

4A Datajoukkojen kuvaajat ja tunnusluvut

Tuntitehtävät

4A1 (Ikäjakauma.) Allaoleva taulukko kuvaa Suomen väestörakennetta ikäluokittain 31.12.2015:

Ikä (v)	Lukumäärä
0–14	896 023
15–24	640 387
25–44	1 363 155
45–64	1 464 640
65–74	642 428
75–	480 675

(Lähde: Tilastokeskus)

- Piirrä ikäjakauman histogrammi taulukon mukaisin luokkaväleihin (viimeisen luokan ylärajaksi voit valita 110) niin, että vaaka-akselin yksikkö on vuosi ja pystyakselin yksikkö on prosenttia per vuosi.
- Laske histogrammin palkkien pinta-alat ja tarkista, että histogrammin alle jäävän alueen pinta-alaksi tulee 100%.

Vastaa histogrammin avulla seuraaviin kysymyksiin:

- Oliko väestössä enemmän 10-vuotiaita vai 30-vuotiaita?
- Oliko väestössä enemmän 1-vuotiaita kuin 66-vuotiaita?

4A2 (Kvantiilit.) Datajoukon $x = (x_1, \dots, x_n)$ kvantiilifunktio määritetään R:ssä oletusarvoisesti seuraavasti. Merkitään $x_{(1)}$ = datajoukon pienin arvo, $x_{(2)}$ = toiseksi pienin, jne. Tällöin data saadaan järjestettyä muotoon $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. Seuraavaksi jaetaan vaaka-akselin yksikköväli $n - 1$ yhtä pitkään osaväliin käyttämällä välien reunapisteinä lukuja $p_k = (k - 1)/(n - 1)$, $k = 1, \dots, n$. Kvantiilifunktion kuvaaja piirretään piirtämällä (x, y) -tasoon ensin pisteet $(p_k, x_{(k)})$, $k = 1, \dots, n$, ja sen jälkeen yhdistämällä pisteet suorilla viivoilla.

Piirrä (käsin) paperille seuraavien datajoukkojen kvantiilifunktiot ja määritä kuvaajasta näille alakvartiili $Q(0.25)$, mediaani $Q(0.50)$ ja yläkvartiili $Q(0.75)$:

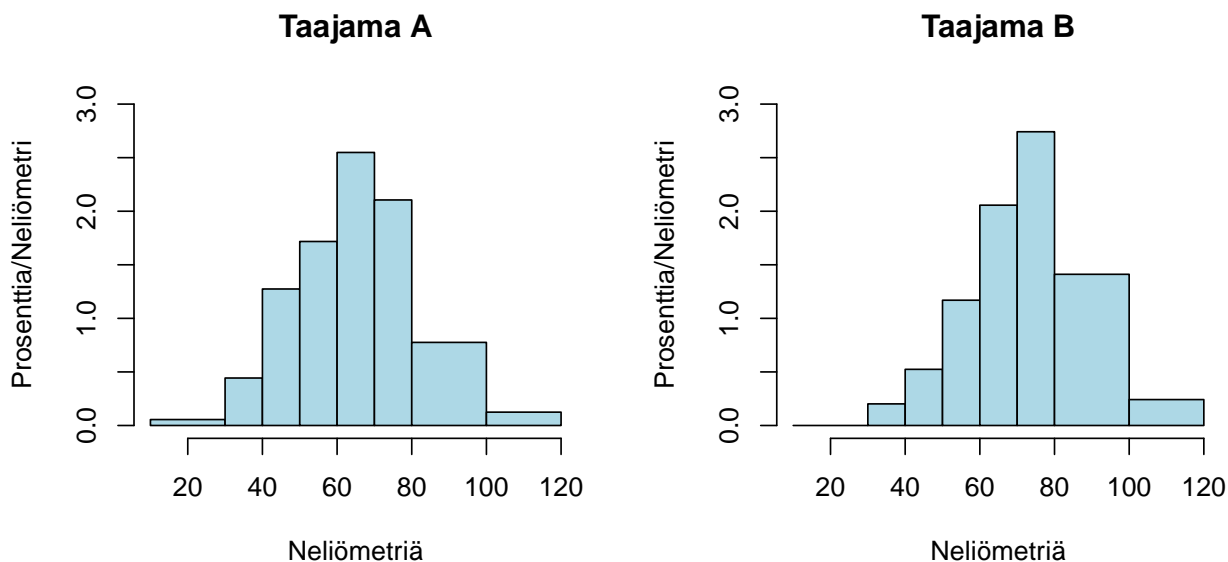
- $x = (1000, 2000, 5000, 9000)$,
- $x = (1000, 2000, 2000, 8000, 9000)$,
- $x = (1, 20, 1, 5, 1)$.

Tutki seuraavaksi seuraavia väitteitä. Perustele miksi väite on tosi tai kehittele vastaesimerkki, jonka perusteella väite on epätosi.

- (d) Datajoukon keskiarvo on aina yhtäsuuri kuin sen mediaani.
- (e) Datajoukon alakvartiili on aina pienempi tai yhtä kuin mediaani.
- (f) Datajoukon alakvartiili on aina pienempi tai yhtä suuri kuin keskiarvo.

Kotitehtävät

4A3 (Asuinneliöt.) Taajamassa A on 361 asuntoa ja taajamassa B on 248 asuntoa. Alla olevissa histogrammeissa on esitettyä kummankin taajaman asuntojen neliömäärät.



Vastaa histogrammien avulla seuraaviin kysymyksiin (oletetaan yksinkertaisuuden vuoksi, että yksikään havaituista neliömääristä ei osu luokkien reunoille).

- (a) Kuinka monessa taajaman B asunnossa neliömetrejä on vähintään 80?
- (b) Kumman taajaman asuntojen neliömäärän mediaani on suurempi?

4A4 (Tiilikone.) Kone tekee tiiliä, joiden massa vaihtelee satunnaisesti jonkin verran. Tarkastettavaksi valitun 25 tiilen erän massat on annettu alla.

5.18	3.75	3.59	3.19	4.38
4.10	6.93	6.23	5.22	5.02
3.94	6.13	5.68	4.42	5.42
5.14	3.69	6.24	6.41	4.56
4.83	4.97	3.38	5.47	5.61

- (a) Määritä datasta seuraavat tunnusluvut: maksimi ja mediaani.

(b) Määritä taulukon varjostettujen solujen keskiarvo ja otoskeskihajonta.