

# Itxura: alborapena eta kurtosia

Josemari Sarasola

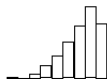
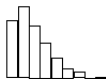
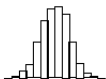
Estatistika eta datuen analisia

Gizapedia



Zentroaz eta sakabanatzeaz gainera, datu multzoetarako alborapena (asimetria) eta kurtosia deritzen ezaugarriak ere azter daitezke, banakuntza baten itxurarekin loturikoak.

- Batezbestekotik gora datu gehiago daudenean, handik behera baino, datu-multzoa *ezkerrera alboratua* dagoela esaten da (hau da,  $\bar{x} < Me$  gertatzen denean). (irudia)
- Batezbestekotik behera datu gehiago daudenean, handik gora baino, datu-multzoa *eskuinera alboratua* dagoela esaten da (hau da,  $\bar{x} > Me$  gertatzen denean). (irudia)
- Batezbestekotik gora eta behera datu kopuru berdinak daudenean, banakuntza *simetrikoa* edo *alboragabea* dela esaten da. (irudia)



Banakuntza simetrikoa

Eskuinera alboratua

Ezkerera alboratua

## Fisher alborapen-koefizientea

$$A_F = \frac{\sum(x_i - \bar{x})^3}{n s_x^3}$$

Abantaila: datu guztiak hartzen ditu. Oztopoa: ez da sendoa.

## Interpretazioa

$A_F < 0 \rightarrow$  ezkerrera alboratua

$A_F > 0 \rightarrow$  eskuinera alboratua

$A_F \approx 0 \rightarrow$  simetrikoa(?)

## Bowley alborapen-koefizientea

$$A_B = \frac{(Q_3 - Me) - (Me - Q_1)}{Q_3 - Q_1}$$

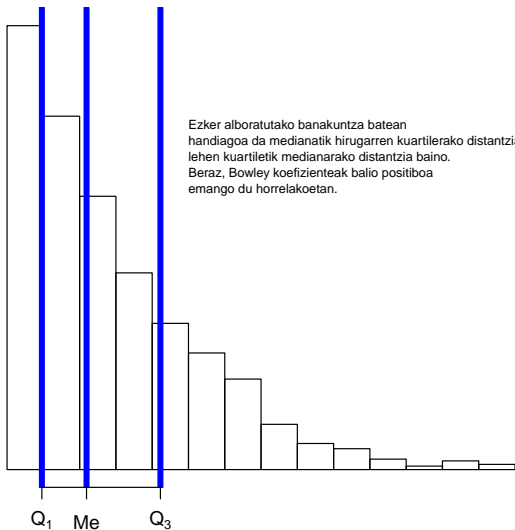
Abantaila: sendoa da. Oztopoa: ez ditu datu guztiak hartzen.

## Interpretazioa

$A_B < 0 \rightarrow$  ezkerrera alboratua

$A_B > 0 \rightarrow$  eskuinera alboratua

$A_B \approx 0 \rightarrow$  simetrikoa(?)



Ezker alboratutako banakuntza batean handiagoa da medianatik hirugarren kuartilerako distantzia lehen kuartiletik medianarako distantzia baino. Beraz, Bowley koefizienteak balio positiboa emango du horrelakoetan.

Ohiko banakuntza alboratuak tamaina eta kopuruei buruzkoak dira, hala nola enpresen langile-kopuruak, herrietako biztanleriak, enpresa ezberdinetako salmentak eta abar. Datu multzo horiek Pareto banakuntzaren eta gamma banakuntzaren bitartez modelizatu ohi dira.

Adibidez, enpresetako langile kopuruari buruzko datu-multzoak eskuinera alboratuak lirateke, langile gutxiko enpresa asko eta langile askoko enpresa gutxi daudelako.



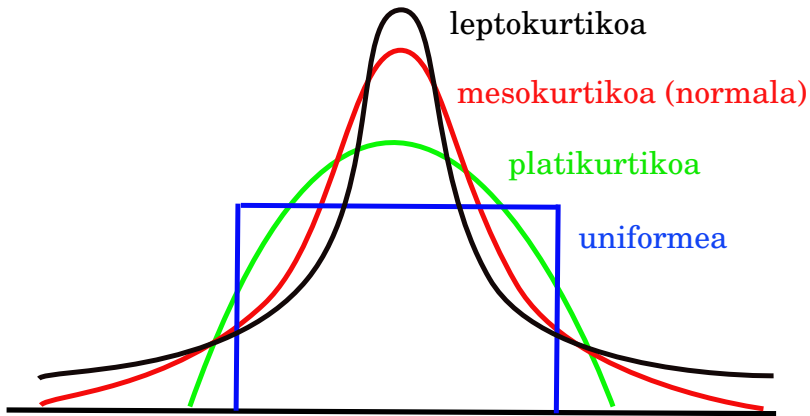
## Kurtosia

datu multzo baten zorrozatasuna (*sharpness*) da, zehatzago muturretan dauden datuen kopuruaren eta erdigunean biltzen diren datuen arteko proportzioa, bariantza konstante atxikirik.

## Kurtosi-maila hauek bereizten dira:

- leptokurtikoa (banakuntza normala baino zorrotzagoa)
- mesokurtikoa (banakuntza normalaren antzeko zorrozatasuna)
- platikurtikoa (banakuntza normala baino zapalagoa)
- uniforme (laua)
- U itxurakoa

## *KURTOSI MAILAK*



## Banakuntza normala

estatistikan maiz erabiltzen den datu-multzo idealizatua da, matematikoki definitua, eredu edo patroia moduan erabiltzen dena, guztiz simetrikoa eta kurtosi maila jakina duena. Horrexegatik erabiltzen da kurtosi mailak sailkatzeko erreferentzia moduan.

- *Normal* hitzak ez du esan nahi beste datu-multzoak anormalak direnik.
- Banakuntza normalaren antzekoak diren datu-multzoekin lan egitea interesgarria da estatistikan, horiekin lan egiteko teknika egokiak daudelako.
- Horregatik komeni da datu-multzoak nahiko simetrikoak eta mesokurtikoak (banakuntza normalaren antzeko kurtosia dutenak) izatea. Horrela denean, banakuntza normala eredu moduan erabil daitekeela esango dugu.

## Kurtosi neurriak

### Pearson koefizientea

$$K_P = \frac{\sum (x_i - \bar{x})^4}{n s_x^4}$$

- Eragozpen moduan, ez da sendoa,
- baina, abantaila moduan, datu guztiak hartzen ditu kontuan.

### Interpretazioa:

- $K_P > 3 \rightarrow$  leptokurtikoa
- $K_P \sim 3 \rightarrow$  mesokurtikoa
- $1.6 < K_P < 3 \rightarrow$  platikurtikoa
- $K_P \sim 1.6 \rightarrow$  uniforme
- $K_P < 1.6 \rightarrow$  U itxurakoa

## Kurtosi neurriak

### Moors koefizientea

$$K_M = \frac{(P_{87.5} - P_{62.5}) + (P_{37.5} - P_{12.5})}{P_{75} - P_{25}}$$

- Abantaila moduan, sendoa da,
- baina, eragozpen moduan, ez ditu datu guztiak kontuan hartzen.

### Interpretazioa:

- $K_M > 1.23 \rightarrow$  leptokurtikoa
- $K_M \sim 1.23 \rightarrow$  mesokurtikoa
- $K_M < 1.23 \rightarrow$  platikurtikoa

## Oharra

Aurreko interpretazioetarako erreferentziazko balioetatik emaitzak aski gertu dauden erabakitzeko, inferentzia-estatistikako tresnak (emaitzen errorea aztertzen dutenak) baliatu behar dira. Berdin esan daiteke, alborapena aztertzean.