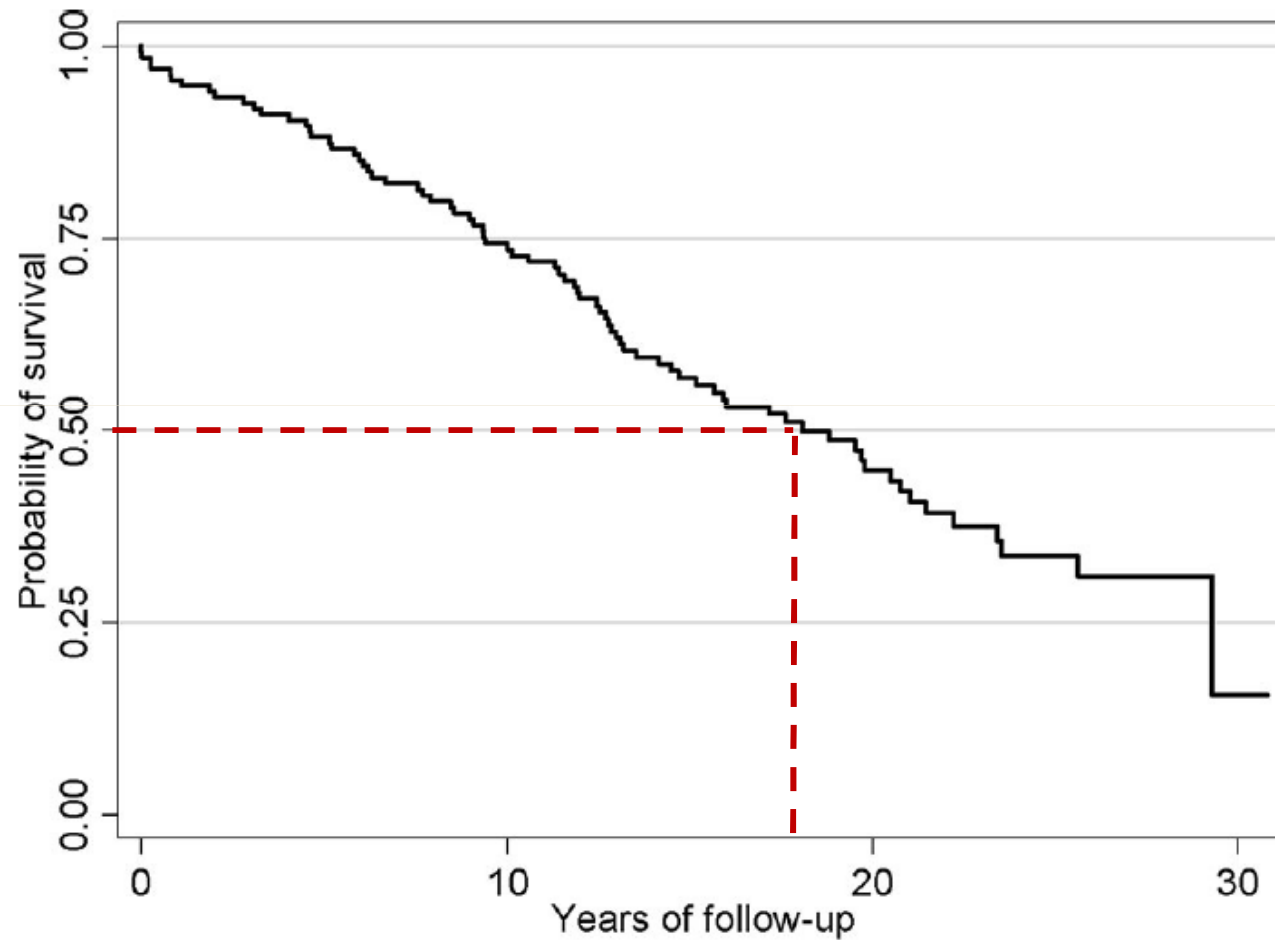
Poco sensible a valores extremos

Medidas de tendencia central



Bernal et al, 2010

Medidas de tendencia central



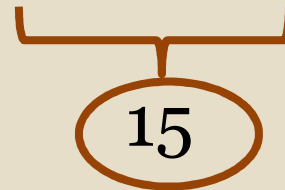
Bernal et al, 2010

Medidas de tendencia central



- Moda
 - El valor más frecuente

- Moda de 10, 15, 15, 15, 20, 20, 35, 100

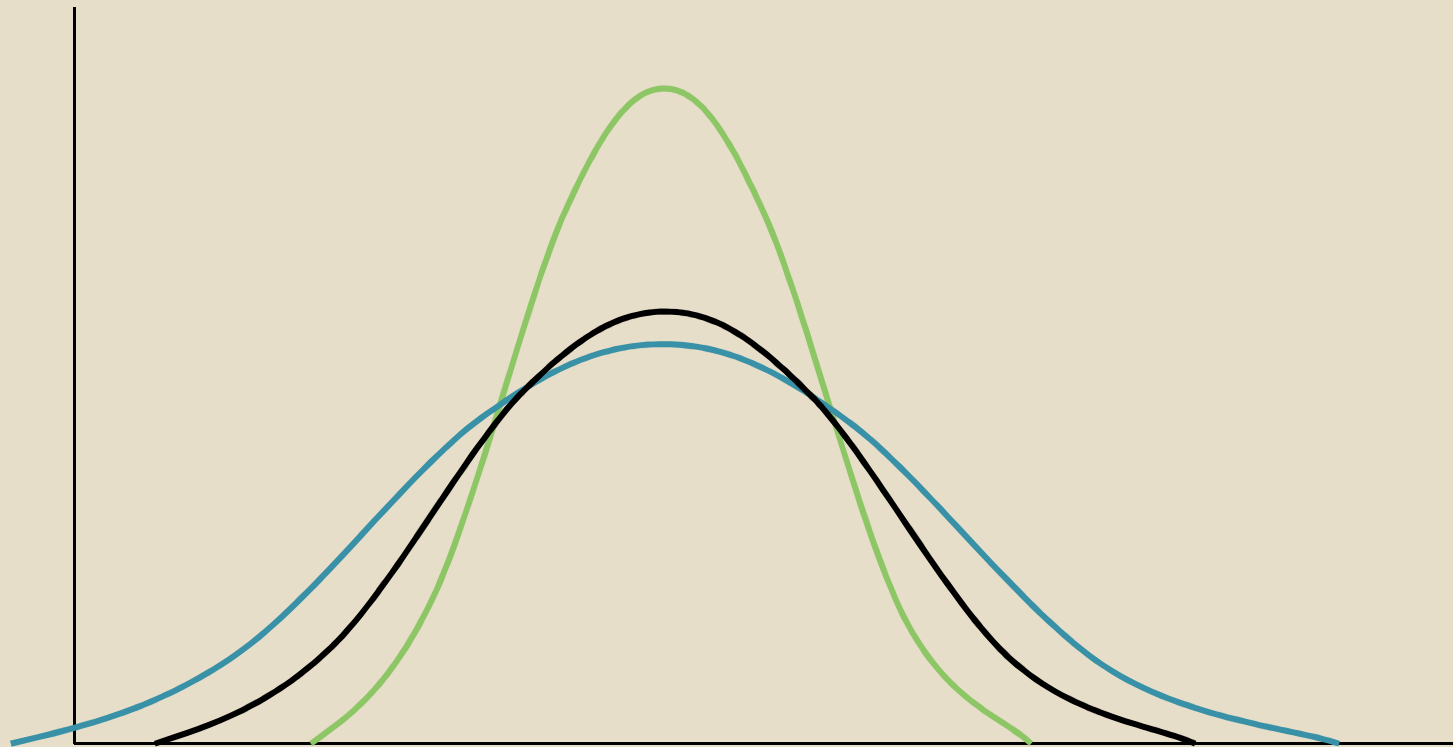


Poco sensible a valores extremos

- Moda de 10, 15, 15, 15, 20, 20, 35



Medidas de dispersión



Medidas de dispersión



- Varianza

- En la población

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n}$$

- En la muestra

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

Medidas de dispersión



- Varianza de 10, 15, 20, 35, 100

$$\sigma^2 = \frac{(10-36)^2 + (15-36)^2 + (20-36)^2 + (35-36)^2 + (100-36)^2}{5} = 1367,5$$

Medidas de dispersión



- Desviación típica (desviación estándar)
 - En la población

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$$

- En una muestra

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

Medidas de dispersión



- Desviación típica de 10, 15, 20, 35, 100

$$\sigma = \sqrt{\frac{(10-36)^2 + (15-36)^2 + (20-36)^2 + (35-36)^2 + (100-36)^2}{5}} = 36,98$$

Medidas de dispersión



- Coeficiente de variación

$$CV = \frac{\sigma}{\mu}$$

$$CV = \frac{s}{m}$$

Medidas de posición



- Percentiles
 - P10: Valor que deja por debajo el 10% de la distribución.
 - Análogamente: P1, P2,..., P99

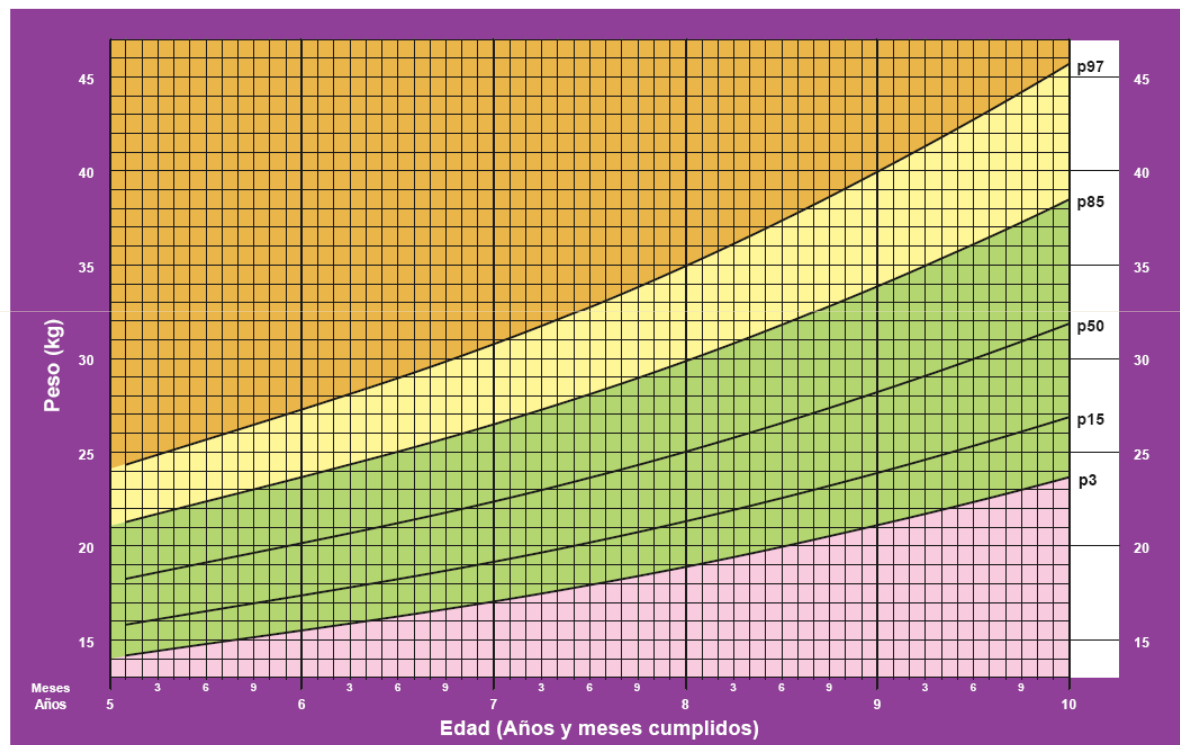
Medidas de posición



- **Percentiles**

Peso para la edad - NIÑAS

Patrones de crecimiento de la OMS 2007 - 5 a 10 años (percentiles)



Nota: Este patrón describe el crecimiento normal de un niño en un ambiente óptimo desde los 5 hasta los 10 años y puede aplicarse a todos los niños en cualquier lugar del mundo, independientemente de su etnia, estatus socioeconómico y tipo de alimentación. Las curvas se basan en el patrón publicado por OMS en el año 2007. Para mayor información visite el sitio oficial de la OMS en <http://www.who.int/childgrowth/en/> - Puede descargar una versión para imprimir en formato PDF en la dirección: <http://www.saludecitura.com/tomuanios/>

Sobrepeso **Riesgo de sobrepeso** **Normal** **Bajo peso**

Medidas de posición



- Deciles
 - P10, P20, P30, ..., P90.
- Quintiles
 - P20, P40, P60, P80
- Cuartiles
 - P25, P50, P75

Medidas de posición



- Rango
 - Observación menor - Observación mayor
- Rango intercuartílico
 - P25 - P75

Diagrama de caja (“Box and whiskers”)

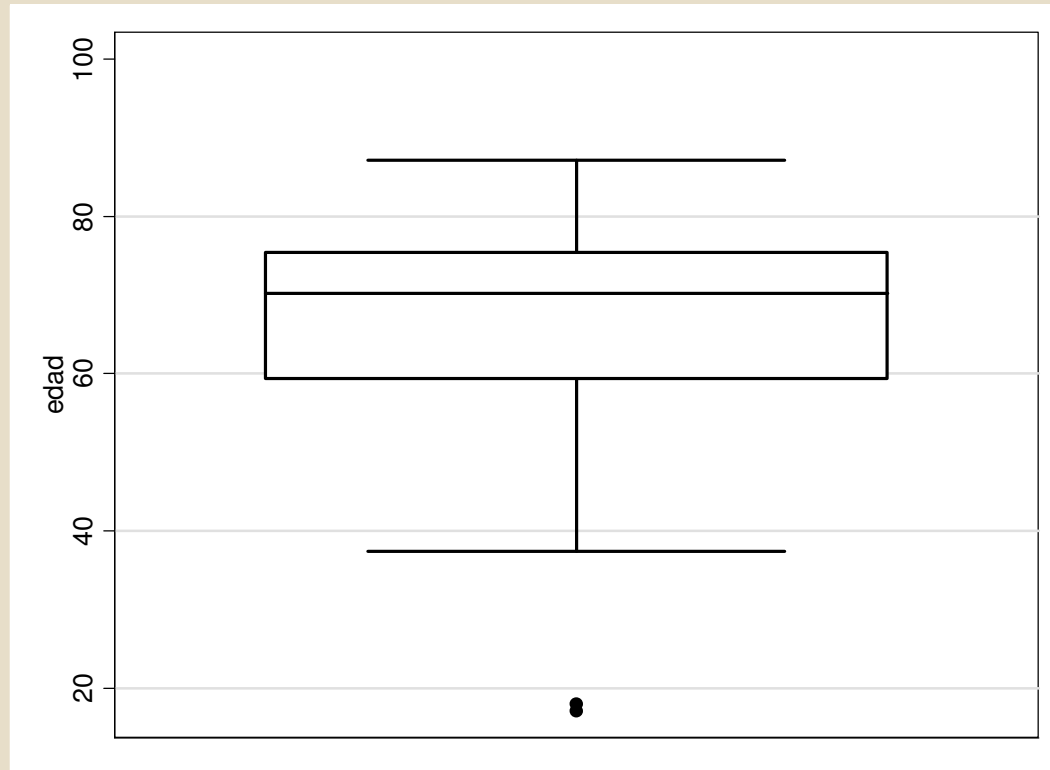
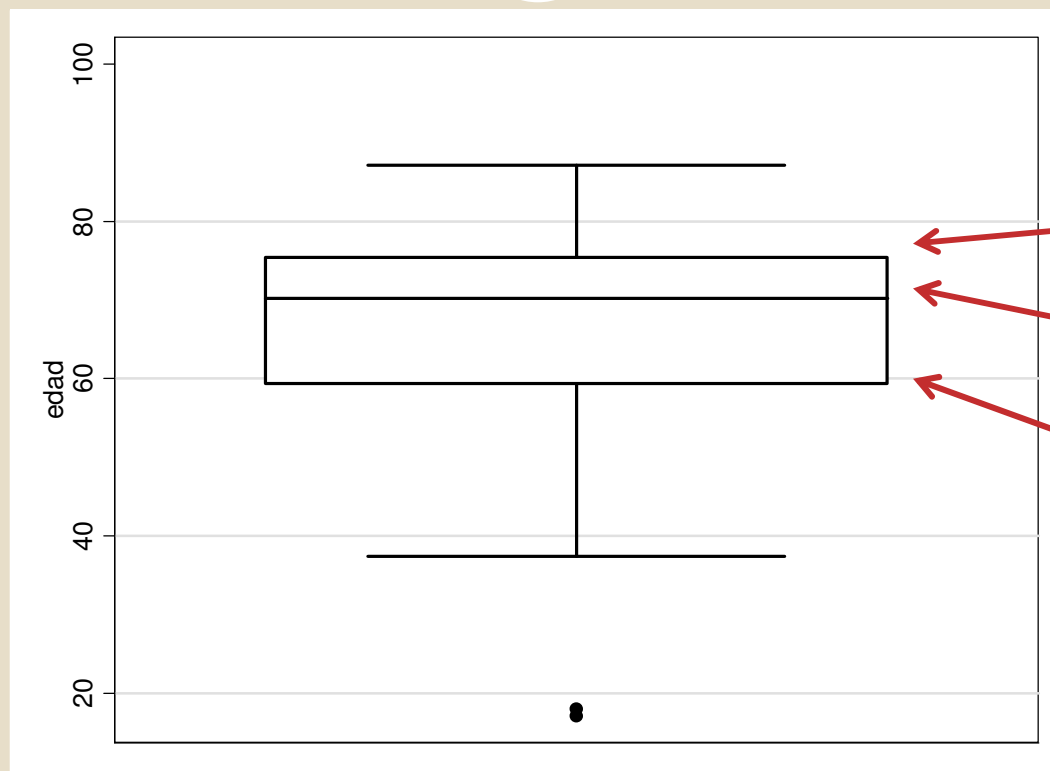


Diagrama de caja (“Box and whiskers”)



Cuartil 3

Mediana

Cuartil 1

Diagrama de caja (“Box and whiskers”)

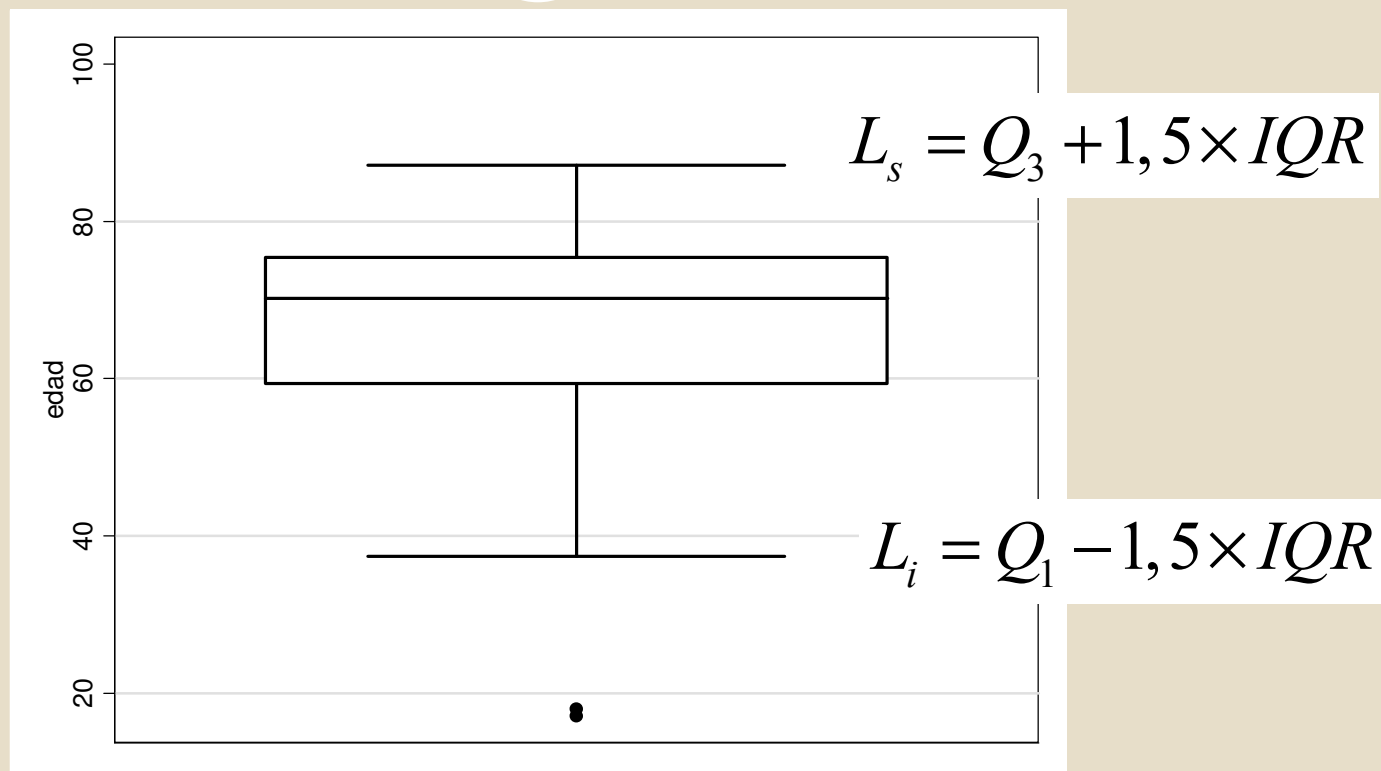
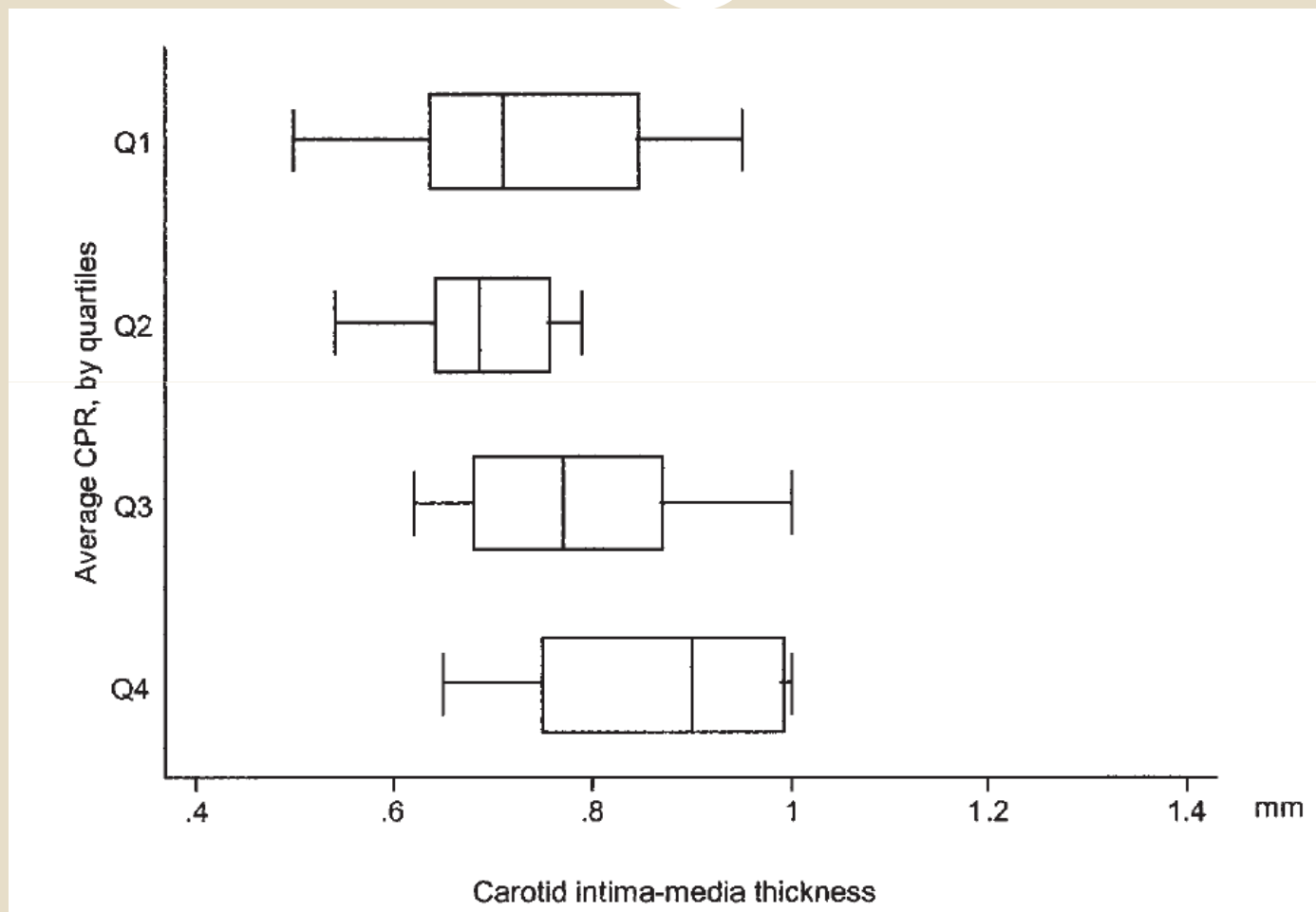


Diagrama de caja (“Box and whiskers”)



Gonzalez-Gay et al, 2005

Medidas de forma



- Asimetría

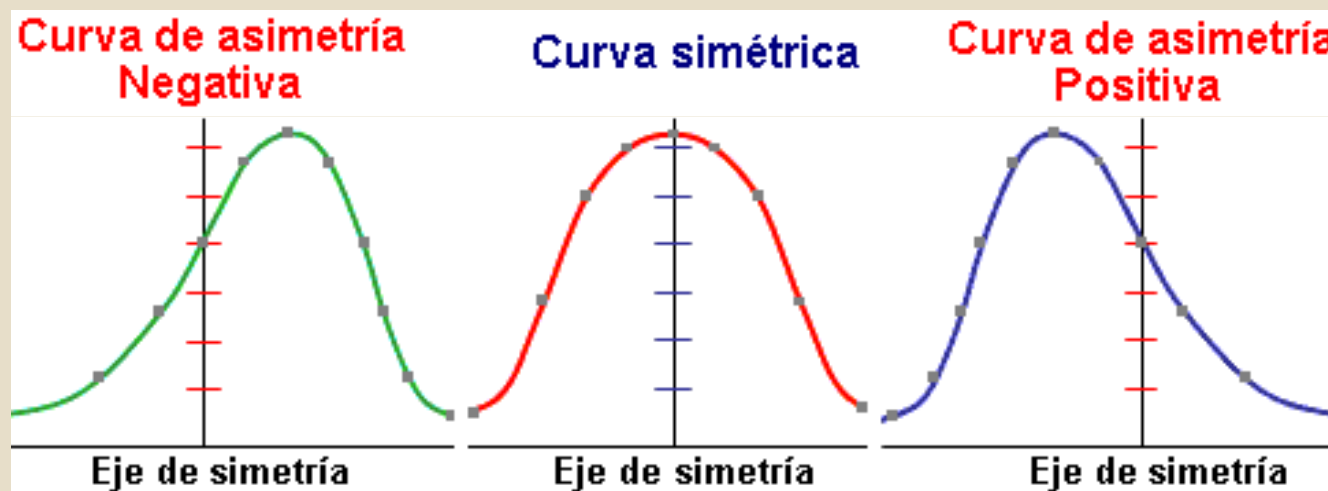
$$\gamma_1 = \frac{\sum_{i=1}^n (x_i - \mu)^3 / n}{\sigma^3}$$

Medidas de forma



- Asimetría

- Figuras con asimetría 0, <0 y >0



Medidas de forma



- Curtosis o apuntamiento

$$g_2 = \frac{\sum_{i=1}^n (x_i - \mu)^4 / n}{\sigma^4} - 3$$

Medidas de forma



- Curtosis
 - Figuras con curtosis 0 , <0 , >0

