Az adatok

Tegyük fel, hogy megmértük 40 ember vérnyomását és a következő értékeket kaptuk (szisztolés értékek, egész számokra kerekítve, Hgmm értékekben):

97, 103, 104, 106, 106, 111, 111, 112, 112, 112, 113, 113, 114, 114, 114, 115, 117, 118, 118, 118, 118, 119, 121, 121, 124, 124, 125, 125, 126, 127, 127, 129, 129, 129, 132, 132, 133, 133, 136, 151

A centrumok:

- átlag: $\bar{x} = 119.725$
- medián: 118 (a 20. és a 21. adat is 118)
- módusz: 118
- középtartomány: $\frac{97+151}{2} = 124$

A minta szórása:

$$s = \sqrt{\frac{n\sum(x^2) - \left(\sum x\right)^2}{n(n-1)}} = 10.595 \tag{1}$$

($s^2 = 112.256$)
variációs együttható:

$$cv = \frac{s}{\bar{x}} = 8.85\% \tag{2}$$

(Van-e az így számolt értéknek értelme? A vérnyomásértékek 0 pontja a légköri nyomás, a vér igazi nyomását a mért értékekhez 760 Hgmm-t hozzáadva kapjuk. Így felmerülhetne, hogy ezekből az igazi nyomásértékekből számolható átlagot használjuk itt. A szív munkáját azonban az jellemzi, hogy mekkora túlnyomást hoz létre, ezért az eredeti számadatokból számolt átlaghoz érdemes a szórást viszonyítani.)

Standard eltérés

$$z = \frac{x - \bar{x}}{s} \tag{3}$$

Néhány adatra kiszámolva:

x	z
97	-2.145
111	-0.823
118	-0.163
124	0.403
132	1.159
151	2.952

Kvartilisek, percentilisek

- első kvartilis: $Q_1 = 112.5$ (a 10. és a 11. adat átlaga)
- második kvartilis (ugyanaz, mint a medián): $Q_2 = 118$
- harmadik (felső) kvartilis: $Q_3 = 127$ (a 30. és a 31. elem átlaga, itt mindkettő 127)

Percentilisek számolása:
P_i számolása: legyen $L = \frac{n \cdot i}{100}$. Ha L egész, akkor P_i az L-ik és az $L+1$-ik elem átlaga, ha L tört, akkor felfele kerekítjük L-et, és P_i a kerekítés után kapott sorszámú elem.
Például:

- $P_{30} = 113.5$ ($\frac{30 \cdot 40}{100} = 12$, így P_{30} a 12. és a 13. elem átlaga
- $P_{63} = 124$ ($\frac{63 \cdot 40}{100} = 25.2$, így P_{63} a 26. elem

Kapcsolat a kvartilisekkel:
$Q_1 = P_{25}$, $Q_2 = P_{50}$, $Q_3 = P_{75}$

Gyakoriság eloszlás készítése

Általában célszerű nagyjából $\sqrt{n}$ osztályt használni, itt például 10 Hgmm szélességű osztályok megfelelőek lehetnek (természetesen lehetne máshogy is felosztani). Ennek megfelelően az alsó osztálykorlátok: 90, 100, 110, 120, 130, 140, 150, a felső osztálykorlátok: 99, 109, 119, 129, 139, 149, 159, az osztályhatárok: 89.5, 99.5, 109.5, 119.5, 129.5, 139.5, 149.5, 159.5. A gyakorisági eloszlás így:

értékek	gyakoriság (f_i)	felezőpont ($x_i^{(C)}$)	relatív gyakoriság
$90 - 99$	1	94.5	2.5%
$100 - 109$	4	104.5	10%
$110 - 119$	17	114.5	42.5%
$120 - 129$	12	124.5	30%
$130 - 139$	5	134.5	12.5%
$140 - 149$	0	144.5	0%
$150 - 159$	1	154.5	2.5%

A kumulatív gyakoriság eloszlás:

értékek	gyakoriság
kisebb, mint 100	1
kisebb, mint 110	5
kisebb, mint 120	22
kisebb, mint 130	34
kisebb, mint 140	39
kisebb, mint 150	39
kisebb, mint 160	40

A gyakorisági eloszlásból számított átlag:

$$\bar{x} = \frac{\sum_i f_i x_i^{(C)}}{\sum_i f_i} = \frac{\sum_i f_i x_i^{(C)}}{n} = 119.5 \tag{4}$$

A gyakorisági eloszlásból számított szórás:

$$s = \sqrt{\frac{n \sum_i \left(f_i \left(x_i^{(C)} \right)^2 \right) - \left(\sum_i f_i x_i^{(C)} \right)^2}{n(n-1)}} = 10.86 \tag{5}$$

Látható, hogy ezek nem térnek el jelentősen az eredeti adatokból számított átlagtól és szórástól.