

Jatkuva tilastollinen jakauma

Esimerkki 1

Alla on esitetty kaikki yli 90 metrin tulokset, jotka on heitetty nykyisellä keihästyypillä. Luokittele tulokset tasavälisiin luokkiin, joista ensimmäinen on 90,00–90,99 m. Havainnollista tulosten jakaumaa kuvion avulla.

98,48 m; 95,66 m; 95,54 m; 94,64 m; 94,44 m; 94,02 m; 93,90 m; 93,88 m; 93,09 m; 92,80 m; 92,72 m; 92,61 m; 92,60 m; 92,42 m; 92,41 m; 92,28 m; 92,28 m; 92,12 m; 92,12 m; 91,82 m; 91,69 m; 91,68 m; 91,59 m; 91,53 m; 91,53 m; 91,50 m; 91,50 m; 91,50 m; 91,46 m; 91,40 m; 91,39 m; 91,36 m; 91,34 m; 91,33 m; 91,31 m; 91,30 m; 91,29 m; 91,28 m; 91,28 m; 91,28 m; 91,27 m; 91,23 m; 91,23 m; 91,23 m; 91,20 m; 91,11 m; 91,07 m; 91,04 m; 90,97 m; 90,88 m; 90,87 m; 90,86 m; 90,81 m; 90,80 m; 90,80 m; 90,76 m; 90,73 m; 90,71 m; 90,71 m; 90,68 m; 90,61 m; 90,60 m; 90,60 m; 90,60 m; 90,59 m; 90,57 m; 90,56 m; 90,56 m; 90,54 m; 90,50 m; 90,44 m; 90,43 m; 90,37 m; 90,34 m; 90,34 m; 90,33 m; 90,33 m; 90,30 m; 90,28 m; 90,24 m; 90,20 m; 90,18 m; 90,17 m; 90,16 m; 90,13 m; 90,12 m; 90,11 m; 90,09 m; 90,06 m; 90,06 m; 90,06 m; 90,02 m

Tilanne 1.1.2018

Ratkaisu

Jakaumaa ei kannata havainnollistaa pylväskuviolla, koska

suurin osa heiton pituuksista esiintyy vain kerran ja mahdollisia arvoja pisimmän ja lyhyimmän heiton välillä on hyvin paljon.

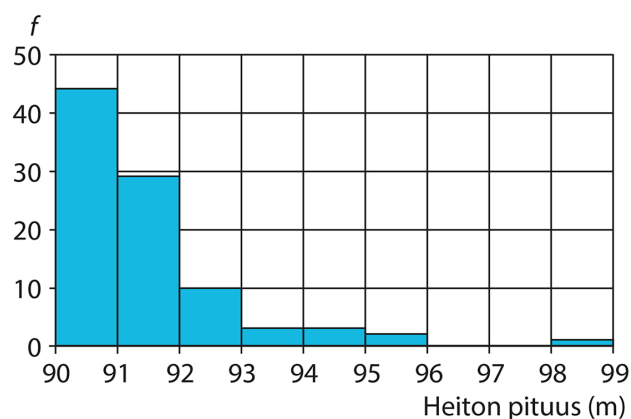
Luokitellaan heittojen pituudet tasavälisiin luokkiin metrin välein.

Heiton pituus (m)	Frekvenssi
90,00–90,99	44
91,00–91,99	29
92,00–92,99	10
93,00–93,99	3
94,00–94,99	3
95,00–95,99	2
96,00–96,99	0
97,00–97,99	0
98,00–98,99	1

Avaa OpenOffice-tiedosto →

Piirretään luokitellusta aineistosta kuvio, jossa pylväät ovat kiinni toisissaan.

Keihäänheiton pituuksia



Kuvio on nimeltään histogrammi, ja se havainnollistaa

pylväskuviota selkeämmin heittojen pituuksien jakaumaa. Yhteen piirretyt pylväät ottavat huomioon sen, että pituus on muuttuja, joka voi saada mitä tahansa arvoja tietyltä väliltä.

Avaa appletti →

Aineiston luokittelu

Useita eri havaintoarvoja taulukoitaessa on havainnollisempaa ryhmitellä arvoja luokkiin kuin kuvata jokainen arvo erikseen. Luokittelu tehdään aina niin, että jokainen havaintoarvo kuuluu vain yhteen luokkaan. Luokittelu tehdään pääsääntöisesti tasavälisenä, ja luokkavälin pituus ja luokkien määrä valitaan tapauskohtaisesti.

Luokille määrätään ala- ja ylärajat. Luokan ala- ja ylärajan keskiarvoa sanotaan luokkakeskukseksi.

Esimerkissä 1 ensimmäisen luokan luokkakeskus on

$$\frac{90,00 + 90,99}{2} = 90,495 \approx 90,50.$$

Luokitellusta aineistosta keskiarvoa ja keskihajontaa laskettaessa havaintoarvoja edustavat luokkakeskukset.

Esimerkki 2

Laske opiskelija-asuntojen pinta-alojen keskiarvo ja keskihajonta.

Pinta-ala (m ²)	Frekvenssi
25–44	29
45–64	10
65–84	12

85–104	7
105–124	8
125–144	2
Yhteensä	68

Avaa OpenOffice-tiedosto →

Ratkaisu

Koska aineisto on valmiiksi luokiteltu, keskiarvoa ei voida laskea yksittäisistä havaintoarvoista vaan se määritetään luokkakeskusten avulla.

Pinta-ala (m²)	Frekvenssi	Luokkakeskus
25–44	29	$\frac{25 + 44}{2} = 34,5$
45–64	10	54,5
65–84	12	74,5
85–104	7	94,5
105–124	8	114,5
125–144	2	134,5
Yhteensä	68	

Määritetään frekvenssitaulukosta keskiarvo ja keskihajonta laskentaohjelman avulla.

Keskiarvo on 63 m² ja keskihajonta 31 m².

Vastaus

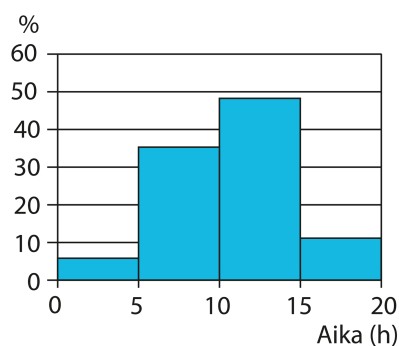
Opiskelija-asuntojen pinta-alojen keskiarvo on 63 m² ja keskihajonta 31 m².

Ohjelma kannattaa valita niin, että tunnuslukujen laskeminen suoraan frekvenssitaulukosta on mahdollista.

Jatkuva ja diskreetti jakauma

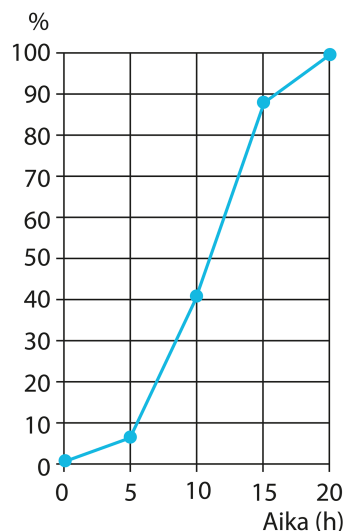
havainnollistaminen

Jatkuva tilastollinen muuttuja voi saada mitä tahansa arvoja tietyllä välillä. Pituuden ja painon jakaumat ovat jatkuvia jakaumia havainnollistetaan histogrammilla, esimerkiksi jatkuvista jakaumista.



a voi saada vain tiettyjä ja kouluarvosanoja kuvaavat kreeteistä jakaumista.

Toinen tapa kuvata jatkuvaa jakaumaa on kertymäkuvaaja eli summafrekvenssikäyrä. Se kertoo, miten paljon havaintoarvoja on kertynyt tiettyyn rajaan asti. Kertymäkuvaajassa käytetään tavallisimmin suhteellisia summafrekvenssejä, ja ne merkitään luokkien ylärajojen kohdalle.



Avaa appletti →

Esimerkki 3

Onko muuttuja diskreetti vai jatkuva?

- a) lemmikkien lukumäärä kotitalouksissa
- b) yönunen pituus
- c) lukiossa suoritettujen kurssien lukumäärä
- d) kehon lämpötila

Ratkaisu

- a) Lemmikkien lukumäärä voi olla vain luonnollinen luku, joten kyseessä on diskreetti muuttuja.
- b) Yönunen pituus voi saada mitä tahansa arvoja tietyltä väliltä, joten kyseessä on jatkuva muuttuja.
- c) Kurssien lukumäärä voi olla vain luonnollinen luku, joten se on diskreetti muuttuja.
- d) Kehon lämpötila voi saada mitä tahansa arvoja tietyltä väliltä, joten kyseessä on jatkuva muuttuja.

Esimerkki 4

Katsastusaseman insinööri kirjasi ylös katsastettavien autojen hiilidioksidipäästöjä.

Hiilidioksidipäästöt (g/km)	<i>f</i>
120–139	102
140–159	196
160–179	49
180–199	201
200–219	18
220–239	94
240–259	60

Avaa OpenOffice-tiedosto →

- a) Piirrä taulukon perusteella autojen

hiilidioksidipäästöistä kertymäkuvaaja.

b) Mikä on hiilidioksidipäästöjen mediaani?

c) Kuinka monella prosentilla autoista päästöt ovat yli 210 g/km?

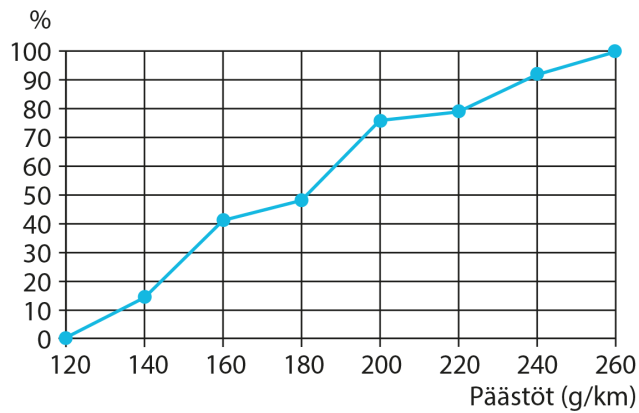
Ratkaisu

a) Kertymäkuvaajan piirtämiseksi lasketaan
summafrekvenssit sf ja suhteelliset summafrekvenssit $sf\%$.

Hiilidioksidipäästöt (g/km)	f	sf	$sf\%$
120–139	102	102	$\frac{102}{720}$ $\approx 14\%$
140–159	196	$102 + 196 = 298$	41 %
160–179	49	347	48 %
180–199	201	548	76 %
200–219	18	566	79 %
220–239	94	660	92 %
240–259	60	720	100 %

Piirretään kertymäkuvaaja suhteellisten
summafrekvenssien avulla. Suhteelliset summafrekvenssit
merkitään luokkien ylärajojen kohdalle.

Kertymäkuvaaja autojen
hiilidioksidipäästöistä



b) Luetaan mediaani kertymäkuvaajasta 50 prosentin kohdalta. Mediaani on noin 181 g/km.

c) Autoista 78 % tuottaa hiilidioksidipäästöjä alle 210 g/km. Autoja, joiden hiilidioksidipäästöt ovat yli 210 g/km, on $100\% - 78\% = 22\%$.

Vastaus

b) Hiilidioksidipäästöjen mediaani on 181 g/km.

c) Autoista 22 % tuottaa hiilidioksidipäästöjä yli 210 g/km.

Kvartiilit

Kvartiilit jakavat aineiston havaintoarvot neljään yhtä suureen osaan.

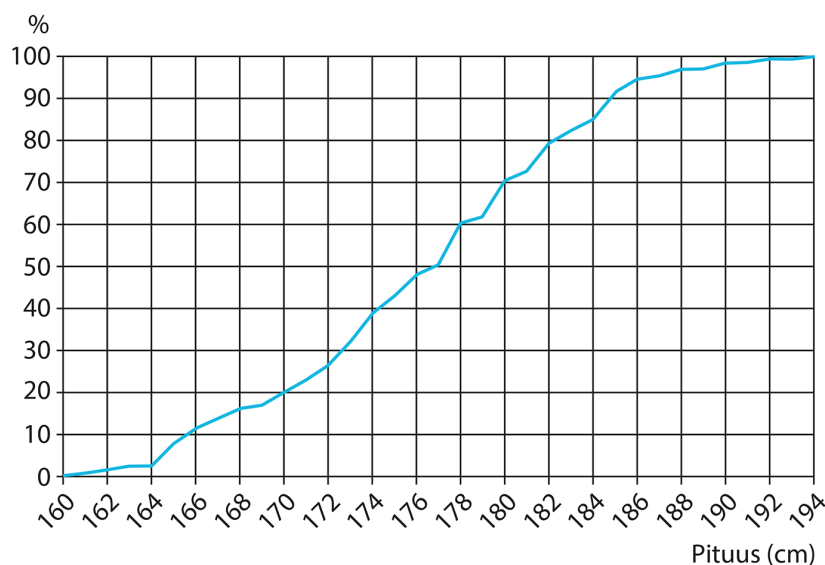
- Alakvartiililla tarkoitetaan sitä muuttujan arvoa, joka rajaa jakauman alusta 25 % havaintoarvoista.
- Mediaani eli keskikvartiili on se muuttujan arvo, jonka ala- ja yläpuolella on 50 % havaintoarvoista.
- Yläkvartiililla tarkoitetaan sitä muuttujan arvoa, joka rajaa jakauman alusta 75 % havaintoarvoista.

Kvartiilit voi määrittää laskentaohjelman tai suhteellisen summafrekvenssin avulla. Laskentaohjelmissa alakvartiilin merkintä on usein Q_1 ja yläkvartiilin Q_3 .

Esimerkki 5

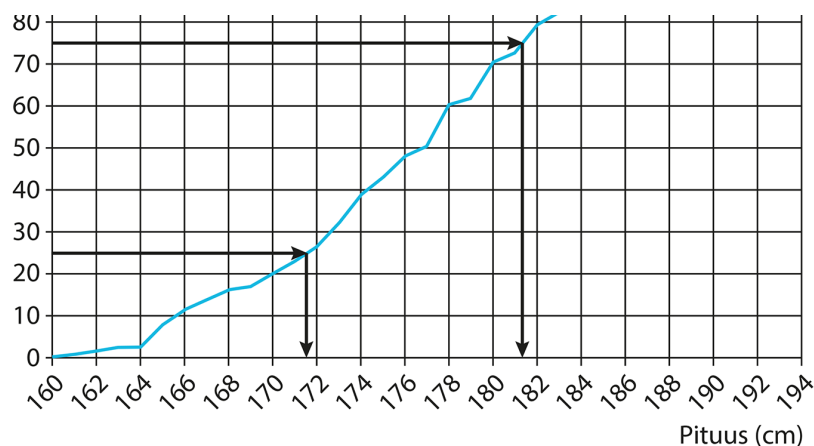
Kertymäkuvaaja esittää Suomen naisten Korisliigassa pelaavien pelaajien pituusjakaumaa syksyllä 2017. Määritä pituuden ala- ja yläkvartiili.

Kertymäkuvaaja naisten korisliigan pelaajien pituuksista



Avainkäsitteet

- luokkakeskus
- diskreetti jakauma
- jatkuva jakauma
- histogrammi
- summafrekvenssikäyrä eli kertymäkuvaaja
- luokan alaraja
- luokan yläraja
- alakvartiili
- yläkvartiili



Kuvaajasta nähdään, että alakvartiili on noin 171,5 cm ja yläkvartiili noin 181,5 cm. Siis neljäsosa naisista on pituudeltaan alle 171,5 cm ja neljäsosa yli 181,5 cm.

Vastaus

Alakvartiili on 171,5 cm ja yläkvartiili 181,5 cm.