Clase 16 de Bioestadística

Cuantiles teóricos y empíricos
Boxplot. QQ-plot

Ernesto Mordecki

CMAT, Facultad de Ciencias, Universidad de la República.

Uruguay

Contenidos de la clase

Cuantiles

Empezamos con un ejemplo:

- Nació Gervasio, y va a la consulta de los 6 meses
- El pediatra le dice a los padres: Gervasio está creciendo muy bien, pesa 8 kilos 200 gramos, está en el cuantil 55 de peso
- Los padres se miran contentos y sale de la consulta
- ¿cuantil...qué? se preguntan ambos
- El cuantil z de un valor $0 < q < 1$ correspondiente a una variable aleatoria X con distribución F es un valor que verifica:

$$P(X \leq z) = q$$

Aquí q es el dato y z es el cuantil.

- A veces (como en la consulta médica) se utiliza $100 \times q$,

¿Que quiere decir entonces lo que comunicó el pediatra?

- El pediatra supone que el peso de un bebé varón de 6 meses es una variable aleatoria X con una distribución F que el conoce
- Mide el peso de Gervasio y compara con su tabla de valores
- La probabilidad de que Gervasio tenga un peso X menor o igual que 8,200 es 0.55, es decir

$$P(X \leq 8,200) = 0{,}55$$

- Si usamos la función de distribución $F(a) = P(X \leq a)$ entonces tenemos

$$F(8,200) = 0,55$$

- Como la incógnita es el 0,55, se utiliza la función inversa de F, que se nota F^{-1} y escribimos

$$F^{-1}(8,200) = 0,55$$

- Este valor es el cuantil teórico, porque se calcula a partir de la F (distribución)

Cuantiles en R

Recordamos los 4 comandos de R para cada distribución (por ejemplo la normal estándar):

- `dnorm(x, mean = 0, sd = 1)` me da la densidad normal en x
- `pnorm(q, mean = 0, sd = 1)` me da la distribución normal en x
- `rnorm(n, mean = 0, sd = 1)` simula n variables normales estándar
- $\boxed{\texttt{qnorm(q, mean = 0, sd = 1)}}$ me da el cuantil q de la normal estándar
- ¿Cuánto da `qnorm(0.5)`?

Cuantiles importantes

Cuando el q es un valor especial, los cuantiles tienen nombres especiales:

- Cuando $q = 0,5$ el percientil se llama la mediana
- La mediana deja la mitad de la probabilidad de cada lado:
- ¿Cuánto vale la mediana si $Z \sim \mathcal{N}(0,1)$
- ¿Cuánto vale la mediana si $X \sim \mathcal{N}(\mu, \sigma^2)$?
- La mediana es una medida de posición de X, alternativa a la esperanza.

Si partimos a la probabilidad en 4, tenemos

- El cuantil de 0,25 es el primer cuartil, que designamos Q_1
- El cuantil de 0,50 es la mediana, designado Q_2
- El cuantil de 0,75 es el tercer cuartil, designado Q_3

Los cuartiles nos permiten definir una medida de dispersión de X, denominada rango inter-cuartílico, que es el número

$$Q_3 - Q_1 .$$

Aplicación en salud pública

Percentiles o Desvíos Estándar (DE)	Longitud-talla/edad	Peso/edad	Peso/longitud-talla	IMC/edad
> Percentil 99 (>+3 DE)	*Ver nota 1*	*Ver nota 2*	Obesidad	Obesidad
> Percentil 97 (>+ 2 DE)			Sobrepeso	Sobrepeso
> Percentil 85 (>+1 DE)			Riesgo de sobrepeso	Riesgo de sobrepeso
Percentil 50 (Media)				
< Percentil 15 (<-1 DE)	Riesgo de retraso de crecimiento *Ver nota 3*	Riesgo de bajo peso *Ver nota 3*	Riesgo de emaciación *Ver nota 3*	Riesgo de emaciación *Ver nota 3*
< Percentil 3 (<-2 DE)	Retraso de crecimiento	Bajo peso	Emaciación	Emaciación
< Percentil 1(<-3 DE)	Retraso de crecimiento severo	Bajo peso severo	Emaciación severa	Emaciación severa

Adaptado de: Patrones de crecimiento del niño de la OMS: Curso de capacitación sobre la evaluación del crecimiento del niño. OMS 2008.

A los cuantiles también se les llama percentiles

Cuartiles empíricos

- Todo esto transcurre en el mundo teórico, cuando conocemos P (o F6.97 7.74 7.84 8.58 8.93 9.58 11.06).
- Supongamos entonces tenemos una muestra aleatoria simple

  $$X_1, \ldots, X_n$$

  de la distribución F desconocida
- ¿Como estimamos los cuantiles?

Ejemplo

- Suponemos que queremos estimar la mediana del peso de los bebes de 6 meses, varones, y contamos con los pesos de los últimos 7 controles de un equipo de pedriatría:

  11,06 8,58 7,74 6,97 8,93 9,58 7,84

- Entonces ordenamos los datos de menor a mayor:

  6,97 7,74 7,84 8,58 8,93 9,58 11,06

- Y definimos la mediana como el dato central

  6,97 7,74 7,84 8,58 8,93 9,58 11,06

- Si tenemos dos datos centrales (n es par), tomamos el promedio de ambos.

Estimación del cuantil q y el rango inter-cuartil

- El cuantil q es el dato que deja a la izquierda de la muestra ordenada una proporción q de los datos.
- Así podemos estimar el rango inter-cuartílico: son los extremos de los datos centrales que abarcan la mitad de los datos:
  - Una cuarta parte queda a la izquierda
  - Una cuarta parte queda a la derecha
  - El rango es el 50 por ciento de los datos centrales

Summary

`Summary` es un comando de R que se aplica a una muestra y presenta un resumen incluyendo las siguientes datos

- El mínimo de la muestra
- El primer cuartil
- La mediana
- El tercer cuartil
- El máximo de la muestra

Ejemplos

- ```
  > summary(rnorm(100))
     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
  -2.2    -0.72     -0.10   0.00664  0.82     2.89
  ```

- ```
  > summary(rnorm(1000))
     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
  -3.29   -0.66     0.036   0.021  0.70     3.19
  ```

Boxplot

El boxplot o diagrama de caja es un gráfico que representa en forma esquemática una muestra. Incluye los siguientes elementos (midiendo verticalmente):

- La mediana de los datos
- Una caja cuya base es el primer cuartil y techo el tercer cuartiles
- Una línea horizontal en $Q_1 - 1{,}5RIC$ (cerca del mínimo)[1]
- Se plotean los datos menores que este nivel
- Una línea horizontal en $Q_3 + 1{,}5RIC$ (cerca del máximo)
- Se plotean los datos por encima de este nivel

[1] *RIC* es el rango intercuartil