

4

Luokittelu

Luokitellun aineiston frekvenssijakauma

Usein muuttuja saa eri arvoja niin paljon, ettei niiden kaikkien luetteleminen frekvenssitaulukossa ole havainnollista eikä edes järkevää.

Tällöin on järkevää **luokitella** aineisto eli ryhmitellä se sopivalla tavalla

elokuvien kestot minuuteissa ovat:

73	5	9	5	85	29	56	76	14	12
82	82	57	30	8	90	10	38	69	20
30	75	57	58	77	60	9	15		

Kesto (min)	f
1–15	9
16–30	4
31–45	1
46–60	5
61–75	3
76–90	6

Luokittelu tiivistää tilaston, mutta samalla tieto yksittäisistä havainnoista menetetään.

Arvosana	f
5	2
6	4
7	10
8	9
9	6
10	1



Arvosana	f
5–6	6
7–8	19
9–10	7

Arvosanaluoan 7 – 8

- alaraja on 7
- yläraja on 8.

Arvosanaluoan 7 – 8 luokkavälin pituus on $8 - 7 = 1$ (arvosanaa).

Luokka ilmoitetaan **luokkavälinä** eli ilmoitetaan luokan **alaraja** ja **yläraja**.

Luokat valitaan **tasavälisiksi** eli yhtä pitkiksi, ellei ole erityistä syytä valita toisin.

Luokan **leveys** eli **luokkavälin pituus** saadaan laskemalla luokan ylärajan ja alarajan erotus.

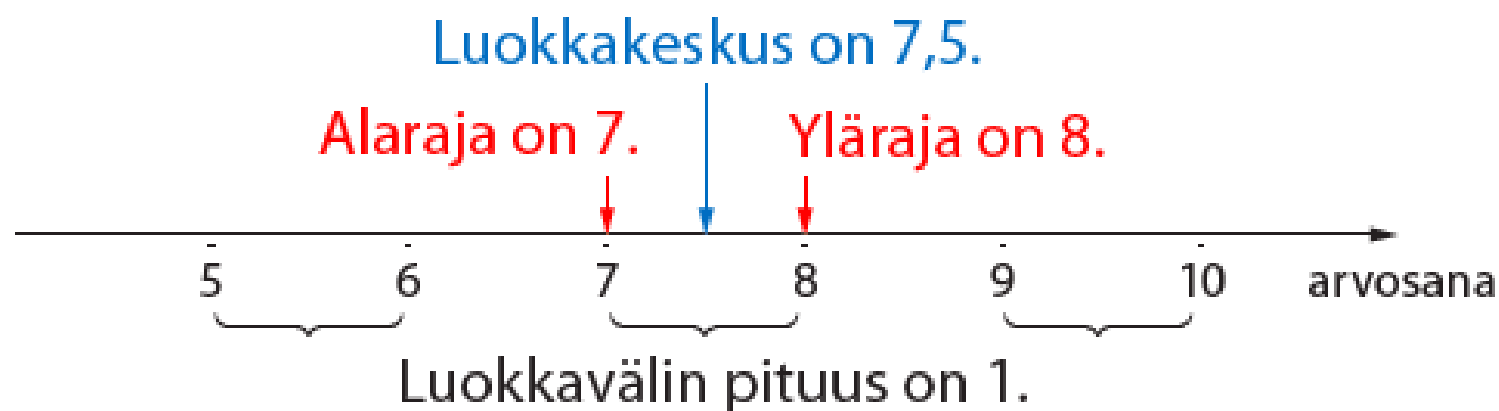
Luokittelun jälkeen yksittäisiä havaintoja vastaavat luokkien **luokkakeskukset**.

Luokkakeskus on luokan ala- ja ylärajan keskiarvo.

$$\text{luokkakeskus} = \frac{\text{alaraja} + \text{yläraja}}{2}$$

Arvosana	f
5–6	6
7–8	19
9–10	7

Esimerkiksi arvosanaluvokan 7–8 luokkakeskus on $\frac{7+8}{2} = \frac{15}{2} = 7,5$.



Peltopyy on suomalaisista pesimälinnuista ahkerin munija. Yhdessä pesässä voi olla jopa yli 20 munaa. Biologi tutki eräällä alueella pesivien peltopyiden pesissä olevia munamääriä. Tulokset olivat seuraavat:



5	12	16	19	23	16	17	11	20	16
23	14	18	19	14	17	15	18	13	17
16	10	21	16	13	12	18	8	17	14

- Mikä oli pienin ja mikä suurin munamäärä tämän alueen peltopyypesissä?
- Taulukkoon on valittu tasavälinen luokittelu. Laske luokkavälin pituus.
- Täydennä frekvenssitaulukko laskemalla, kuinka monta pesää kuuluu munamäärältään mihinkin luokkaan.
- Missä luokassa on eniten havaintoja? Laske tämän luokan luokkakeskus.

Munia	f
5–9	
10–14	
15–19	
20–24	

Ratkaisu

Tarkastellaan aineistoa taulukkolaskentaohjelmassa.

- a) Lajitellaan aineisto nousevaan järjestykseen. Lajittelun jälkeen nähdään, että pienin arvo on 5 ja suurin arvo on 23. Pesissä oli siis 5–23 munaa.

- b) Luokkavälin pituus saadaan laskemalla luokan ylä- ja alarajan erotus. Luokkavälin pituus on $9 - 5 = 4$ (munaa).

Munia	f
5–9	2
10–14	9
15–19	15
20–24	4

- c) Lasketaan kuhunkin luokkaan kuuluvien havaintojen lukumäärä. Kootaan tieto taulukkoon.

- d) Eniten havaintoja eli 15 havaintoa on luokassa 15–19. Luokkakeskus saadaan laskemalla luokan ala- ja ylärajan keskiarvo.

$$\text{luokkakeskus} = \frac{15 + 19}{2} = \frac{34}{2} = 17$$

Luokkakeskus on 17 (munaa).

Luokittelu on tasavälinen eli jokaisen luokan leveys on sama.

Frekvenssien laskeminen voidaan tehdä esimerkiksi luokkien ylärajojen ja ohjelman Taajuus-funktion avulla.

Vastaus

- a) pienin 5, suurin 23
b) 4 (munaa)
d) 15–19, luokkakeskus 17 (munaa)

**Tee tehtävät 4.1
ja 4.2**

Jatkuvan muuttujan tilasto

Diskreetti muuttuja saa vain tiettyjä arvoja. Kaikki mahdolliset arvot voidaan luetella.

Jatkuva muuttuja voi periaatteessa saada mitä tahansa arvoja tietyltä väliltä.

Jatkuvia tilastollisia muuttujia käsitellään yleensä luokiteltuina.

Elokuvan kesto on jatkuva muuttuja. Suomen vuoden 2020 katsotuimmat 40 elokuvaa on oheisessa taulukossa luokiteltu keston mukaan kuuteen tasaväliseen luokkaan.

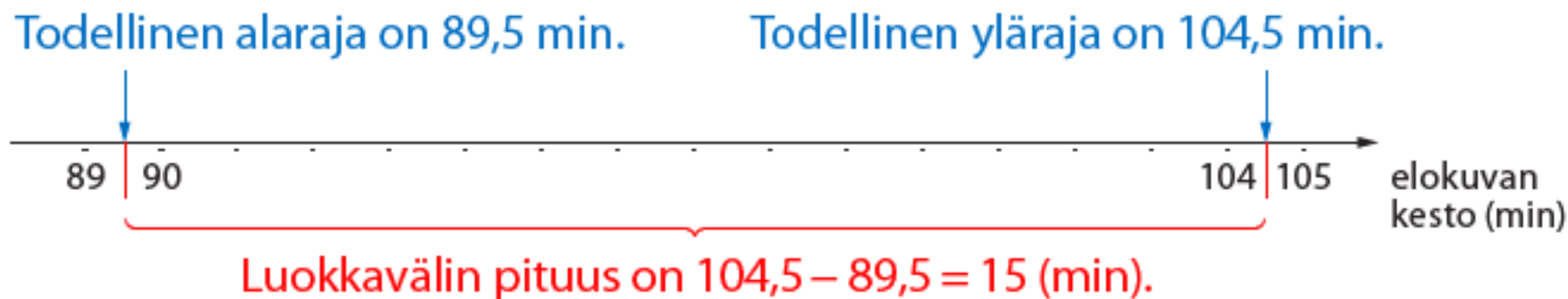
Elokuvan kesto (min)	f
60–74	2
75–89	8
90–104	15
105–119	5
120–134	7
135–149	3

Myös luokkarajat on taulukossa pyöristetty minuutin tarkkuuteen.

Tämän luokan **todellinen alaraja** on 89,5 min
ja **todellinen yläraja** 104,5 min.

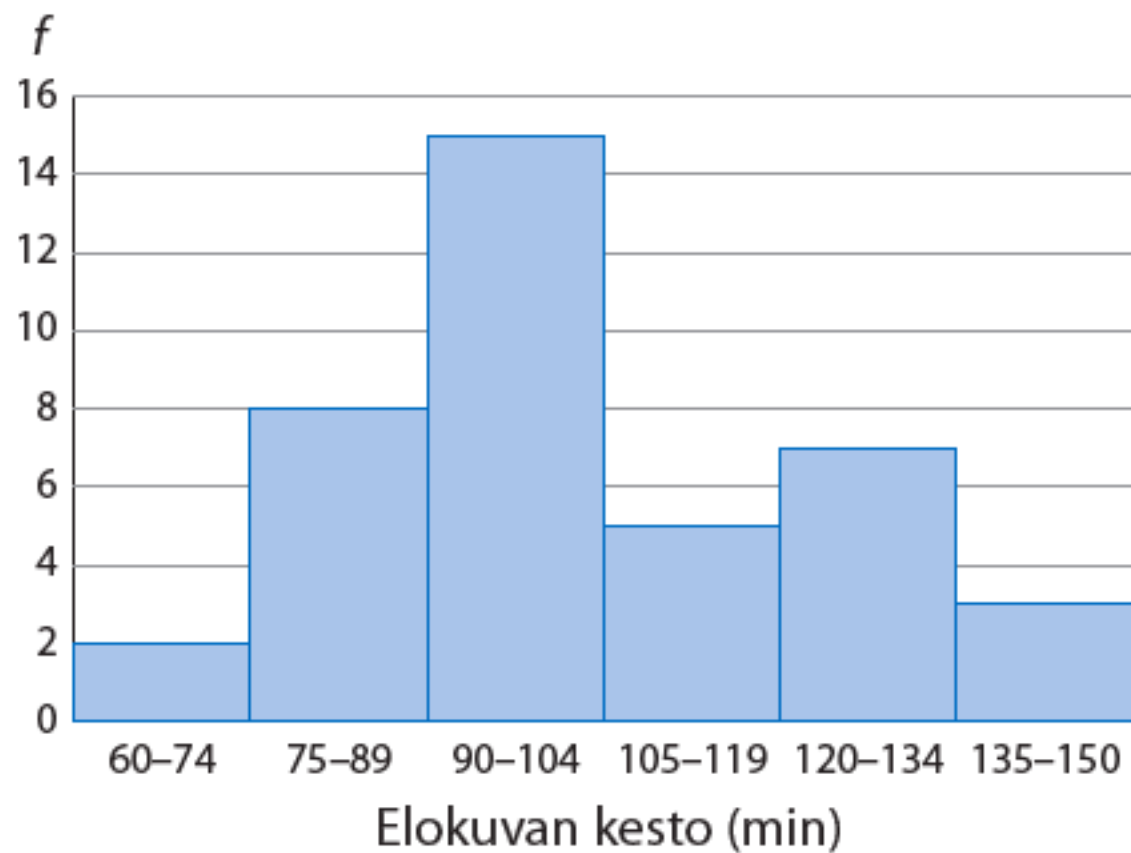
Todellinen yläraja on
 $104,4999... = 104,5$.
 Todellisia luokkarajoja
 käytetään mm. graafisissa
 esityksissä ja kun
 lasketaan tunnuslukuja.

Elokuvan kesto (min)	f
60–74	2
75–89	8
90–104	15
105–119	5
120–134	7
135–149	3



Jatkuvan muuttujan luokiteltua frekvenssijakaumaa kuvataan usein **histogrammilla**.

Elokuvan kesto (min)	f
60–74	2
75–89	8
90–104	15
105–119	5
120–134	7
135–149	3



Histogrammissa pylväät piirretään yhteen.

B Esimerkki 2

Taulukossa on esitetty EU-jäsenvaltioiden pinta-alat (/1000 km²) vuonna 2021. *Lähde: EU*



42	31	111	506	70	302	84	132	88
9	65	65	2,6	0,3	92	313	633	238
439	357	49	20	338	43	79	93	45

- a) Luokittele aineisto tasavälisiin luokkiin niin, että ensimmäinen luokka on 0–149 (1000 km²).
- b) Havainnollista luokiteltua jakaumaa histogrammilla.
- c) Havainnollista luokiteltua jakaumaa ympyräkaaviolla.
- d) Kuvaile EU-jäsenmaiden pinta-alojen jakaumaa sanallisesti.

Ratkaisu

- a) Aloitetaan selvittämällä aineiston pienin ja suurin arvo. Lajitellaan aineisto nousevaan järjestykseen. Pienin arvo on 0,3 ja suurin 633.

Ratkaisu

Pinta-ala voi saada periaatteessa mitä tahansa arvoja tietyltä väliltä eli kyseessä on jatkuva muuttuja. Jatkuvan muuttujan luokittelussa luokkavälin pituus saadaan laskemalla peräkkäisten luokkien alarajojen erotus.

- a) Aloitetaan selvittämällä aineiston pienin ja suurin arvo. Lajitellaan aineisto nousevaan järjestykseen. Pienin arvo on 0,3 ja suurin 633. Aineistossa on tieto yhteensä 27 maan pinta-alasta.

Pienin arvo kuuluu ensimmäiseen luokkaan **0–149**, jonka yläraja on 149. Luokkarajat on annettu kokonaislukuina. Seuraavan luokan alaraja on 150. Luokkavälin pituus on alarajojen erotus eli $150 - 0 = 150$.

Seuraavan luokan ala- ja yläraja saadaan lisäämällä edellisen luokan ala- ja ylärajaan luokkavälin pituus eli luku **150**.

Luokan alaraja	Luokan yläraja
0	149
150	$149 + 150 = 299$
$150 + 150 = 300$	$299 + 150 = 449$
$300 + 150 = 450$	$449 + 150 = 599$
$450 + 150 = 600$	$599 + 150 = 749$

Suurin arvo 633 kuuluu viimeiseen luokkaan 600–749. Määritetään luokkien frekvenssit ohjelmistolla.

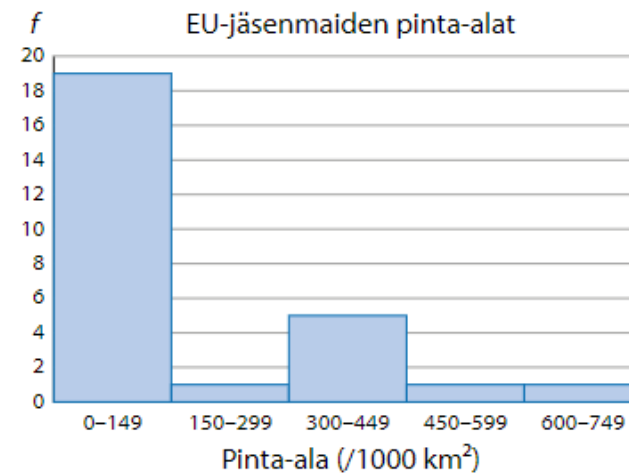
Frekvenssien laskeminen voidaan tehdä esimerkiksi luokkien ylärajojen ja ohjelman Taajuus-funktion avulla.

Pinta-ala (/1000 km ²)	<i>f</i>
0–149	
150–299	
300–449	
450–599	
600–749	

Määritetään frekvenssit.

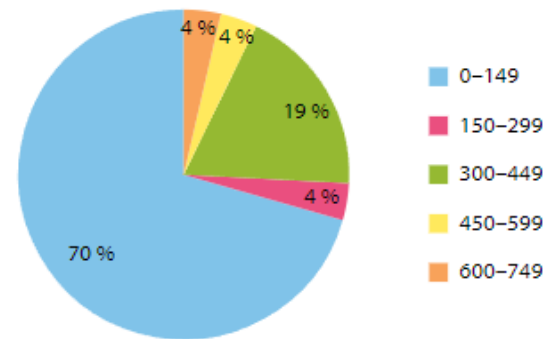
Pinta-ala (/1000 km ²)	<i>f</i>
0–149	19
150–299	1
300–449	5
450–599	1
600–749	1

b) Piirretään histogrammi ohjelmistolla.



c) Piirretään ympyräkaavio ohjelmistolla.

EU-jäsenmaiden pinta-alat (/1000 km²)



d) Pinta-alaltaan pienin EU-jäsenmaa on kooltaan 300 km² ja suurin 633 000 km². Ylivoimainen enemmistö (19 kpl eli 70 %) maista on pinta-alaltaan verraten pieniä eli alle 150 000 km². Pinta-alaltaan keskikokoisia maita on muutama (5 kappaletta) ja pinta-alaltaan suuria valtioita vielä vähemmän.

**Tee tehtävät
4.3 ja 4.4**

B**Esimerkki 3**

Laskentataulukossa on tiedot Yhdysvaltojen presidenttien virkaanastumis-
iästä.

- a) Mikä on presidenttien virkaanastumisiän vaihteluväli?
- b) Luokittele iät tasavälisiin luokkiin niin, että alimman luokan alaraja on aineiston pienin arvo ja luokkavälin pituus on 8 vuotta.
- c) Barak Obama oli virkaan astuessaan 47-vuotias. Kuinka monta presidenttiä kuuluu Obaman kanssa samaan ikäluokkaan b-kohdan luokittelussa?
- d) Yhdysvaltojen nykyinen presidentti, Joe Biden, oli virkaan astuessaan 78-vuotias. Kuinka monta prosenttia presidenteistä kuuluu Bidenin kanssa samaan ikäluokkaan b-kohdan luokittelussa? Ilmoita vastaus 0,1 prosentin tarkkuudella.

- a) Lajitellaan aineisto taulukkolaskentaohjelmassa nousevaan järjestykseen. Lajittelun jälkeen nähdään, että pienin arvo on 42 ja suurin 78. Presidenttien virkaanastumisiän vaihteluväli on siis 42–78 (vuotta).
- b) Ikä on jatkuva muuttuja, joten luokkavälin pituus on peräkkäisten luokkien alarajojen erotus. Ensimmäisen luokan alaraja on 42 (vuotta). Luokkavälin pituus on 8, joten seuraavan luokan alaraja on $42 + 8 = 50$. Samoin saadaan muiden luokkien alarajat. Lopuksi päätellään ylärajat.

Aineiston suurin arvo 78 kuuluu viimeiseen luokkaan 74–81.

Ikä (vuotta)	f
42–49	
50–57	
58–65	
66–73	
74–81	

Määritetään frekvenssit.

Ikä (vuotta)	f
42–49	9
50–57	24
58–65	9
66–73	3
74–81	1

- c) Obaman ikä, 47 vuotta, on välillä 42–49. Frekvenssitaulukosta nähdään, että tämän luokan frekvenssi on 9, joten tähän ikäluokkaan kuuluu Obama mukaan lukien 9 presidenttiä.
- d) Bidenin ikä, 78 vuotta, kuuluu viimeiseen luokkaan 74–81. Tämän luokan frekvenssi on 1 eli Biden on ainoa presidentti tässä ikäluokassa. Suhteellinen frekvenssi saadaan laskemalla frekvenssin prosenttiosuus kaikista havainnoista. Lasketaan havaintojen kokonaismäärä.
- $$9 + 24 + 9 + 3 + 1 = 46$$

Havaintojen kokonaismäärä on frekvenssien summa.

Lasketaan ikäluokan 74–81 suhteellinen frekvenssi.

$$\frac{1}{46} = 0,0217... \approx 2,2 \%$$

Vastaus

- a) 42–78 vuotta c) Obama mukaan lukien 9 presidenttiä d) 2,2 %

**Tee tehtävät 4.5,
4.7 ja 4.9**

5

Tilastollisia tunnuslukuja

Moodi ja mediaani

Moodi on yleisin muuttujan arvo ja mediaani on suuruusjärjestyksessä olevan aineiston keskimäinen arvo.

Muuttuja x	f	$f\%$	$sf\%$
1	18	36	36
2	5	10	46
3	12	24	70
4	15	30	100

Muuttujan arvolla **1** on

- suurin frekvenssi (18)
- suurin suhteellinen frekvenssi (36 %).

$Mo = \mathbf{1}$

Muuttujan arvo **3** on se, johon mennessä kertymä (70 %) on ensimmäisen kerran yli 50 % havainnoista.

$Md = \mathbf{3}$

Luokitellun aineiston moodi ja mediaani ilmoitetaan usein **moodiluokkana** ja **mediaaniluokkana**.

Työmatkan pituus (km)	f	$f\%$	$sf\%$	Luokka- keskus (km)
15,0–15,5	1	5	5	15,25
15,6–16,1	8	40	45	15,85
16,2–16,7	4	20	65	16,45
16,8–17,3	7	35	100	17,05

Työmatkan pituuden

- moodiluokka on **15,6–16,1** (km) ja moodi $M_o = 15,85$ (km)
- mediaaniluokka on **16,2–16,7** (km) ja mediaani $M_d = 16,45$ (km).

Jos tilastoaineisto on pieni, moodi ja mediaani voidaan määrittää ilman frekvenssijakaumaa, yksittäisten havaintoarvojen avulla.

Moodi määritetään etsimällä muuttujan arvoista se, jota on eniten.

3, **11**, 3, 10, **11**, 8, **11**, 8, 2 Mo = **11**

Mediaanin määrittämistä varten aineisto asetetaan suuruusjärjestykseen.

Tällöin mediaani on se muuttujan arvo, joka on aineistossa keskimmäisenä.

2, 3, 3, 8, **8**, 10, 11, 11, 11 Md = **8**

Jos keskimmäisiä arvoja on kaksi, mediaani on näiden keskiarvo.

2, 3, **5**, **7**, 8, 8 $Md = \frac{5+7}{2} = 6$

A Esimerkki 1

Opiskelijat tekivät pienissä ryhmissä tilastollisia tutkimuksia.
Yksi tutkimuksen kohde oli ryhmän jäsenten serkkujen lukumäärä.
Ohessa on lueteltu kahden ryhmän tulokset.

Ryhmä A

2, 2, 3, 1, 3, 0, 3

Ryhmä B

0, 4, 2, 2, 6, 4, 1, 5

Määritä serkkujen lukumäärän

a) moodi

b) mediaani.

Ratkaisu

a) Moodi on se muuttujan arvo, jota on eniten.

Ryhmä A

2, 2, **3**, 1, **3**, 0, **3**

$$\text{Mo}_A = \mathbf{3}$$

Ryhmä B

0, **4**, **2**, **2**, 6, **4**, 1, 5

$$\text{Mo}_B = \mathbf{2} \text{ ja } \text{Mo}_B = \mathbf{4}$$

b) Mediaani on suuruusjärjestykseen laitetun aineiston keskimäinen arvo.

Ryhmä A

0, 1, 2, **2**, 3, 3, 3

Serkkujen lukumäärän mediaani ryhmässä A on $Md_A = \mathbf{2}$.

Ryhmä B

0, 1, 2, **2**, **4**, 4, 5, 6

$$Md_B = \frac{\mathbf{2} + \mathbf{4}}{2} = 3$$

Serkkujen lukumäärän mediaani ryhmässä B on $Md_B = 3$.

Keskiarvo

Kun puhutaan keskiarvosta, tarkoitetaan yleensä aritmeettista keskiarvoa.

Yksi eniten käytetyistä tunnusluvusta on **keskiarvo** $\bar{x}$.

Keskiarvo kuvaa keskimääräistä muuttujan arvoa.

Se lasketaan jakamalla muuttujan arvojen summa arvojen lukumäärällä.

6, 7, 8, 9, 9

$$\bar{x} = \frac{6 + 7 + 8 + 9 + 9}{5} = 7,8$$

Muuttuja x	f
1	18
2	5
3	12

$$\sum x = 18 \cdot 1 + 5 \cdot 2 + 12 \cdot 3 = 64$$

$$n = 18 + 5 + 12 = 35$$

$$\bar{x} = \frac{\sum x}{n} = \frac{18 \cdot 1 + 5 \cdot 2 + 12 \cdot 3}{35} = \frac{64}{35} \approx 1,8$$

Matematiikan ryhmässä tutkittiin ryhmän opiskelijoiden matematiikan arvosanoja.

- a) Millan matematiikan opinnoista saamat arvosanat ovat 8, 7, 8, 9, 8, 7. Laske Millan arvosanojen keskiarvo yhden desimaalin tarkkuudella.

Ratkaisu

- a) Keskiarvo lasketaan jakamalla muuttujan arvojen summa arvojen lukumäärällä.

$$\bar{x} = \frac{\sum x}{n}$$



$$\bar{x} = \frac{8 + 7 + 8 + 9 + 8 + 7}{6} = \frac{47}{6} = 7,833... \approx 7,8$$

Millan matematiikan arvosanojen keskiarvo on 7,8.

- b) Rinnakkaisryhmän opiskelijoiden saamat arvosanat on esitetty oheisessa taulukossa. Laske arvosanojen keskiarvo yhden desimaalin tarkkuudella.

Ratkaisu

- b) Frekvenssijakaumassa on yhteensä muuttujan arvoja

$$8 + 23 + 34 + 39 + 25 + 19 + 2 = 150$$

Lasketaan keskiarvo frekvenssijakauman avulla.

$$\bar{x} = \frac{\sum f \cdot x}{n}$$

$$\bar{x} = \frac{8 \cdot 10 + 23 \cdot 9 + 34 \cdot 8 + 39 \cdot 7 + 25 \cdot 6 + 19 \cdot 5 + 2 \cdot 4}{150} = \frac{1085}{150} = 7,233... \approx 7,2$$

Arvosana x_i	Frekvenssi f_i
10	8
9	23
8	34
7	39
6	25
5	19
4	2

Rinnakkaisryhmän matematiikan arvosanojen keskiarvo on 7,2.

- c) Pärjäsi Milla keskimäärin paremmin vain huonommin matematiikan opinnoissaan kuin rinnakkaisryhmän opiskelijat? Perustele.

Ratkaisu

- c) Millan matematiikan arvosanojen keskiarvo on 7,8.

Rinnakkaisryhmän matematiikan arvosanojen keskiarvo on 7,2.

Millan arvosanojen keskiarvo on $7,8 > 7,2$, joten Milla pärjäsi matematiikan opinnoissaan keskimäärin paremmin kuin rinnakkaisryhmän opiskelijat.

**Tee tehtävät 5.1,
5.2 ja 5.5**

A Esimerkki 3

Lukiolaisilta tiedusteltiin, kuinka pitkä jonotusaika koulun ruokalassa oli eräänä päivänä lounasaikaan. Tulokset on esitetty taulukossa.

a) Laske jonotusajan keskiarvo.

b) Määritä jonotusajan moodi ja mediaani.

Anna vastaukset yhden desimaalin tarkkuudella.

Jonotusaika (min)	f
0–2	67
3–5	52
6–8	24
9–11	18

Ratkaisu

a) Keskiarvon laskemiseksi tarvitaan luokkien luokkakeskukset.

Lasketaan luokkakeskukset.

$$\text{luokkakeskus} = \frac{\text{todellinen alaraja} + \text{todellinen yläraja}}{2}$$

Odotusaika (min)	Luokkakeskus x (min)	f
0–2	$\frac{0 + 2,5}{2} = 1,25$	67
3–5	$\frac{2,5 + 5,5}{2} = 4,0$	52
6–8	7,0	24
9–11	10	18

Lasketaan keskiarvo luokkakeskusten ja frekvenssien avulla.

$$\bar{x} = \frac{\sum f \cdot x}{n}$$

$$\begin{aligned}\bar{x} &= \frac{67 \cdot 1,25 + 52 \cdot 4,0 + 24 \cdot 7,0 + 18 \cdot 10,0}{67 + 52 + 24 + 18} \\ &= \frac{639,75}{161} = 3,973... \approx 4,0 \text{ (min)}\end{aligned}$$

Laske ensin osoittajassa ja nimittäjässä olevat laskutoimitukset.

Moodiluokka on se luokka, jonka frekvenssi on suurin.

- b) Suurin frekvenssi on luokalla 0–2 (min), joten tämä on moodiluokka. Moodi on tämän luokan luokkakeskus eli $Mo = 1,25 \approx 1,3$ (min).

Mediaaniluokka on se luokka, jossa keskimäinen muuttujan arvo sijaitsee.

Määritetään mediaaniluokka.

- Aineistossa on muuttujan arvoja yhteensä $67 + 52 + 24 + 18 = 161$ kpl.
- Keskimäinen arvo on suuruusjärjestyksen 81. arvo ($161 = 80 + 1 + 80$).
- Kahdessa ensimmäisessä luokassa on yhteensä $67 + 52 = 119$ arvoa.

Suuruusjärjestyksessä 81. arvo sijaitsee siis luokassa 3–5 (min), joten tämä on mediaaniluokka. Mediaani on tämän luokan luokkakeskus.

$$Md = 4,0 \text{ (min)}$$

Luokitellussa aineistossa havaintoja vastaavat luokkakeskukset.

$1,25 \ 1,25 \dots 1,25 \ 4 \ 4 \dots 4 \ 7 \ 7 \dots$
67 havaintoa
 $67 + 52 = 119$ havaintoa

Vastaus

a) 4,0 min

b) $Mo = 1,3$ min ja $Md = 4,0$ min

**Tee tehtävät
5.3 ja 5.7**

Urheiluseurojen välisessä mäkihyppykilpailussa annettiin pisteitä sekä hypyn pituudesta että tyylistä. Tyylipisteitä antoivat viisi tuomaria asteikolla 0–10 pisteen tarkkuudella, ja pisteistä laskettiin keskiarvo yhden desimaalin tarkkuudella. Juho ja Aapo osallistuivat kilpailuun. He saivat molemmat kolmelta ensimmäiseltä tuomarilta tyylipisteet 6, 6 ja 8.

- a) Neljäs tuomari antoi Aapolle vain 4 pistettä, mutta Aapon tyylipisteiden keskiarvo oli kuitenkin 6,4. Kuinka monta tyylipistettä viides tuomari antoi?
- b) Kaksi viimeistä tuomaria antoivat Juholle täsmälleen samat tyylipisteet. Juhon tyylipisteiden keskiarvo oli tällöin 6,8. Mitkä olivat kahden viimeisen tuomarin antamat tyylipisteet?

Ratkaisu

- a) Merkitään viidennen tuomarin antamaa pistemäärää kirjaimella x . Tuomarien antama yhteispistemäärä on

$$6 + 6 + 8 + 4 + x = 24 + x.$$

Muodostetaan keskiarvon avulla yhtälö ja ratkaistaan x .

$$\frac{24 + x}{5} = 6,4 \quad | \cdot 5$$

$$24 + x = 32$$

$$x = 32 - 24$$

$$x = 8$$

Viides tuomari antoi Aapolle 8 pistettä.

Kerro yhtälön molemmat puolet nimittäjässä olevalla luvulla 5.

- b) Merkitään kahden viimeisen tuomarin antamaa pistemäärää kirjaimella x . Tuomarien antama yhteispistemäärä on

$$6 + 6 + 8 + x + x = 20 + 2x.$$

Muodostetaan keskiarvon avulla yhtälö ja ratkaistaan x .

$$\frac{20 + 2x}{5} = 6,8 \quad | \cdot 5$$



Kerro ensin nimittäjä pois.

$$20 + 2x = 34$$

$$2x = 34 - 20$$

$$2x = 14 \quad | : 2$$

$$x = 7$$

Neljäs ja viides tuomari antoivat kumpikin 7 pistettä.

Vastaus

a) 8 pistettä

b) 7 pistettä

**Tee tehtävät
5.10 ja 5.13**