

2 Jatkuva tilastollinen jakauma

2.1 Jatkuvan muuttujan frekvenssijakauma

71.

Summafrekvenssi (sf) ilmaisee, kuinka monta havaintoa yhteensä kuuluu kyseiseen luokkaan ja sitä edeltäviin luokkiin.

Suhteellinen frekvenssi ($f\%$) ilmaisee frekvenssin prosenttiosuuden kaikista havainnoista.

Suhteellinen summafrekvenssi ($sf\%$) ilmaisee summafrekvenssin prosenttiosuuden. Se saadaan myös kyseisen luokan ja sitä edeltävien luokkien suhteellisten frekvenssien ($f\%$) summana.

x	f	sf	$f\%$	$sf\%$
1-3	2	2	$\frac{2}{20} \cdot 100 = 10$	$\frac{2}{20} \cdot 100 = 10$
4-6	5	$2 + 5 = 7$	$\frac{5}{20} \cdot 100 = 25$	$\frac{7}{20} \cdot 100 = 35$
7-9	4	$7 + 4 = 11$	$\frac{4}{20} \cdot 100 = 20$	$\frac{11}{20} \cdot 100 = 55$
10-12	9	$11 + 9 = 20$	$\frac{9}{20} \cdot 100 = 45$	$\frac{20}{20} \cdot 100 = 100$

Viimeisen luokan summafrekvenssi kertoo aineiston koon. Havaintoja on yhteensä 20 eli $N = 20$. Tätä lukumäärää tarvitaan suhteellisten frekvenssien määrittämiseen.

Suhteellinen summafrekvenssi ($sf\%$) saadaan myös laskemalla suhteellisten frekvenssien ($f\%$) summa. Esimerkiksi: $10 + 25 + 20 = 55$.

72.

Annettujen tietojen perusteella saadaan täytettyä osa taulukon puuttuvista tiedoista.

x	f	sf	$f\%$	$sf\%$
0,1–0,5	3	3	10	10
0,6–1,0	$6 - 3 = 3$	6		
1,1–1,5		x		90
1,6–2,0		N		100

Merkitään havaintojen lukumäärää (eli aineiston kokoa) kirjaimella N .

Ensimmäisen luokan frekvenssi on $f = 3$ ja suhteellinen frekvenssi on $f\% = 10\% = 0,1$ eli

$$\frac{3}{N} = 0,1.$$

Ratkaistaan yhtälö: $N = 30$.

Merkitään kolmannen luokan (1,1–1,5) summafrekvenssiä (sf) kirjaimella x . Luokan suhteellinen summafrekvenssi on $90\% = 0,9$ eli

$$\frac{k}{30} = 0,9.$$

Yhtälön ratkaisuna saadaan $k = 27$.

Täytetään puuttuvat tiedot taulukosta.

<i>x</i>	<i>f</i>	<i>sf</i>	<i>f</i> %	<i>sf</i> %
0,1–0,5	3	3	10	10
0,6–1,0	$6 - 3 = 3$	6	$\frac{3}{30} \cdot 100 = 10$	$10 + 10 = 20$
1,1–1,5	$27 - 6 = 21$	27	$\frac{21}{30} \cdot 100 = 70$	90
1,6–2,0	$30 - 27 = 3$	30	$\frac{3}{30} \cdot 100 = 10$	100

Suhteellinen summafrekvenssi ($f\%$)
saadaan myös laskemalla vastaavan
summafrekvenssin (sf) prosenttiosuus
kaikista havainnoista.

$$\frac{6}{30} \cdot 100 = 20$$

73.

a)

Taimen pituus (cm)	Todellinen alaraja (cm)	Todellinen yläraja (cm)	Luokkakeskus
5-10	4,5	10,5	$\frac{4,5 + 10,5}{2} = 7,5$
11-16	10,5	16,5	$\frac{10,5 + 16,5}{2} = 13,5$
17-22	16,5	22,5	$\frac{16,5 + 22,5}{2} = 19,5$

LISÄTIETO:

Todelliset luokkarajat määräytyvät mittaustarkkuudesta.

- todellinen alaraja = alaraja - 0,5 · mittaustarkkuus
- todellinen yläraja = yläraja + 0,5 · mittaustarkkuus

Tässä on oletettu, että mittaustarkkuus on 1 cm.

b) Huomaa, että taimen pituus on positiivinen luku eli ensimmäisen luokan todellinen alaraja on 0 (cm).

Taimen pituus (cm)	Todellinen alaraja (cm)	Todellinen yläraja (cm)	Luokkakeskus
0-10	0	10,5	$\frac{0 + 10,5}{2} = 5,25$
11-21	10,5	21,5	$\frac{10,5 + 21,5}{2} = 16$
22-32	21,5	32,5	$\frac{21,5 + 32,5}{2} = 27$

74.

a) Vaihteluvälin pituus on $198 - 147 = 51$ (cm).

Luokkien lukumäärä on kuusi. Jaetaan vaihteluvälin pituus luokkien lukumäärällä.

$$\frac{51}{6} = 8,5$$

Pyöristetään saatu luku sopivasti ylöspäin, esimerkiksi lukuun 10. Sopiva luokkaväli on esimerkiksi 10 cm.

Valitaan ensimmäisen luokan alarajaksi jokin sopiva luku pienimmän arvon 147 cm alapuolelta, esimerkiksi arvo 145 cm (myös pienin arvo 147 cm on mahdollinen valinta).

Luokkavälin pituus on 10 cm, joten seuraavan luokan alaraja on
 $145 \text{ cm} + 10 \text{ cm} = 155 \text{ cm}.$

Ensimmäisen luokan yläraja on tällöin 154 cm.

Ensimmäinen luokka on $145 \text{ cm} - 154 \text{ cm}.$

Luokitus on tällöin

145 cm – 154 cm

155 cm – 164 cm

165 cm – 174 cm

175 cm – 184 cm

185 cm – 194 cm

195 cm – 204 cm

Pienin arvo 147 cm kuuluu ensimmäiseen eli alimpaan luokkaan.

Suurin arvo 198 cm kuuluu viimeiseen eli ylimpään luokkaan.

b) Vaihteluvälin pituus on $3200 - 452 = 2748$ (€).

Luokkien lukumäärä on kuusi. Jaetaan vaihteluvälin pituus luokkien lukumäärällä.

$$\frac{2748}{6} = 458$$

Pyöristetään saatu luku sopivasti ylöspäin, esimerkiksi lukuun 460. Sopiva luokkaväli on esimerkiksi 460 €.

Valitaan ensimmäisen luokan alarajaksi jokin sopiva luku pienimmän arvon 452 € alapuolelta, esimerkiksi arvo 450 €.

Luokkavälin pituus on 460 €, joten seuraavan luokan alaraja on

$$450 \text{ €} + 460 \text{ €} = 910 \text{ €}.$$

Ensimmäisen luokan yläraja on tällöin 909 €.

Ensimmäinen luokka on 450 € – 909 €.

Luokitus on tällöin

450 € – 909 €

910 € – 1369 €

1370 € – 1829 €

1830 € – 2289 €

2290 € – 2749 €

2750 € – 3209 €

Pienin arvo 452 € kuuluu ensimmäiseen eli alimpaan luokkaan.

Suurin arvo 3200 € kuuluu viimeiseen eli ylimpään luokkaan.

75.

Kirjoitetaan havainnot esimerkiksi matematiikkaohjelmiston taulukkosovellukseen ja tehdään yhden muuttujan analyysi.

a) Tilastojen yhteenvetotaulukosta nähdään, että aineiston koko on $N = 20$.

Pienin havainto on $\min = 324$ (l/min) ja suurin havainto on $\max = 618$ (l/min).

Muuttujan arvojen vaihteluväli on 324 l/min – 618 l/min.

b) Ensimmäisen luokan alarajaksi on valittu **320** (l/min).

- Todellinen alaraja on **319,5** (l/min).
- Luokkavälin pituudeksi on valittu **60** (l/min).

Seuraavan luokan alaraja on

$$320 + 60 = \mathbf{380} \text{ (l/min)}.$$

Edellisen luokan ylärajan tulee olla pienempi kuin seuraavana oleva alaraja. Ensimmäisen luokan yläraja on siis **379**.

Saadaan taulukon mukainen luokitus.

Uloshengityksen huippuarvo (l/min)	Frekvenssi
320–379	4
380–439	4
440–499	3
500–559	4
560–619	5

Matematiikkaohjelmisto tekee automaattisesti luokittelun ja laskee luokkien frekvenssit, kun sovellukseen syötetään:

- luokituksen alku 319,5
- luokan leveys 60.

Muista tarkistaa:

Pienin arvo 324 l/min kuuluu alimpaan luokkaan ja suurin arvo 618 l/min ylimpään luokkaan.

c) Todelliset ala- ja ylärajat nähdään sovelluksen tuottamasta frekvenssitaulukosta, kun sovellukseen on syötetty luokituksen alku (319,5) sekä luokkavälin leveys (60).

Luokkakeskukset kannattaa määrittää todellisten luokkarajojen avulla laskentataulukossa soluviittauksilla.

Uloshengityksen huippuarvo (l/min)	Todellinen alaraja (l/min)	Todellinen yläraja (l/min)	Luokkakeskus (l/min)
320–379	319,5	379,5	$\frac{319,5 + 379,5}{2} = 349,5$
380–439	379,5	439,5	409,5
440–499	439,5	499,5	469,5
500–559	499,5	559,5	529,5
560–619	559,5	619,5	589,5

76.

Kirjoitetaan tai kopioidaan havainnot esimerkiksi matematiikkaohjelmiston taulukkosovellukseen ja tehdään yhden muuttujan analyysi.

a) Tilastojen yhteenvetotaulukosta nähdään, että aineiston koko on $N = 21$.

Pienin havainto on $\min = 613$ (cm) ja suurin havainto on $\max = 776$ (cm).

Vaihteluvälin pituus on $776 - 613 = 163$ (cm).

Luokkien lukumääräksi on valittu kuusi. Jaetaan vaihteluvälin pituus luokkien lukumäärällä.

$$\frac{163}{6} = 27,166 \dots$$

Pyöristetään saatu luku sopivasti ylöspäin, esimerkiksi lukuun 30.

➤ Valitaan luokkaväliksi 30 (cm).

Valitaan ensimmäisen luokan alarajaksi jokin sopiva arvo, joka on korkeintaan aineiston pienin arvo 613 cm, esimerkiksi 613 cm.

➤ Todellinen alaraja on 612,5 cm.

Luokkavälin pituudeksi valittiin 30 (cm), joten seuraavan luokan alaraja on

$$613 + 30 = 643 \text{ (cm)}.$$

Ensimmäisen luokan yläraja on tällöin 642 (cm).

Saadaan taulukon mukainen luokitus.

Pituushypyn tulos (cm)	Frekvenssi
613-642	4
643-672	2
673-702	6
703-732	4
733-762	3
763-792	2

Matematiikkaohjelmisto tekee automaattisesti luokittelun ja laskee luokkien frekvenssit, kun sovellukseen syötetään:

- luokituksen alku 612,5 ja
- luokan leveys 30.



Muista tarkistaa:
Pienin arvo 613 cm kuuluu alimpaan luokkaan ja suurin arvo 776 cm ylimpään luokkaan.



c) Todelliset ala- ja ylärajat nähdään sovelluksen tuottamasta frekvenssitaulukosta, kun sovellukseen on syötetty luokituksen alku (609,5) sekä luokkavälin leveys (30).

Luokkakeskukset kannattaa määrittää todellisten luokkarajojen avulla laskentataulukossa soluviittauksilla.

Pituushypyn tulos (cm)	Todellinen alaraja (cm)	Todellinen yläraja (cm)	Luokkakeskus (cm)
613–642	612,5	642,5	$\frac{612,5 + 642,5}{2} = 627,5$
643–672	642,5	672,5	657,5
673–702	672,5	702,5	687,5
703–732	702,5	732,5	717,5
733–762	733,5	762,5	747,5
763–792	763,5	792,5	777,5

77.

Kirjoitetaan tai kopioidaan havainnot esimerkiksi matematiikkaohjelmiston taulukkosovellukseen ja tehdään yhden muuttujan analyysi.

a) Aineiston koko on $N = 27$. Aineistossa on siis mukana 27 EU-maata.

Pienin havainto on $\min = 34,5$ (%) ja suurin havainto on $\max = 97,9$ (%).

Vaihteluvälin pituus on

$$97,9 - 34,5 = 63,4 \text{ (\%-yksikköä)}.$$

Huomaa: Muuttujan arvot on ilmaistu prosenttilukuina, mutta vaihteluvälin pituus eli kahden prosenttiluvun erotus on 63,4 **prosenttiyksikköä**.

b) Alimman eli ensimmäisen luokan alarajaksi on valittu **34** (%). Todellinen alaraja on **33,5** (%).

Alimman luokan yläraja on **43** (%), joten seuraavan luokan alaraja on **44** (%) ja luokkavälin pituus on

$$44 - 34 = **10** \text{ (\%-yksikköä)}$$

Saadaan taulukon mukainen luokitus.

Opiskelevien osuus 18-vuotiaista (%)	Frekvenssi
34–43	2
44–53	0
54–63	0
64–73	3
74–83	7
84–93	10
94–100	5

Matematiikkaohjelmisto tekee automaattisesti luokittelun ja laskee luokkien frekvenssit, kun sovellukseen syötetään:

- luokituksen alku **33,5**
- luokan leveys **10**

Muista tarkistaa:

Pienin arvo 34,5 % kuuluu alimpaan luokkaan ja suurin arvo 97,9 % ylämpään luokkaan.

Huomaa: Prosenttiosuus on korkeintaan 100, joten viimeisen luokan (todellinen) yläraja on 100.

Huomaa: Jos b-kohdassa valitaan alimman luokan todelliseksi alarajaksi arvo **34** (%), luokkien frekvenssit poikkeavat taulukossa ilmoitetuista. Tällöin luokitus ei kuitenkaan ole tasavälinen, sillä ensimmäisen luokan luokkavälin pituus on $43,5 - 34 = 9,5$, kun muiden luokkien leveys on **10**.

c) Ensimmäisen luokan frekvenssi on 2, mutta seuraavat kaksi luokkaa ovat tyhjiä (frekvenssi on 0). Aineistossa on siis kaksi poikkeuksellisen pientä havaintoa. Muut $27 - 2 = 25$ havaintoa sijoittuvat neljään ylimpään luokkaan eli välille 64 % – 100 %.

Siis, kahdessa EU-maassa 18-vuotiaiden opiskelevien nuorten prosenttiosuus ikäluokastaan on poikkeuksellisen pieni (aineiston pienimmät arvot: 34,5 % ja 40,7 %). Muissa EU-maissa opiskelevien nuorten osuus ikäluokasta on vähintään 64 %.

d) Todelliset ala- ja ylärajat nähdään sovelluksen tuottamasta frekvenssitaulukosta, kun sovellukseen on syötetty luokituksen alku (33,5) sekä luokkavälin leveys (10).

Luokkakeskukset kannattaa määrittää todellisten luokkarajojen avulla laskentataulukossa soluviittauksilla.

	A	B	C	D
1	Luokka (%)	Todellinen alaraja	Todellinen yläraja	Luokkakeskus
2	34-43	33.5	43.5	38.5
3	44-53	43.5	53.5	48.5
4	54-63	53.5	63.5	58.5
5	64-73	63.5	73.5	68.5
6	74-83	73.5	83.5	78.5
7	84-93	83.5	93.5	88.5
8	94-100	93.5	100	96.75

Soluun D2 on kirjoitettu laskukaava:

$$=(B2+C2)/2$$
 Kaava on kopioitu sarakkeessa D alaspäin.

78.

a) Laaditaan frekvenssijakaumat taulukkolaskentasovelluksella.

	A	B	C	D	E
1	Nopeus (km/h)	f	sf	f(%)	sf(%)
2	80-85	3	3	2.113	2.113
3	86-91	10	13	7.042	9.155
4	92-97	30	43	21.127	30.282
5	98-103	46	89	32.394	62.676
6	104-109	39	128	27.465	90.141
7	110-115	14	142	9.859	100

Soluun **C3** on kirjoitettu laskukaava:
=C2+B3
Kaava on kopioitu sarakkeessa C alaspäin.

Soluun **D2** on kirjoitettu laskukaava:
=B2/142*100
Kaava on kopioitu sarakkeessa D alaspäin.

Soluun **E2** on kirjoitettu laskukaava:
=C2/142*100
Kaava on kopioitu sarakkeessa E alaspäin.

Pyöristetään suhteelliset frekvenssit prosentin tarkkuuteen.

Nopeus (km/h)	f	sf	f (%)	sf (%)
80–85	3	3	2	2
86–91	10	13	7	9
92–97	30	43	21	30
98–103	46	89	32	63
104–109	39	128	27	90
110–115	14	142	10	100

b) Viimeisen luokan summafrekvenssistä (solu C7) nähdään, että aineistossa on yhteensä **142** havaintoa eli $N = 142$. Mittauspisteen ohi ajoi mittauksen aikana 142 autoa.

c) Luokan 98–103 (km/h) todellinen yläraja on 103,5 km/h. Niiden autojen, joiden nopeus oli alle 103,5 km/h, lukumäärä nähdään tämän luokan summafrekvenssistä (solu C5). Autoja oli yhteensä 89 kappaletta.

d) Luokan 98–103 (km/h) todellinen alaraja on 97,5 km/h. Autot, joiden nopeus on vähintään 97,5 km/h kuuluvat kolmeen ylimpään luokkaan. Lasketaan näiden autojen prosenttiosuus luokkien suhteellisten frekvenssien ($f\%$) avulla.

$$32,39\ldots \% + 27,46\ldots \% + 9,85\ldots \% = 69,71\ldots \% \approx 70 \%$$



Lasku voidaan tehdä laskentataulukossa kirjoittamalla sopivaan tyhjään soluun laskukaava:
 $=D5 + D6 + D7$

79.

Tehtävä voidaan ratkaista ladattavan aineiston Excel-taulukossa tai aineisto voidaan ratkaisua varten kopioida toisen ohjelmiston taulukko-sovellukseen.

a) Etsitään luokka, jonka todellinen alaraja on 46,25 kg: tämä on luokka 46,3–49,4 (kg). Asiakkaat, joiden tulos oli vähintään 46,25 kg kuuluvat neljään ylimpään luokkaan. Näitä asiakkaita on

$14 + 10 + 4 + 1 = 29$ (asiakasta).

Puristusvoima	f
30,3 – 33,4	1
33,5 – 36,6	2
36,7- 39,8	15
39,9 – 43,0	10
43,1 – 46,2	12
46,3 – 49,4	14
49,5 – 52,6	10
52,7 – 55,8	4
55,9 – 59,0	1

b) Määritetään suhteelliset frekvenssit laskentataulukkoon.

	A	B	C	D
1	Puristusvoima	f	sf	sf (%)
2	30,3 – 33,4	1	1	1,449
3	33,5 – 36,6	2	3	4,348
4	36,7- 39,8	15	18	26,087
5	39,9 – 43,0	10	28	40,580
6	43,1 – 46,2	12	40	57,971
7	46,3 – 49,4	14	54	78,261
8	49,5 – 52,6	10	64	92,754
9	52,7 – 55,8	4	68	98,551
10	55,9 – 59,0	1	69	100,000

Soluun D2 on kirjoitettu laskukaava: =C2/69*100
Kaava on kopioitu sarakkeessa D alaspäin.

Soluun C3 on kirjoitettu laskukaava: =C2 + B3
Kaava on kopioitu sarakkeessa C alaspäin.

Viimeisen luokan summafrekvenssistä nähdään, että aineistossa on yhteensä 69 havaintoa eli N = 69.

Etsitään luokka, jonka todellinen yläraja on 52,65 kg: tämä on luokka 49,5–52,6 (kg). Asiakkaat, joiden tulos oli alle 52,65 kg kuuluvat 7 alimpaan luokkaan. Näiden asiakkaisen prosenttiosuus nähdään luokan 49,5–52,6 suhteellisesta summafrekvenssistä. Asiakkaita on 92,75... % $\approx$ 93 %.

Puristusvoima	f	sf	sf (%)
30,3 – 33,4	1	1	1,449
33,5 – 36,6	2	3	4,348
36,7- 39,8	15	18	26,087
39,9 – 43,0	10	28	40,580
43,1 – 46,2	12	40	57,971
46,3 – 49,4	14	54	78,261
49,5 – 52,6	10	64	92,754
52,7 – 55,8	4	68	98,551
55,9 – 59,0	1	69	100,000

c) Asiakkaat, joiden tulos oli vähintään 39,85 kg mutta alle 49,45 kg, kuuluvat luokkiin 39,9–43,0 tai 43,1–46,2 tai 46,3–49,4. Näihin luokkiin kuuluvia asiakkaita oli yhteensä

$$10 + 12 + 14 = 36 \text{ (asiakasta)}$$

eli prosentteina

$$\frac{36}{69} = 0,521 \dots \approx 52 \text{ \%}.$$

Puristusvoima	f	sf	sf (%)
30,3 – 33,4	1	1	1,449
33,5 – 36,6	2	3	4,348
36,7- 39,8	15	18	26,087
39,9 – 43,0	10	28	40,580
43,1 – 46,2	12	40	57,971
46,3 – 49,4	14	54	78,261
49,5 – 52,6	10	64	92,754
52,7 – 55,8	4	68	98,551
55,9 – 59,0	1	69	100,000

80.

a) Lasketaan frekvenssien summa laskin- tai matematiikkaohjelmiston taulukkosovelluksessa.

	A	B
1	Ikä	Lkm/1000 henkilöä
2	15-24	50
3	25-34	57
4	35-44	36
5	45-54	47
6	55-64	48
7		238

Soluun B7 on laskettu yllä olevissa soluissa olevien frekvenssien summa. Frekvenssit ilmaisevat tuhansia ihmisiä, joten summa on 238 000 ihmistä.

Aineistossa on yhteensä 238 000 havaintoa, eli työttömiä oli tammikuussa 2018 yhteensä 238 000.

b) Lasketaan suhteelliset frekvenssit. Frekvenssit jaetaan havaintojen kokonaismäärällä eli frekvenssien summalla, joka määritettiin aikohdassa ($N = 238$).

	A	B	C
	Ikä	Lkm/1000 henkilöä	f %
1			
2	15-24	50	21,00840
3	25-34	57	23,94958
4	35-44	36	15,12605
5	45-54	47	19,74790
6	55-64	48	20,16807
7		238	

Soluun C2 on kirjoitettu laskukaava $=B2/238*100$
Kaava on kopioitu sarakkeessa C alaspäin

Aineistossa on yhteensä 238 (tuhatta) havaintoa eli $N = 238\ 000$.

Pyöristetään prosenttiluvut yhden desimaalin tarkkuuteen.

Ikä	f %
15-24	21,0
25-34	23,9
35-44	15,1
45-54	19,7
55-64	20,2

Työttömät muodostavat perusjoukon, jonka koko on $N = 238\ 000$. Suhteelliset frekvenssit ilmaisevat kuhunkin ikäluokkaan kuuluvien työttömien prosenttiosuuden kaikista työttömistä.

Jos halutaan vertailla työttömien ikärakennetta eri perusjoukoissa, esimerkiksi eri vuosina tai eri maissa, perusjoukkojen koko voi poiketa toisistaan suuresti. Tällöin prosenttiosuudet sopivat vertailuun paremmin kuin absoluuttiset lukumäärät.

c) Ikä ilmaistaan alaspäin pyöristettynä kokonaislukuna. Esimerkiksi luokkaan 35–44 kuuluvat ne henkilöt, jotka ovat täyttäneet 35 vuotta mutta ovat alle 45-vuotiaita. Luokan todelliset luokkarajat ovat siis 35 ja 45.

Vähintään 45-vuotiaat kuuluvat kahteen ylimpään luokkaan. Näihin luokkiin kuuluvia havaintoja on yhteensä

$$47 + 48 = 95 \text{ (tuhatta).}$$

Työttömistä vähintään 45-vuotiaita oli yhteensä 95 000.

d) Alle 35-vuotiaat kuuluvat kahteen alimpaan ryhmään. Näihin luokkiin kuuluvia havaintoja on yhteensä

$$21,008 \dots \% + 23,950 \dots \% = 44,958 \dots \% \approx 45,0 \%$$

	A	B	C	D
1	Ikä	Lkm/1000 henkilöä	f %	
2	15-24	50	21,00840	
3	25-34	57	23,94958	44,958
4	35-44	36	15,12605	
5	45-54	47	19,74790	
6	55-64	48	20,16807	

Summa voidaan laskea laskentataulukossa kirjoittamalla sopivaan tyhjään soluun laskukaava
=C2+C3

81.

Luokitellun aineiston histogrammi voidaan piirtää esimerkiksi matematiikkaohjelmiston taulukkosovelluksella, kun luokat ovat tasalevyisiä.

Kirjoitetaan tai kopioidaan suhteelliset frekvenssit ($f\%$) matematiikkaohjelmiston taulukkosovellukseen.

	A
1	$f\%$
2	25
3	10
4	15
5	40
6	5
7	5

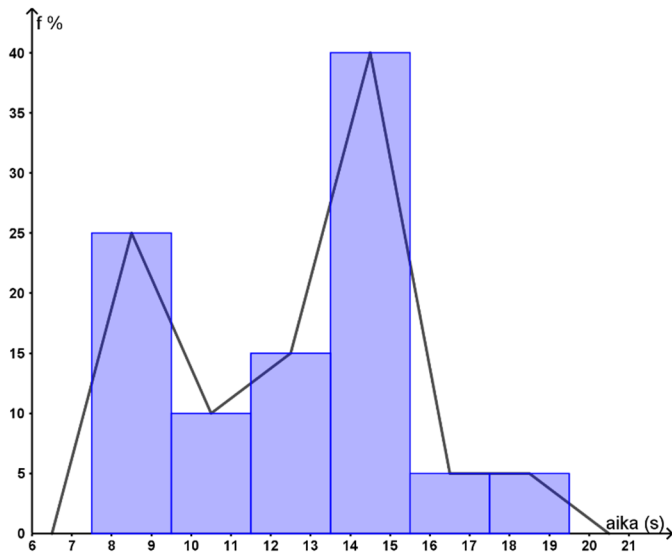
Määritetään luokituksen alku ja luokkavälin leveys:

1. Ensimmäisen luokan 8–9 (s) todellinen alaraja on **7,5** (s).
2. Luokkavälin leveys on kahden peräkkäisen luokan alarajojen (tai ylärajojen) erotus eli $10 - 8 =$ **2**.

Tehdään suhteellisille frekvensseille yhden muuttujan analyysi ja valitaan asetuksista vaihtoehto ”Luokat ja frekvenssi”. Asetetaan:

- luokituksen alku: **7,5**
- luokan leveys: **2**.

Ohjelmisto kuvaa jakauman histogrammilla. Lisätään asetuksista vaihtoehto ”Frekvenssimonikulmio”.



Kopioidaan kuvaaja piirtoalueelle muokkaamista varten.

- Siirretään y -akseli vaakasuunnassa sopivaan kohtaan, esimerkiksi kohtaan $x = 6$.
- Lisätään akseleille nimet ja sopiva asteikko.

82.

Laaditaan pyydetty taulukko esimerkiksi matematiikkaohjelmiston taulukkosovelluksella. Luokkakeskukset lasketaan todellisten luokkarajojen avulla, keskiarvona.

Soluun D2 on kirjoitettu laskukaava: $= (B2 + C2) / 2$
Kaava on kopioitu sarakkeessa D alaspäin.

	A	B	C	D	E	F
1	Pituus (cm)	tod.alaraja	tod.yläraja	luokkakeskus	f	f %
2	5-10	4.5	10.5	7.5	15	13.64
3	11-16	10.5	16.5	13.5	20	18.18
4	17-22	16.5	22.5	19.5	33	30
5	23-28	22.5	28.5	25.5	27	24.55
6	29-34	28.5	34.5	31.5	10	9.09
7	35-40	34.5	40.5	37.5	5	4.55
8					110	

Soluun E8 on laskettu sarakkeessa yllä olevien solujen eli frekvenssien summa. Aineiston koko on $N = 110$.

Suhteelliset frekvenssit on laskettu sarakkeeseen F. Soluun F2 on kirjoitettu laskukaava: $= E2 / 110 * 100$

a) Pyöristetään suhteelliset frekvenssit prosentin tarkkuuteen.

Pituus (cm)	Luokkakeskus (cm)	f (%)
5-10	7,5	14
11-16	13,5	18
17-22	19,5	30
23-28	25,5	25
29-34	31,5	9
35-40	37,5	5

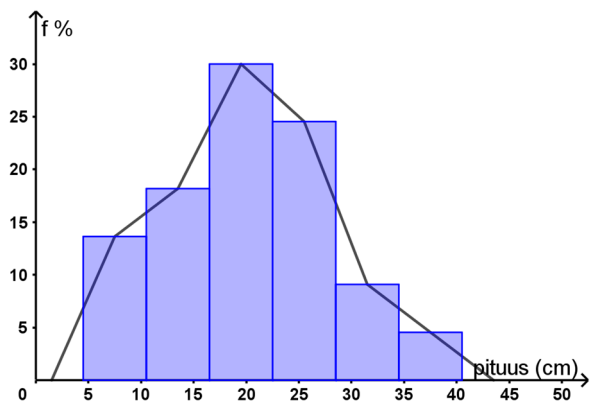
b) Luokitus alkaa ensimmäisen luokan todelliselta alarajalta, joka on 4,5.

Luokkavälin leveys on kahden peräkkäisen luokan alarajojen (tai ylärajojen) erotus eli $11 - 5 = 6$.

Tehdään suhteellisille frekvensseille (sarake F) yhden muuttujan analyysi ja valitaan asetuksista vaihtoehto "Luokat ja frekvenssi". Asetetaan:

- luokituksen alku: 4,5
- luokan leveys: 6

Ohjelmisto kuvaa jakauman histogrammilla. Lisätään asetuksista vaihtoehto "Frekvenssimonikulmio".



Kopioidaan kuvaaja piirtoalueelle muokkaamista varten.

- Lisätään akseleille nimet ja sopiva asteikko.

83.

a) Luokat ovat tasavälisiä. Luokkien todelliset ala- ja ylärajat luetaan kuvaajasta pylväiden reunojen kohdalta. Esimerkiksi ensimmäisen luokan todellinen alaraja on 56,5 kg ja todellinen yläraja 59,5 kg.

Luokkavälin pituus on $59,5 - 56,5 = 3$ (kg).

b) Luetaan kuvaajasta luokkien todelliset ala- ja ylärajat. Luokkakeskukset lasketaan todellisten luokkarajojen keskiarvoina. Luokkia vastaavat frekvenssit luetaan pystyakselilta pylvään korkeuden kohdalta.

Paino (kg)	tod.alaraja	tod.yläraja	luokkakeskus	f
57 - 59	56.5	59.5	58	3
60 - 62	59.5	62.5	61	6
63 - 65	62.5	65.5	64	5
66 - 68	65.5	68.5	67	2
69 - 71	68.5	71.5	70	1

c) Luokalla 60–62 (kg) on suurin frekvenssi ($f = 6$), joten tämä on moodiluokka. Moodina voidaan ilmoittaa moodiluokan luokkakeskus eli $M_o = 61$ kg. Moodiluokan luokkakeskus antaa mielikuva siitä, minkä arvon läheisyyteen suurin osa muuttujan arvoista sijoittuu.

Tyypillisin paino ryhmässä oli 61 kg.

84.

a) Luetaan kuvaajasta luokkien todelliset ala- ja ylärajat.

Luokkakakeskukset lasketaan todellisten luokkarajojen keskiarvoina.

Luokkia vastaavat suhteelliset frekvenssit luetaan pystyakselilta pylvään korkeuden kohdalta.

Nopeus (km/h)	tod.alaraja	tod.yläraja	luokkakeskus	f %
80 - 84	79.5	84.5	82	2
85 - 89	84.5	89.5	87	7
90 - 94	89.5	94.5	92	22
95 - 99	94.5	99.5	97	34
100 - 104	99.5	104.5	102	25
105 - 109	104.5	109.5	107	10

b) Luokalla 95–99 (km/h) on suurin suhteellinen frekvenssi ($f\% = 34$), joten tämä on moodiluokka eli tyyppiarvoluokka. Moodina eli tyyppiarvona voidaan ilmoittaa moodiluokan luokkakeskus eli $M_o = 97$ km/h. Moodiluokan luokkakeskus antaa mielikuva siitä, minkä arvon läheisyyteen suurin osa muuttujan arvoista eli autojen nopeuksista sijoittuu.

Nopeuden tyyppiarvo oli 97 km/h.

c) Vähintään nopeutta 100 km/h ajoivat autot, jotka kuuluvat kahteen ylimpään luokkaan. Näihin luokkiin kuuluvien autojen prosenttiosuus on yhteensä

$$25 \% + 10 \% = 35 \%$$

Mittauspisteen ohitti yhteensä 142 autoa, joten vähintään nopeutta 100 km/h ajavia autoja oli

$$0,35 \cdot 142 = 49,7 \approx 50 \text{ (autoa).}$$

85.

Kirjoitetaan tai kopioidaan iät matematiikkaohjelmiston taulukkosovellukseen. Tehdään yhden muuttujan analyysi.

Yhteenvetotaulukosta nähdään mm.:

- vastaajia on yhteensä 30 ($N = 30$)
- nuorin vastaaja on 18-vuotias ($\min = 18$)
- vanhin vastaaja on 40-vuotias ($\max = 40$)

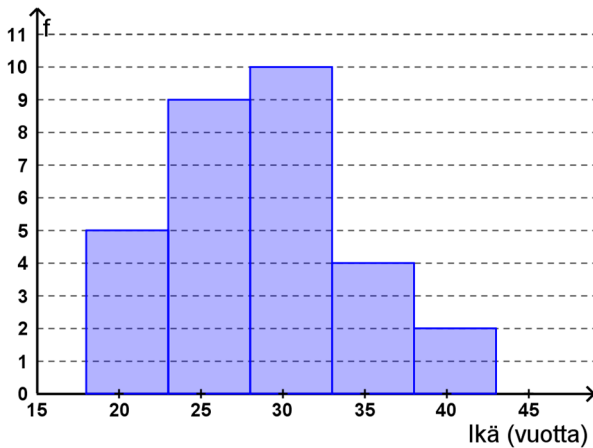
a) Muutetaan luokittelua asettamalla luokat käsin. Luokkien lukumääräksi on valittu viisi. Vaihteluvälin pituus on $40 - 18 = 22$ (vuotta).

$$\frac{22}{5} = 4,4 \approx 5$$

Valitaan ensimmäisen luokan alarajaksi vastaajien pienin ikä 18, joka on myös luokan todellinen alaraja. Valitaan luokkavälin leveydeksi 5. Ohjelmistoon syötetään:

- luokituksen alku: 18
- luokkavälin leveys: 5

Matematiikkaohjelmisto tekee frekvenssitaulukon ja kuvaa jakauman histogrammilla.



Ikä (vuotta)	<i>f</i>
18–22	5
23–27	9
28–32	10
33–37	4
38–42	2

Ohjelmiston frekvenssitaulukossa ovat luokkien todelliset ala- ja ylärajat. Esimerkiksi ensimmäisen luokan todellinen alaraja on 18 ja todellinen yläraja on 23, mutta frekvenssitaulukossa luokka ilmoitetaan välinä 18–22. Luokkaan kuuluvat siis vastaajat, jotka ovat täyttäneet 18-vuotta mutta ovat alle 23-vuotiaita.

b) Vastaajien ikä vaihtelee välillä 18–40 vuotta. Suurin osa vastaajista ($9 + 10 = 19$ vastaajaa 30:stä) on iältään 23–32 vuotta.

86.

Moodi on se muuttujan arvo, jolla on suurin frekvenssi. Arvo 85,3 esiintyy kaksi kertaa ja muut havainnot vain kerran.

Siis, $Mo = 85,3$ (dB).

Asetetaan havainnot suuruusjärjestykseen. Tämä voidaan tehdä esimerkiksi laskentataulukossa Lajittele-toiminnolla.

	A
1	75,2
2	79,2
3	82,4
4	84,3
5	85,3
6	85,3
7	86,2

Mediaani on suuruusjärjestykseen asetetussa aineistossa keskimäinen arvo.

Siis, $Md = 84,3$ (dB)

87.

Määritetään viiden luvun yhteenveto esimerkiksi laskimen tilastotoiminnoilla.

Viiden luvun yhteenvedoksi saadaan

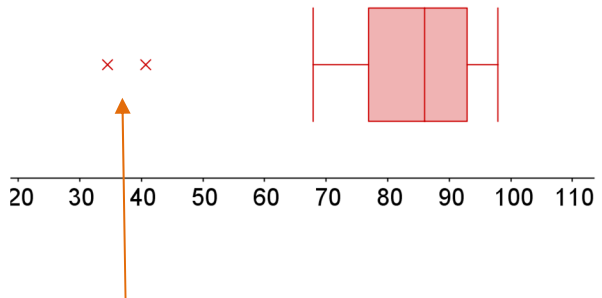
min = 34,5 %

Q_1 = 76,9 %

M_d = 86 %

Q_3 = 92,9 %

max = 97,9 %



Matematiikkaohjelmisto kuvaa viiden luvun yhteenvedon laatikkokuviolla. Kuviosta nähdään, että aineistossa on kaksi poikkeuksellisen pientä havaintoa. Poikkeavat havainnot näkyvät laatikkokuviossa rasteina (X).

88.

a) Asumismenot olivat pienimmillään $\min = 250$ € ja suurimmillaan $\max = 780$ €.

Vastaajat käyttivät asumiseen 250–780 euroa/kk.

b) Asumismenojen mediaani oli 420 €.

Puolet vastaajista käytti asumiseen korkeintaan 420 euroa/kk.

c) Asumismenojen yläkvartiili eli yläneljännes oli $Q_3 = 540$ €.

Neljännes vastaajista käytti asumiseen yli 540 euroa/kk.

d) Asumismenojen kvartiiliväli oli 390 € – 540 €. Kvartiilivälillä on puolet havainnoista.

Puolet vastaajista käytti asumiseen 390–540 euroa/kk.

e) Sana ”keskimäärin” viittaa yleensä (aritmeettiseen) keskiarvoon. Asumismenojen keskiarvo oli 408 €.

Joissakin tilanteissa ”keskimääräisellä” arvolla voidaan tarkoittaa myös mediaania, eli keskimmäistä havaintoa. Asumismenojen mediaani oli 420 €.

Keskimäärin vastaajat käyttivät asumiseen 408 euroa/kk.

TAI

Keskimäärin vastaajat käyttivät asumiseen 420 euroa/kk.

89.

a) Moodiluokka on se luokka (tai luokat), jonka frekvenssi (f tai $f\%$) on suurin.

Luokalla 86–87 (dB) on suurin frekvenssi, joten se on moodiluokka.

Mediaaniluokka on se luokka, jonka suhteellinen summafrekvenssi ($sf\%$) ensimmäisen kerran ylittää arvon 50 (%). Muodostetaan äänenvoimakkuuden summafrekvenssijakauma.

Äänen voimakkuus (dB)	f	sf	$sf\%$
78–79	3	3	10
80–81	4	7	23,33...
82–83	6	13	43,33...
84–85	7	20	66,66...
86–87	9	29	96,66...
88–89	1	30	100

Esimerkiksi:

$$\frac{7}{30} \cdot 100 = 23,33 \dots$$

Viimeisen luokan
summafrekvenssistä nähdään, että
havaintoja on yhteensä 30.

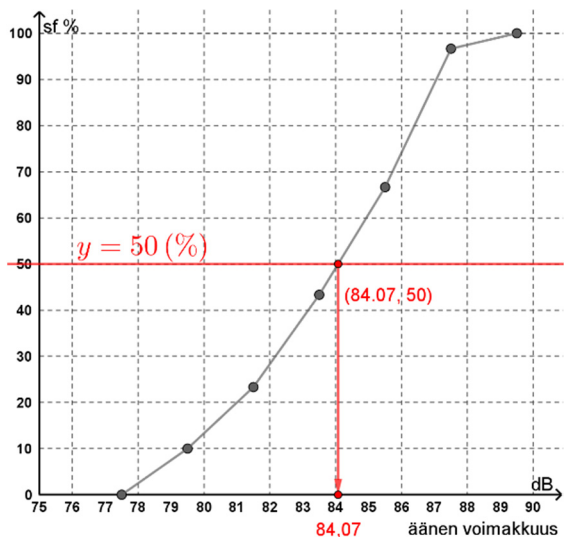
Frekvenssijakaumataulukon perusteella mediaaniluokka on 84–85 (dB). Äänenvoimakkuuden mediaani on siis välillä 84 dB – 85 dB.

b) Piirretään summakäyrä taulukkolaskentasovelluksessa frekvenssipisteiden avulla. Frekvenssipisteet muodostuvat luokkien todellisista ylärajoista ja luokkia vastaavista summafrekvensseistä (sf tai $sf\%$). Summakäyrä alkaa nollatasolta ensimmäisen luokan todellisen alarajan 77,5 dB kohdalta.

Piirretään suhteellisten summafrekvenssien ($sf\%$) summakäyrä: tällöin mediaani luetaan kertymän 50 (%) kohdalta.

Muodostetaan frekvenssipisteistä taulukko ja piirretään sen perusteella viivakaavio.

	A	B
	Äänen voimakkuus (dB)	$sf\%$
1		
2	77,5	0
3	79,5	10
4	81,5	23,33
5	83,5	43,33
6	85,5	66,66
7	87,5	96,66
8	89,5	100



Summakäyrästä arvioidaan mediaaniksi $Md \approx 84$ dB.

Huomautus: Mediaani voidaan arvioida myös summafrekvenssien (sf) mukaan piirretystä summakäyrästä. Tällöin mediaani luetaan sen frekvenssin kohdalta, joka vastaa puolta (50 %) kaikista havainnoista, eli arvon $\frac{30}{2} = 15$ kohdalta.

90.

a) Lasketaan summafrekvenssit taulukkolaskentasovelluksessa.

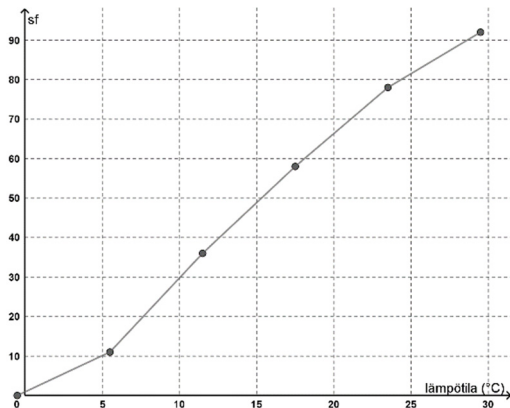
	A	B	C
1	Lämpötila (°C)	f	sf
2	0-5	11	11
3	6-11	25	36
4	12-17	22	58
5	18-23	20	78
6	24-29	14	92

Summakäyrä piirretään taulukkolaskentasovelluksessa frekvenssipisteiden avulla. Frekvenssipisteet muodostuvat luokkien todellisista ylärajoista ja luokkia vastaavista summafrekvensseistä (*sf*).

Summakäyrä alkaa nollatasolta ensimmäisen luokan todellisen alarajan -0.5 °C kohdalta.

Muodostetaan frekvenssipisteistä taulukko ja piirretään sen perusteella viivakaavio.

	A	B
1	Lämpötila (°C)	sf
2	-0,5	0
3	5,5	11
4	11,5	36
5	17,5	58
6	23,5	78
7	29,5	92

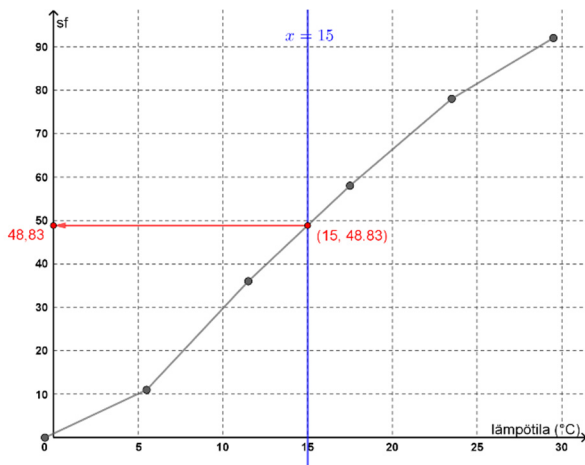


Summakäyrän y -arvo (sf) ilmaisee niiden päivien lukumäärän, joina lämpötila jäi alle tietyn x -arvon.

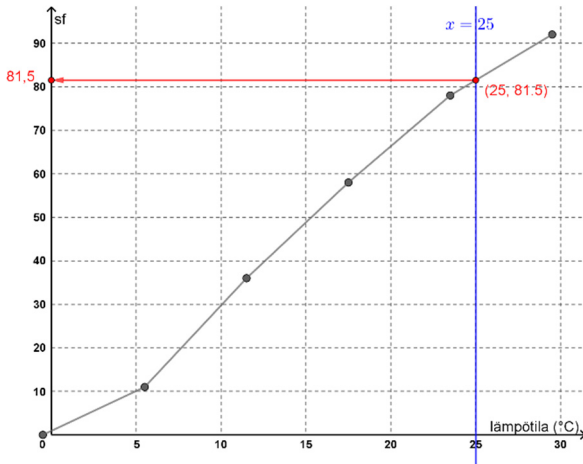
b) Kuvaajan perusteella lämpötilaa $x = 15$ °C vastaava summafrequenssi on 48,83....

Siis päivistä arviolta 48 oli sellaisia, joina lämpötila ei noussut yli 15 asteen.

Huomaa, että lukumäärä on pyöristettävä alaspäin.



c) Kuvaajan perusteella lämpötilaa $x = 25$ °C vastaava summafrekvenssi on 81,5. Siis, päivistä arviolta 81 oli sellaisia, jolloin lämpötila jäi alle 25 asteen.



Viimeisen luokan summafrekvenssistä nähdään, että havaintoja on yhteensä 92, eli lämpötilamittaus tehtiin yhteensä 92 päivänä.

Hellerajan ylittäviä päiviä oli tällöin

$$92 - 81 = 11 \text{ (päivää).}$$

Lämpötila nousi yli hellerajan 11 päivänä.

91.

a) Lasketaan summafrekvenssit taulukkolaskentasovelluksessa.

	A	B	C	D
1	Palkka (€/kk)	f	sf	sf (%)
2	2000-2999	9	9	8,910891
3	3000-3999	11	20	19,80198
4	4000-4999	30	50	49,50495
5	5000-5999	25	75	74,25743
6	6000-6999	15	90	89,10891
7	7000-7999	8	98	97,0297
8	8000-8999	3	101	100

Viimeisen palkkaluokan summafrekvenssistä (solu C8) nähdään, että yrityksessä on 101 työntekijää.

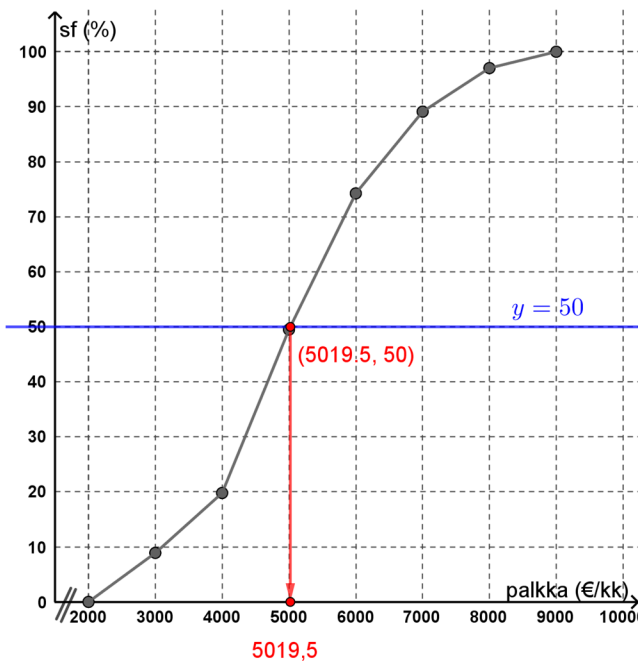
b) Summakäyrä piirretään taulukkolaskentasovelluksessa frekvenssipisteiden avulla. Frekvenssipisteet muodostuvat palkkaluokkien todellisista ylärajoista ja luokkia vastaavista summafrekvensseistä (sf tai $sf\%$).

Summakäyrä alkaa nollatasolta ensimmäisen luokan todellisen alarajan 1999,50 (€/kk) kohdalta. Piirretään summakäyrä suhteellisten summafrekvenssien mukaan ($sf\%$).

Muodostetaan frekvenssipisteistä taulukko ja piirretään sen perusteella viivakaavio.

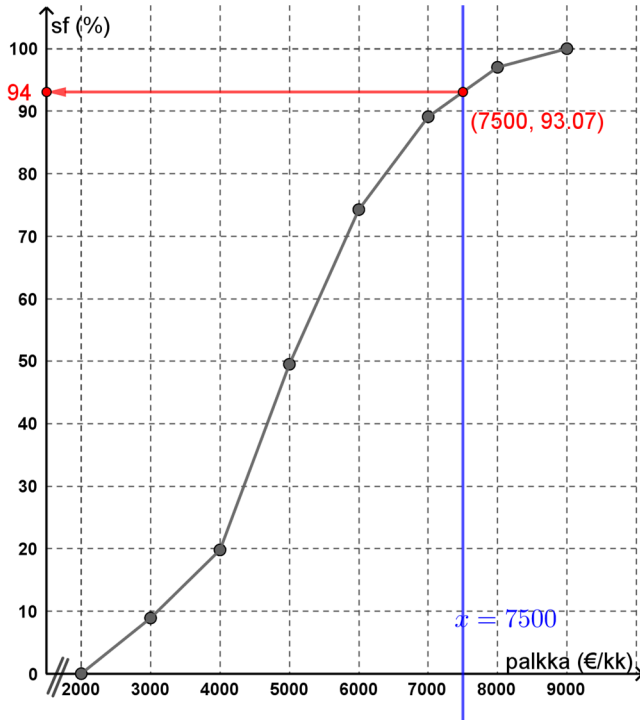
	A	B
1	Palkka (€/kk)	sf (%)
2	1999,50	0
3	2999,50	8,910891
4	3999,50	19,80198
5	4999,50	49,50495
6	5999,50	74,25743
7	6999,50	89,10891
8	7999,50	97,0297
9	8999,50	100

Todelliset luokkarajat ovat puolen euron eli 50 sentin tarkkuudella.



Summakäyrästä arvioidaan mediaaniksi $M_d \approx 5000$ €/kk. Puolet yrityksen työntekijöistä ansaitsee korkeintaan 5000 €/kk ja puolet vähintään 5000 €/kk.

c) Summafrekvenssin arvo ($sf\%$) ilmaisee niiden työntekijöiden prosenttiosuuden, jotka ansaitsevat kuukaudessa korkeintaan tietyn summan. Luetaan kuvaajasta arvoa $x = 7500 \text{ €/kk}$ vastaava summafrekvenssi.



Kuukausiansiota $x = 7500 \text{ €/kk}$ vastaava summafrekvenssi on 93,07 (%) eli arviolta 93 % työntekijöistä ansaitsee alle 7500 €/kk.

Vähintään 7500 €/kk ansaitsevia työntekijöitä on

$$100 \% - 93 \% = 7 \%.$$

92.

Kirjoitetaan asuntojen pinta-alat esimerkiksi matematiikkaohjelmiston taulukkosovellukseen ja tehdään yhden muuttujan analyysi. Tunnusluvut nähdään yhteenvetotaulukosta.

Viiden luvun yhteenveto on:

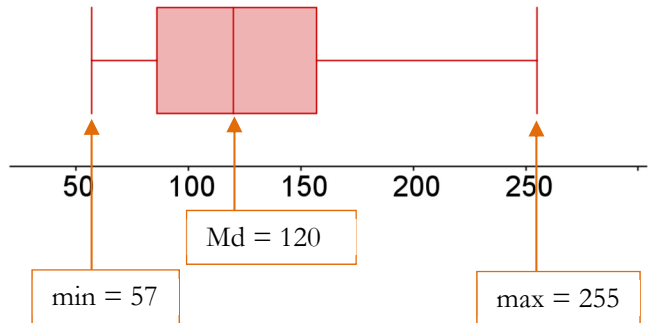
$$\mathbf{min} = 57 \text{ m}^2$$

$$Q_1 = 86 \text{ m}^2$$

$$Md = 120 \text{ m}^2$$

$$Q_3 = 157 \text{ m}^2$$

$$\mathbf{max} = 255 \text{ m}^2$$



a) Pinta-alan yläneljännes on $Q_3 = 157 \text{ m}^2$ eli asunnoista 75 % on alle 157 m^2 . Tällöin asunnoista 25 %, eli suurin neljännes, on vähintään 157 m^2 .

b) Pinta-alan mediaani on $Md = 120 \text{ m}^2$. Asunnoista puolet on korkeintaan 120 m^2 ja puolet vähintään 120 m^2 .

c) Pinta-alan alaneljännes on $Q_1 = 86 \text{ m}^2$ eli asunnoista 25 % on alle 86 m^2 .

93.

Kirjoitetaan tai kopioidaan maiden pinta-alat esimerkiksi matematiikkaohjelmiston taulukkosovellukseen ja tehdään yhden muuttujan analyysi. Tunnusluvut nähdään yhteenvetotaulukosta.

$$N = 27$$

$$\mathbf{min} = 0,3 \text{ (/1000 km}^2\text{)}$$

$$Q_1 = 43 \text{ (/1000 km}^2\text{)}$$

$$Md = 84 \text{ (/1000 km}^2\text{)}$$

$$Q_3 = 302 \text{ (/1000 km}^2\text{)}$$

$$\mathbf{max} = 633 \text{ (/1000 km}^2\text{)}$$

a) Aineiston koko on $N = 27$ eli maita on yhteensä 27 kappaletta.

Pienin arvo on $0,3 \cdot 1000 = 300 \text{ (km}^2\text{)}$.

Suurin arvo on $633 \cdot 1000 = 633\,000 \text{ (km}^2\text{)}$.

Vaihteluväli on siis $300 \text{ km}^2 - 633\,000 \text{ km}^2$.

b) Pinta-alan

- mediaani on $Md = 84 \cdot 1000 = 84\,000 \text{ (km}^2\text{)}$
- alakvartiili on $Q_1 = 43 \cdot 1000 = 43\,000 \text{ (km}^2\text{)}$
- yläkvartiili on $Q_3 = 302 \cdot 1000 = 302\,000 \text{ (km}^2\text{)}$.

Mediaani on $84\,000 \text{ km}^2$. EU-jäsenmaista puolet on pinta-alaltaan korkeintaan tämän suuruisia ja puolet vähintään tämän suuruisia.

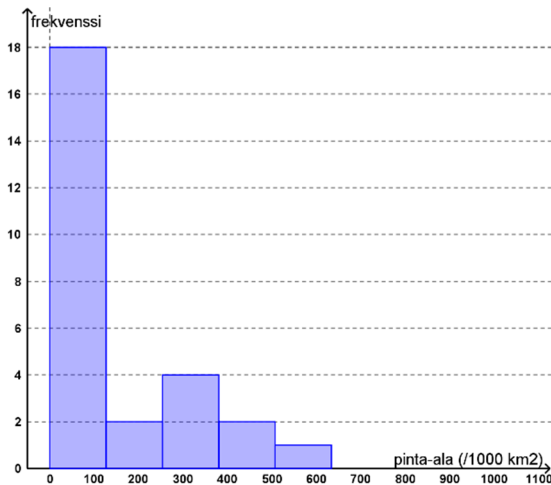
Kvartiiliväli on $[43\,000 \text{ km}^2, 302\,000 \text{ km}^2]$.

EU-jäsenmaista pinta-alaltaan keskimäinen 50 % (puolet) on tällä välillä.

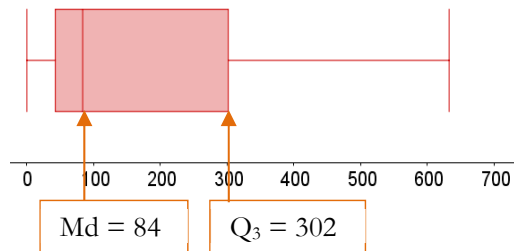
c) Jakaumaa voidaan kuvata histogrammilla tai laatikkokuvioilla. Histogrammi kuvaa frekvenssijakaumaa ja laatikkokuva viiden luvun yhteenvetoa.

Histogrammi riippuu aineiston luokittelusta. Histogrammissa luokkaväliksi voidaan valita esimerkiksi $100 \text{ (km}^2\text{)}$ ja ensimmäisen luokan todelliseksi alarajaksi aineiston pienin arvo 0,3.

Histogrammi



Laatikkokuvaio



Molemmista kuvaajista nähdään, että jakauma on vahvasti painottunut vasemmalle, eli pienet pinta-alat ovat yleisempiä kuin suuret pinta-alat. Sanotaan, että jakauma on *oikealle vino*: jakaumassa on pitkä matala "häntä" tai pitkä "viiksi" oikealle.

Histogrammista nähdään esimerkiksi, että 18 EU-maata eli

$$\frac{18}{27} = 0,666 \dots \approx 67 \%$$

maista on pinta-alaltaan korkeintaan 100 000 km².

Laatikkokuviosta nähdään esimerkiksi, että maista puolet (50 %) on pinta-alaltaan alle 84 000 km² ja yli 302 000 km²:n maita on 25 %.

94.

Kirjoitetaan tai kopioidaan syyskuun sademäärät esimerkiksi matematiikkaohjelmiston taulukkosovellukseen ja tehdään yhden muuttujan analyysi. Tunnusluvut nähdään yhteenvetotaulukosta.

$$N = 30$$

$$\mathbf{min} = 0 \text{ mm/vrk}$$

$$Q_1 = 1,2 \text{ mm/vrk}$$

$$Md = 3,3 \text{ mm/vrk}$$

$$Q_3 = 7,6 \text{ mm/vrk}$$

$$\mathbf{max} = 17,4 \text{ mm/vrk}$$

Havaintoja on yhteensä 30.

a) Sademäärän mediaani on $Md = 3,3 \text{ mm/vrk}$.

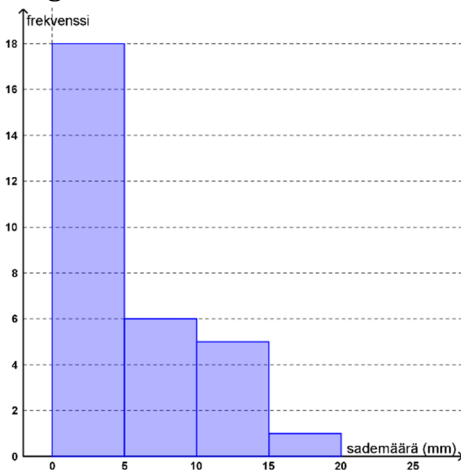
Kyseisen syyskuun päivistä puolet, eli $\frac{30}{2} = 15$ päivää olivat sellaisia, joina satoi korkeintaan $3,3 \text{ mm/vrk}$ ja puolet sellaisia, joina satoi vähintään tämän verran.

b) Sademäärän yläneljännes on $Q_3 = 7,6 \text{ mm/vrk}$ eli 75 % päivistä oli sellaisia, jolloin satoi alle $7,6 \text{ mm/vrk}$. Syyskuun sateisimmalla neljänneksellä (25 % päivistä) satoi vähintään $7,6 \text{ mm/vrk}$.

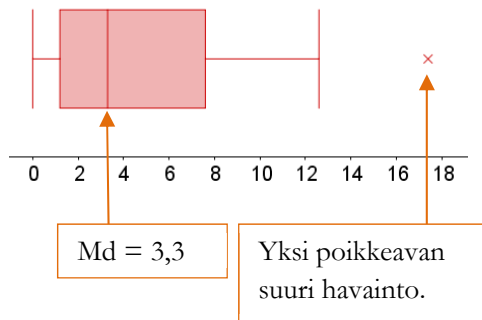
c) Jakaumaa voidaan kuvata histogrammilla tai laatikkokuvioilla. Histogrammi kuvaa frekvenssijakaumaa ja laatikkokuva viiden luvun yhteenvetoa.

Histogrammi riippuu valitusta luokittelusta. Luokkaväliksi voidaan valita esimerkiksi 5 (mm) ja ensimmäisen luokan todelliseksi alarajaksi aineiston pienin arvo 0.

Histogrammi



Laatikkokuva



Molemmista kuvaajista nähdään, että jakauma on vahvasti painottunut vasemmalle, eli pienet sademäärät ovat yleisempiä kuin isot sademäärät. Sanotaan, että jakauma on *oikealle vino*.

Histogrammista nähdään esimerkiksi, että 18 päivää eli

$$\frac{18}{30} = 0,6 = 60 \%$$

päivistä oli sellaisia, joina satoi alle 5 mm/vrk.

Laatikkokuvioista nähdään, että joukossa on yksi poikkeavan suuri havainto: tämä on aineiston suurin arvo 17,4 mm/vrk. Yhden vuorokauden aikana satoi siis poikkeuksellisen paljon. Muina päivinä sademäärä jäi välille 0 mm/vrk – 12,6 mm/vrk.

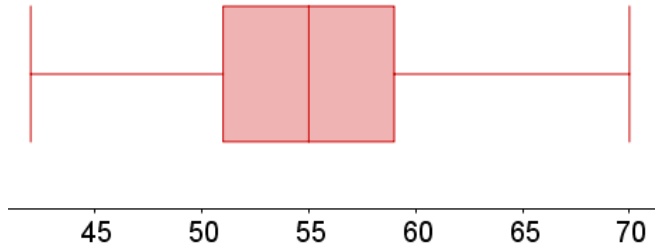
Laatikkokuvioista nähdään lisäksi esimerkiksi, että yläneljänneksen $Q_3 = 7,6$ mm/vrk ja suurimman arvon 17,6 mm/vrk (ja myös toiseksi suurimman arvon 12,6 mm/vrk) välinen neljännes on levein. Suuret sademäärät jakautuvat siis leveämmälle välille kuin pienemmät sademäärät. Tämä näkyy histogrammissa matalana ”häntänä” oikealle.

95.

a) Kirjoitetaan tai kopioidaan iät esimerkiksi matematiikkaohjelmiston taulukkosovellukseen ja tehdään yhden muuttujan analyysi.

Tunnusluvut nähdään yhteenvetotaulukosta. Havainnollistetaan yhteenvetoa laatikkokuvalla.

$N = 45$
min = 42
 $Q_1 = 51$
 $Md = 55$
 $Q_3 = 59$
max = 70



b) Virkaanastumisiän alaneljännes on $Q_1 = 51$ ja ylaneljännes on $Q_3 = 59$. Presidenteistä siis iältään keskimmäinen 50 % on virkaan astuessaan ollut 51–59-vuotias.

c) Virkaanastumisiän alaneljännes on $Q_1 = 51$ eli 25 % (nuorin neljännes) presidenteistä on virkaan astuessaan ollut alle 51-vuotias. Barak Obama oli virkaan asuessaan 47-vuotias, joten hän kuuluu presidenteistä nuorimpaan neljännekseen.

96.

Kirjoitetaan havainnot matematiikkaohjelmiston taulukkosovellukseen ja tehdään yhden muuttujan analyysi. Muutetaan luokitusta asettamalla luokat käsin.

Ensimmäisen luokan 0,4–1,5 todelliset luokkarajat ovat 0,35 (m/s) ja 1,55 (m/s). Luokkavälin pituus on $1,55 - 0,35 = 1,2$ (m/s)

Asetetaan:

- luokituksen alku: 0,35
- luokkavälin leveys: 1,2

Ohjelmisto tuottaa frekvenssitaulukon todellisten luokkarajojen mukaisesti. Laaditaan frekvenssijakaumat laskentataulukossa.

	A	B	C	D	E
1	Nopeus (m/s)	f	sf	f %	sf %
2	0,4 - 1,5	4	4	19.05	19.05
3	1,6 - 2,7	10	14	47.62	66.67
4	2,8 - 3,9	1	15	4.76	71.43
5	4,0 - 5,1	0	15	0	71.43
6	5,2 - 6,3	2	17	9.52	80.95
7	6,4 - 7,5	3	20	14.29	95.24
8	7,6 - 8,7	1	21	4.76	100

b) Frekvenssijakauman perusteella tuulen maksiminopeus oli alle 4,0 m/s kolmessa alimmassa luokassa. Näitä päiviä oli yhteensä $4 + 10 + 1 = 15$.

c) Frekvenssijakauman perusteella tuulen maksiminopeus oli vähintään 5,2 m/s kolmessa ylimmässä luokassa. Näitä päiviä oli yhteensä

$$2 + 3 + 1 = 6$$

eli prosentteina

$$\frac{6}{21} = 0,285 \dots \approx 29 \, \%.$$

d) Luokan 1,6–2,7 (m/s) frekvenssi on suurin ($f = 10$), joten tämä on moodiluokka. Lasketaan moodiluokan luokkakeskus.

$$\frac{1,55 + 2,75}{2} = 2,15 \approx 2,2 \, (\text{m/s}).$$

Siis $M_o \approx 2,2 \, \text{m/s}$.

97.

a) Luokan 0–379 frekvenssi on suurin. Moodiluokka on siis luokka 0–379 (g).

Mediaaniluokka (eli luokka, jossa mediaani on) määritetään suhteellisen summafrekvenssijakauman avulla. Laaditaan tarvittavat frekvenssijakaumat taulukkolaskentasovelluksessa.

	A	B	C	D
1	Kalan paino (g)	f	sf	sf %
2	0 - 379	52	52	26
3	380 - 759	10	62	31
4	760 - 1139	43	105	52.5
5	1140 - 1519	20	125	62.5
6	1520 - 1899	44	169	84.5
7	1900 - 2279	31	200	100

Suhteelliset summafrekvenssit on laskettu sarakkeeseen D. Soluun D2 on kirjoitettu laskukaava:
 $=C2/200*100$
Kaava on kopioitu sarakkeessa D alaspäin.

Viimeisen luokan summafrekvenssistä nähdään, että havaintoja on yhteensä 200.

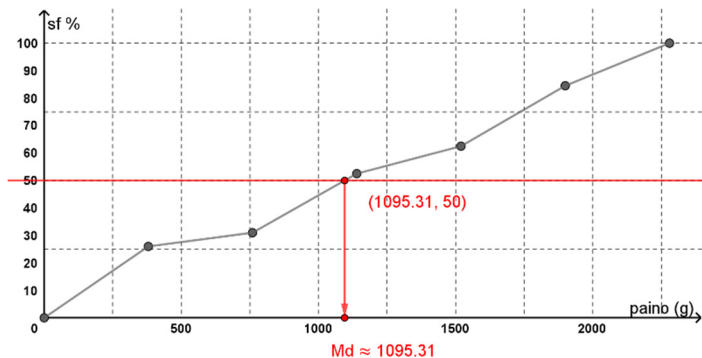
Suhteellinen summafrekvenssi ylittää ensimmäisen kerran arvon 50 % luokka 760–1139 (g), joten tämä on mediaaniluokka.

b) Summakäyrä piirretään taulukkolaskentasovelluksessa frekvenssipisteiden avulla. Frekvenssipisteet muodostuvat painoluokkien todellisista ylärajoista ja luokkia vastaavista summafrekvensseistä (sf tai $sf\%$). Mediaanin arvioimista varten kannattaa piirtää suhteellisten summafrekvenssien ($sf\%$) summakäyrä.

Summakäyrä alkaa nollatasolta ensimmäisen luokan todellisen alarajan 0 (g) kohdalta.

Muodostetaan frekvenssipisteistä taulukko ja piirretään sen perusteella viivakaavio.

	A	B
1	Paino (g)	sf %
2	0	0
3	379.5	26
4	759.5	31
5	1139.5	52.5
6	1519.5	62.5
7	1899.5	84.5
8	2279.5	100



Summakäyrän perusteella mediaani on $Md = 1095,3... \approx 1100$ (g). Kilpailussa saaduista kaloista puolet painoi alle 1100 g ja puolet yli 1100 g.

c) Moodina voidaan ilmoittaa moodiluokan 0–379 (g) luokkakeskus.

$$\frac{0 + 379,5}{2} = 189,75 \approx 200 \text{ (g)}$$

Siis $M_o = 200$ g. Moodiluokan luokkakeskus antaa käsitystä siitä, minkä arvon läheisyyteen suuri osa muuttujan arvoista sijoittuu.

Frekvenssijakauman perusteella suuri osa muuttujan arvoista on myös esimerkiksi välillä 1520–1899 g (luokan frekvenssi on $f = 44$).

Yksittäisen moodiluokan tai moodin ilmaiseminen ei tässä jakaumassa siis kerro oleellista tietoa kalojen painon jakaumasta.

98.

Muodostetaan suhteellinen summafrekvenssijakauma ($sf\%$) taulukkolaskentasovelluksessa.

	A	B	C	D
1	Ikä (vuotta)	f	sf	sf %
2	21 - 25	28	28	6.0475
3	26 - 30	35	63	13.6069
4	31 - 35	51	114	24.622
5	36 - 40	109	223	48.1641
6	41 - 45	80	303	65.4428
7	46 - 50	90	393	84.8812
8	51 - 55	43	436	94.1685
9	56 - 60	17	453	97.8402
10	61 - 65	10	463	100

Suhteelliset summafrekvenssit on laskettu sarakkeeseen D. Soluun D2 on kirjoitettu laskukaava: $=C2/463*100$
Kaava on kopioitu sarakkeessa D alaspäin.

Viimeisen luokan summafrekvenssistä nähdään, että havaintoja on yhteensä 463.

Pyydetty kuvaajat voidaan piirtää esimerkiksi matematiikkaohjelmiston taulukkosovelluksella.

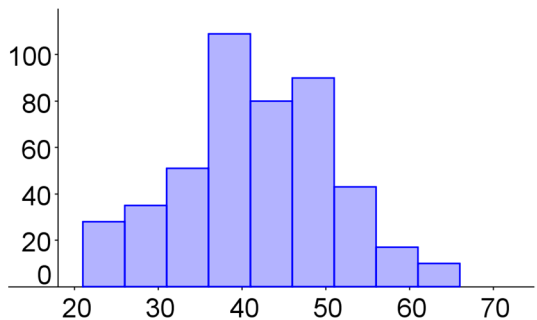
a) Ensimmäisen luokan 21–25 vuotta todellinen alaraja on **21** vuotta. Luokitus on tasavälinen ja luokkavälin leveys on $26 - 21 = 5$ vuotta.

Histogrammi voidaan piirtää frekvenssien (f), suhteellisten frekvenssien ($f\%$) tai summafrekvenssien (sf tai $sf\%$) perusteella.

Havainnollistetaan frekvenssejä (f , sarake B) histogrammilla. Tehdään yhden muuttujan analyysi ja valitaan asetuksista vaihtoehto ”Luokat ja frekvenssi”.

Asetetaan:

- luokituksen alku: **21**
- luokan leveys: **5**



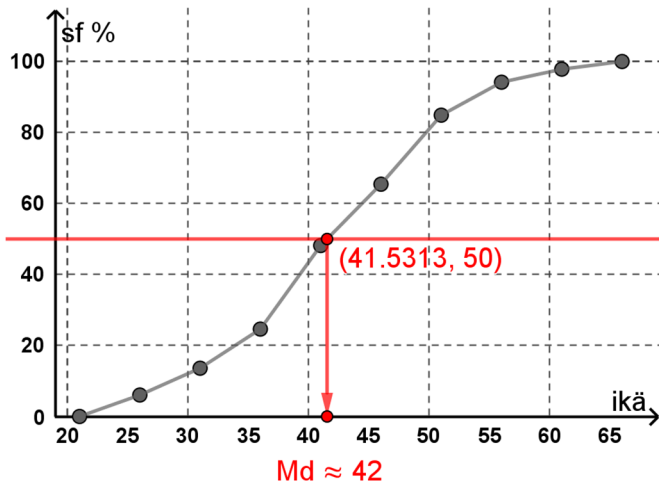
b) Summakäyrä piirretään frekvenssipisteiden avulla.

Frekvenssipisteet muodostuvat ikäluokkien todellisista ylärajoista ja luokkia vastaavista summafrekvensseistä ($sf\%$).

Summakäyrä alkaa nollatasolta ensimmäisen luokan todellisen alarajan 21 (vuotta) kohdalta. Huomaa, että luokkien todelliset ylärajat ovat kokonaislukuja: esimerkiksi ensimmäiseen luokkaan kuuluvat alle 26-vuotiaat, eli todellinen yläraja on 26 (vuotta).

Muodostetaan frekvenssipisteistä taulukko ja piirretään sen perusteella viivakaavio.

	A	B
1	ikä (vuotta)	sf %
2	21	0
3	26	6.0475
4	31	13.6069
5	36	24.622
6	41	48.1641
7	46	65.4428
8	51	84.8812
9	56	94.1685
10	61	97.8402
11	66	100



Summakäyrän perusteella mediaani-ikä on $Md = 41,53... \approx 42$ (vuotta).

99.

Kirjoitetaan tai kopioidaan havainnot matematiikkaohjelmiston taulukkosovellukseen ja tehdään yhden muuttujan analyysi.

a) Tilastojen yhteenvetotaulukosta nähdään mm.:

- aineiston koko on $N = 27$
- pienin arvo on $\min = 1,0$ (%)
- suurin arvo on $\max = 15,6$ (%).

Vaihteluvälin pituus on $15,6 - 1,0 = 14,6$ (prosenttiyksikköä).

Sopiva luokkien lukumäärä on esimerkiksi viisi.

$$\frac{14,6}{5} = 2,92 \approx 3$$

Valitaan luokkavälin pituudeksi 3 (prosenttiyksikköä).

Valitaan ensimmäisen luokan alarajaksi esimerkiksi pienin arvo 1,0 (%).

Asettamalla luokat käsin saadaan taulukon mukainen luokitus.

Työttömyysaste (%)	Frekvenssi
1,0–3,9	19
4,0–6,9	6
7,0–9,9	1
10,0–12,9	0
13,0–15,9	1

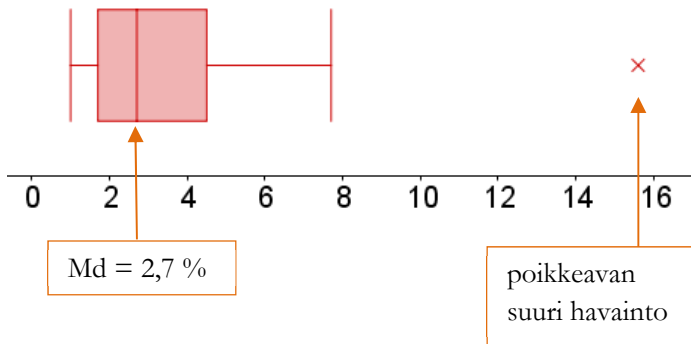
Ensimmäisen luokan todellinen alaraja on 0,95 %. Ohjelmisto tekee automaattisesti luokittelun ja laskee luokkien frekvenssit, kun sovellukseen syötetään:

- luokituksen alku 0,95
- luokan leveys 3

Huomautus: Jos luokkarajoiksi valitaan **kokonaislukuja**, ensimmäisen luokan todelliseksi alarajaksi tulee **0,5**. Tällöin luokitus on:

Työttömyysaste (%)	Frekvenssi
1–3	18
4–6	6
7–9	2
10–12	0
13–15	0
16–18	1

b) Laatikkokuva:



c) Frekvenssitaulukosta näkyy sama kuin laatikkokuvasta: joukossa on yksi poikkeuksellisen suuri havainto (aineiston suurin arvo 15,6 %).

Frekvenssitaulukon mukaan, suurimmassa osassa EU-maita työttömyysaste on korkeintaan 3,9 %. Laatikkokuvion mukaan puolessa EU-maista työttömyysaste on alle 2,7 %.

2.2 Keskiarvo ja keskihajonta

100.

Kirjoitetaan iät laskimen tai matematiikkaohjelmiston taulukkosovellukseen ja määritetään tunnusluvut tilastotoiminnoilla.

a) Virkaanastumisiän keskiarvo on $\bar{x} = 59,833 \dots \approx 60$ (vuotta).

Tieto virkaanastumisiästä on kaikista Suomen presidenteistä, joten kyseessä on kokonaisaineisto. Keskihajonta on $\sigma = 9,796 \dots \approx 9,8$ (vuotta).

b) Virkaanastumisen mediaani-ikä on $M_d = 57$ vuotta.

Suomen presidenteistä puolet ovat virkaan astuessaan olleet alle 57-vuotiaita ja puolet yli 57-vuotiaita.

Huomautus: Mediaani määritettiin yllä laskimen tilastotoiminnoilla. Mediaani on suuruusjärjestykseen asetetun aineiston keskimmäinen arvo (tai kahden keskimmäisen arvon keskiarvo). Iät ovat suuruusjärjestyksessä

41 51 54 55 56 56 58 63 63 69 75 77

Kaksi keskimmäistä ikää ovat 56 ja 58 (vuotta). Mediaani-ikä on näiden keskiarvo eli

$$M_d = \frac{56 + 58}{2} = 57 \text{ (vuotta).}$$

101.

a) Suurin frekvenssi on arvoilla 20 (%) ja 21 (%).

Tyypillisin alv-kanta, eli tyyppiarvo, on $M_o = 20\%$ ja $M_o = 21\%$.

b) Kirjoitetaan verokannat ja frekvenssit laskimen tai matematiikkaohjelmiston taulukkosovellukseen ja määritetään tunnusluvut tilastotoiminnoilla.

Arvonlisäverokannan keskiarvo on $\bar{x} = 21,22 \dots \approx 21\%$.

Tieto verokannoista on kaikista EU-maista, joten kyseessä on kokonaisaineisto.

Keskihajonta on $\sigma = 2,078 \dots \approx 2,1\%$.

c) Arvonlisäverokannan mediaani on $M_d = 21\%$.

Puolet EU-maista ovat sellaisia, joissa alv-kanta on alle 21 % ja puolet maista sellaisia, joissa verokanta on yli 21 %.

Huomautus: Mediaani määritettiin yllä laskimen tilastotoiminnoilla. Mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi ($sf\%$) ensimmäisen kerran ylittää arvon 50 %. Laaditaan tarvittavat frekvenssijakaumat. Niistä nähdään, että verokannan mediaani on $M_d = 21\%$.

	A	B	C	D
1	ALV (%)	f	sf	sf %
2	17	1	1	3.7
3	18	1	2	7.41
4	19	3	5	18.52
5	20	6	11	40.74
6	21	6	17	62.96
7	22	2	19	70.37
8	23	4	23	85.19
9	24	1	24	88.89
10	25	3	27	100

102.

a) Tehtävässä on annettu luokiteltu aineisto. Tunnusluvut arvioidaan luokkakeskusten ja niitä vastaavien frekvenssien perusteella.

Määritetään luokkien luokkakeskukset laskimen tai matematiikkaohjelmiston taulukkosovelluksessa. Määritetään tunnusluvut tilastotoiminnoilla.

Luokkakeskukset
määritetään todellisten
luokkarajojen avulla,
niiden keskiarvona.

	A	B	C	D	E
1	Lämpötila	tod. alaraja	tod. yläraja	luokkakeskus	frekvenssi
2	0-5	0	5.5	2.75	11
3	6-11	5.5	11.5	8.5	25
4	12-17	11.5	17.5	14.5	22
5	18-23	17.5	23.5	20.5	20
6	24-29	23.5	29.5	26.5	14

Tunnusluvut määritetään
luokkakeskusten ja
luokkia vastaavien
frekvenssien perusteella.

Lämpötilan keskiarvoksi saadaan $\bar{x} = 14,595 \dots \approx 15 \text{ (}^\circ\text{C)}.$

Tieto lämpötiloista on kaikilta kyseisen kolmen kuukauden päiviltä, joten kyseessä on kokonaisaineisto. Keskihajonnaksi saadaan $\sigma = 7,484 \dots \approx 7 \text{ (}^\circ\text{C)}.$

Huomaa, että vastaukset pyydettiin asteen tarkkuudella.

b) Kun aineisto luokitellaan, tieto yksittäisten päivien lämpötiloista menetetään*. Tällöin luokitellun aineiston keskiarvoa ja muita tunnuslukuja laskettaessa kunkin luokan luokkakeskus edustaa luokan jokaista yksittäistä havaintoa. Tämän vuoksi luokitellusta aineistosta lasketut tunnusluvut ovat vain arvioita.

**Huomautus:* Jatkuvan muuttujan aineisto luokitellaan esimerkiksi tutkimusraportissa tiedon esittämistä varten. Tieto yksittäisistä havainnoista on edelleen raportin kirjoittajan tiedossa. Tunnusluvut määritetään luokittelemattomasta aineistosta aina, kun sellainen on käytettävissä, vaikka aineisto olisi esittämistä varten luokiteltu.

103.

Frekvenssit ovat suhteellisia eli ne ilmaisevat prosenttiosuuksia. Tunnuslukujen laskeminen tapahtuu samaan tapaa kuin absoluuttisten frekvenssien tilanteessa.

Määritetään luokkien luokkakeskukset laskimen tai matematiikkaohjelmiston taulukkosovelluksessa. Määritetään tunnusluvut tilastotoiminnoilla.

Luokkakeskukset määritetään todellisten luokkarajojen avulla, niiden keskiarvona.

	A	B	C	D	E
1	Koulumatka (km)	tod.alaraja	tod.yläraja	luokkakeskus	f %
2	1-4	0.5	4.5	2.5	13
3	5-8	4.5	8.5	6.5	42
4	9-12	8.5	12.5	10.5	8
5	13-16	12.5	16.5	14.5	13
6	17-20	16.5	20.5	18.5	17
7	21-24	20.5	24.5	22.5	7

Tunnusluvut määritetään luokkakeskusten ja luokkia vastaavien frekvenssien perusteella.

Koulumatkan keskiarvoksi saadaan $\bar{x} = 10,5$ (km).

Kyseessä on kokonaisaineisto eli tieto koulumatkoista on kaikilta kyseisen kurssin opiskelijoilta. Keskihajonnaksi saadaan $\sigma = 6,170 \dots \approx 6,2$ (km).

104.

a) Arvioidaan sademäärän keskiarvo (mm/vrk). Luokitellun aineiston tunnusluvut lasketaan luokkakeskusten ja niitä vastaavien frekvenssien perusteella.

Määritetään luokkien luokkakeskukset laskimen tai matematiikkaohjelmiston taulukkosovelluksessa. Määritetään tunnusluvut tilastotoiminnoilla.

Luokkakeskukset määritetään todellisten luokkarajojen avulla, niiden keskiarvona.

	A	B	C	D	E
1	Sademäärä (mm/vrk)	tod.alaraja	tod.yläraja	luokkakeskus	f
2	0.0 - 2.0	0	2.05	1.025	6
3	2.1 - 4.0	2.05	4.05	3.05	3
4	4.1 - 6.0	4.05	6.05	5.05	6
5	6.1 - 8.0	6.05	8.05	7.05	4
6	8.1 - 10.0	8.05	10.05	9.05	5
7	10.1 - 12.0	10.05	12.05	11.05	3
8	12.1 - 14.0	12.05	14.05	13.05	1
9	14.1 - 16.0	14.05	16.05	15.05	2

Tunnusluvut määritetään luokkakeskusten ja luokkia vastaavien frekvenssien perusteella.

Sademäärän keskiarvoksi saadaan $\bar{x} = 6,511 \dots \approx 6,5$ (mm/vrk).

Kesäkuussa satoi keskimäärin 6,5 mm/vrk.

b) Kunkin luokan luokkakeskus ilmaisee sademäärää ja luokan frekvenssi päivien lukumäärää. Esimerkiksi ensimmäisen luokan luokkakeskus on 1,025 mm/vrk ja luokan frekvenssi on 6, joten kesäkuussa oli kuusi päivää, jolloin satoi noin 1,025 mm/vrk.

Kokonaissademäärä saadaan kertomalla kunkin luokan luokkakeskus luokan frekvenssillä ja laskemalla tulot yhteen. Tehdään lasku laskentataulukossa.

D	E	F
luokkakeskus	f	tulot
2. rivi	6	6.15
3.05	3	9.15
5.05	6	30.3
7.05	4	28.2
9.05	5	45.25
11.05	3	33.15
13.05	1	13.05
15.05	2	30.1
		195.35

Soluun F2 on kirjoitettu laskukaava: $=D2 \cdot E2$
Kaava on kopioitu sarakkeessa F alaspäin.

Soluun on laskettu sarakkeessa yläpuolella olevien solujen summa.

Summasta nähdään, että kesäkuussa satoi kaiken kaikkiaan noin $195,35 \approx 195$ mm.

Tapa 2: Kesäkuun sademäärä voidaan arvioida myös a-kohdan perusteella. Kesäkuussa satoi keskimäärin $\bar{x} = 6,511 \dots$ mm/vrk. Kesäkuussa on yhteensä 30 päivää, joten kesäkuussa satoi kaiken kaikkiaan noin

$$30 \cdot 6,511 \dots = 195,349 \dots \approx 195 \text{ (mm)}.$$

105.

a) Kirjoitetaan tai kopioidaan sähkönkulutukset laskimen tai matematiikkaohjelmiston taulukkosovellukseen ja määritetään tunnusluvut tilastotoiminnoilla. Aineistot ovat otoksia, joten käytetään otoskeskihajontaa (merkintä on: s tai s_{n-1} tai σ_{n-1}).

Tunnusluvut voidaan määrittää myös taulukkolaskenta-ohjelmistossa (esim. ladattava tiedosto Excel) yksitellen, sopivilla komennoilla.

Kysely 1:

Keskiarvo on $\bar{x} = 1986,73 \dots \approx 1987$ (kWh/vuosi).

Otoskeskihajonta on $s = 421,96 \dots \approx 422$ (kWh/vuosi).

Kysely 2:

Keskiarvo on $\bar{x} = 3815,42 \dots \approx 3815$ (kWh/vuosi).

Otoskeskihajonta on $s = 426,47 \dots \approx 426$ (kWh/vuosi).

Kysely 3:

Keskiarvo on $\bar{x} = 7137,15 \dots \approx 7137$ (kWh/vuosi).

Otoskeskihajonta on $s = 3936,73 \dots \approx 3937$ (kWh/vuosi).

b) Kyselyihin 1 ja 2 ovat vastanneet rajatut, keskenään samanlaiset kotitaloudet (kerrostalo- ja rivitaloasukkaat). Esimerkiksi kerrostaloissa on keskimäärin pienempi sähkönkulutus kuin rivi- ja omakotitaloissa. Sähkönkulutuksen hajonta on keskenään samanlaisissa kotitalouksissa keskimäärin pienempää kuin kaikkien kotitalouksien joukossa.

Kysely 3 edustaa kaikkia kyseisen asuntoalueen kotitalouksia ja antaa siis parhaan arvion asuntoalueen kaikkien kotitalouksien keskimääräisestä sähkönkulutuksesta.

Asuntoalueen keskimääräinen sähkönkulutus oli siis 7137,15 ... $\approx$ 7137 (kWh/vuosi) ja kulutuksen keskihajonta oli 3936,73 ... $\approx$ 3937 (kWh/vuosi).

106.

a) Kirjoitetaan lämpötilat laskimen tai matematiikkaohjelmiston taulukkosovellukseen ja määritetään tunnusluvut tilastotoiminnoilla. Aineistot ovat kokonaisaineistoja (tieto on juhannusviikon jokaiselta päivältä), joten käytetään keskihajontaa (merkintä on: σ tai s_n tai σ_n).

Kaupunki A:

Keskiarvo on $\bar{x} = 21,571 \dots \approx 21,6$ (°C)

Keskihajonta on $\sigma = 1,399 \dots \approx 1,4$ (°C)

Kaupunki B:

Keskiarvo on $\bar{x} = 21,142 \dots \approx 21,1$ (°C)

Keskihajonta on $\sigma = 3,440 \dots \approx 3,4$ (°C)

Kaupunki C:

Keskiarvo on $\bar{x} = 16,571 \approx 16,6$ (°C)

Keskihajonta on $\sigma = 0,903 \dots \approx 0,9$ (°C)

b) Keskimäärin lämpimintä oli kaupungissa A, missä päivän ylimmän lämpötilan viikkokeskiarvo oli $21,6\text{ }^{\circ}\text{C}$. Keskimäärin kylmintä oli kaupungissa C, missä päivän ylimmän lämpötilan viikkokeskiarvo oli $16,6\text{ }^{\circ}\text{C}$.

Kaupunkien A ja B keskilämpötilat $21,6\text{ }^{\circ}\text{C}$ ja $21,1\text{ }^{\circ}\text{C}$ ovat lähellä toisiaan, mutta lämpötilan keskihajonta on kaupungissa B suurempaa: $3,4\text{ }^{\circ}\text{C} > 1,4\text{ }^{\circ}\text{C}$. Kaupungissa B päivän ylimmissä lämpötiloissa on keskimäärin suuremmat erot kuin kaupungissa A.

Kaupungissa C lämpötilan keskihajonta $0,9\text{ }^{\circ}\text{C}$ on pienintä. Kaupungissa C päivien ylimmissä lämpötiloissa oli siis viikon aikana keskimäärin pienin vaihtelu.

107.

a) Kirjoitetaan tai kopioidaan pituudet laskimen tai matematiikkaohjelmiston taulukkosovellukseen ja määritetään tunnusluvut tilastotoiminnoilla. Kyseessä on kokonaisaineisto, joten käytetään perusjoukon keskihajontaa (merkintä on: σ tai s_n tai σ_n).

Keskiarvo on $\bar{x} = 172,95 \dots \approx 173,0$ (cm).

Keskihajonta on $\sigma = 8,724 \dots \approx 8,7$ (cm).

b) Luokitellun aineiston tunnusluvut lasketaan luokkakeskusten ja luokkia vastaavien frekvenssien perusteella.

Kirjoitetaan luokkakeskukset ja frekvenssit laskin- tai matematiikkaohjelmiston taulukkosovellukseen ja määritetään tunnusluvut tilastotoiminnoilla.

Luokitellun aineiston

keskiarvo on $\bar{x} = 173,83 \dots \approx 173,8$ (cm)

keskihajonta on $\sigma = 9,51 \dots \approx 9,5$ (cm).

c) Kun aineisto luokitellaan, tieto yksittäisistä pituuksista menetetään. Luokitellussa aineistossa kunkin luokan luokkakeskus edustaa kaikkia kyseisen luokan havaintoja. Siis esimerkiksi ensimmäisen luokan 158–165 (cm) luokkakeskus 161,5 (cm) edustaa kaikkia kyseiseen luokkaan kuuluvia kuutta havaintoa (havainnot ovat: 158 cm, 159 cm, 163 cm, 164 cm, 165 cm ja 165 cm).

Tunnuslukua laskettaessa yksittäisten havaintojen arvot korvataan kyseisen luokan luokkakeskuksella. Tästä syystä a- ja b-kohdissa lasketut tunnuslukujen arvot poikkeavat toisistaan.

Tunnusluvut lasketaan aina luokittelemattomasta aineistossa, jossa on tieto kaikista havainnoista. a-kohdassa laskettiin tunnuslukujen tarkat arvot.

108.

a) Luetaan kuvaajasta pylväiden reunojen kohdalta luokkien todelliset ala- ja ylärajat. Luokkakeskukset voidaan lukea frekvenssimonikulmion frekvenssipisteistä tai laskea todellisten luokkarajojen keskiarvoina. Luokkia vastaavat frekvenssit luetaan pystyakselilta pylvään korkeuden kohdalta.

tod.alaraja (s)	tod.yläraja (s)	luokkakeskus (s)	f
4.5	6.5	5.5	6
6.5	8.5	7.5	5
8.5	10.5	9.5	3
10.5	12.5	11.5	4
12.5	14.5	13.5	1
14.5	16.5	15.5	1

b) Lasketaan frekvenssien summa.

$$6 + 5 + 3 + 4 + 1 + 1 = 20$$

Havaintojen lukumäärä on $N = 20$ eli 20 opiskelijaa suoritti testin.

c) Tunnusluvut määritetään luokkakeskusten ja frekvenssien perusteella tilastotoiminnoilla.

Testin tulosten keskiarvoksi saadaan $\bar{x} = 8,7$ s ja mediaaniksi $M_d = 7,5$ s.

Tieto testituloksista on kaikilta ryhmän opiskelijoilta, joten kyseessä on kokonaisaineisto. Tulosten keskihajonnaksi saadaan $\sigma = 2,925 \dots \approx 2,9$ s.

109.

a) Luetaan kuvaajasta pylväiden reunojen kohdalta luokkien todelliset ala- ja ylärajat. Luokkakeskukset voidaan lukea frekvenssimonikulmion frekvenssipisteistä tai laskea todellisten luokkarajojen keskiarvoina. Luokkia vastaavat frekvenssit luetaan pystyakselilta pylvään korkeuden kohdalta.

tod.alaraja (€/vko)	tod.alaraja (€/vko)	luokkakeskus (€/vko)	f
7.5	11.5	9.5	9
11.5	15.5	13.5	6
15.5	19.5	17.5	5
19.5	23.5	21.5	7
23.5	27.5	25.5	3

Frekvenssijakaumataulukossa luokkarajat ilmaistaan euron tarkkuuteen pyöristettyinä kokonaislukuina.

Viikkoraha (€/ vko)	f
8–11	9
12–15	6
16–19	5
20–23	7
24–27	3

Lasketaan frekvenssien summa.

$$9 + 6 + 5 + 7 + 3 = 30$$

Havaintojen lukumäärä on $N = 30$ eli 30 opiskelijaa vastasi kyselyyn.

b) Tunnusluvut määritetään luokkakeskusten ja frekvenssien perusteella tilastotoiminnoilla.

Viikkorahan keskiarvoksi saadaan $\bar{x} = 16,033 \dots \approx 16,0$ (€/vko).

Viikkorahan mediaaniksi saadaan $M_d = 15,50 \approx 15,5$ (€/vko).

Keskimääräinen viikkoraha oli noin 16 €/vko.



Tässä jakaumassa keskiarvo ja mediaani ovat varsin lähellä toisiaan. Molemmat keskiluvut kuvaavat hyvin keskimääräistä muuttujan arvoa.

c) Jakauman moodiluokka on 7,5–11,5 (€/vko). Viikkorahan moodina eli yleisimpänä viikkorahan voidaan pitää moodiluokan luokkakeskusta, joka on 9,5 (€/vko).

Yleisin viikkoraha vastaajien joukossa oli 9,5 €/vko.

110.

a) Luetaan kuvaajasta pylväiden reunojen kohdalta luokkien todelliset ala- ja ylärajat. Luokkakeskukset voidaan lukea frekvenssimonikulmion frekvenssipisteistä tai laskea todellisten luokkarajojen keskiarvoina. Luokkia vastaavat frekvenssit luetaan pystyakselilta pylvään korkeuden kohdalta.

tod.alaraja (m/s)	tod.yläraja (m/s)	luokkakeskus (m/s)	f
0.5	2.5	1.5	13
2.5	4.5	3.5	2
4.5	6.5	5.5	3
6.5	8.5	7.5	3

b) Tunnusluvut määritetään luokkakeskusten ja frekvenssien perusteella tilastotoiminnoilla. Tuulennopeuden keskiarvoksi saadaan $\bar{x} = 3,119 \dots \approx 3,1 \text{ m/s}$.

Kyseessä on kokonaisaineisto, joten keskihajonnaksi saadaan $\sigma = 2,277 \dots \approx 2,3 \text{ m/s}$.

c) Tuulennopeuden mediaaniksi saadaan $M_d = 1,5 \text{ m/s}$.

Kyseisen kolmen viikon aikana puolet päivistä (10 päivää) olivat sellaisia, jolloin tuulen maksiminopeus oli alle $1,5 \text{ m/s}$ ja puolet päivistä (10 päivää) sellaisia, jolloin tuulen maksiminopeus oli yli $1,5 \text{ m/s}$.

d) Tuulen päivittäisen maksiminopeuden keskiarvoksi saatiin 3,1 m/s kun mediaani, 1,5 m/s, on selvästi pienempi.

Frekvenssijakauman kuvaajasta nähdään, että jakauma on voimakkaasti vasemmalle painottunut eli pienet muuttujan arvot ovat yleisimpiä. Jakauman oikealla oleva ”häntä”, eli suuret muuttujan arvot, joita on vähemmän, siirtää keskiarvoa mediaanista oikealle.

Tällaisessa voimakkaasti vinossa jakaumassa keskiarvo on huono tunnusluku kuvaamaan keskimääräistä muuttujan arvoa. Nyt keskiarvoksi saatiin 3,1 m/s, mutta suurin osa havainnoista: 13 havaintoa eli $\frac{13}{21} = 0,619 \dots \approx 62 \%$ sijoittuu ensimmäiseen luokkaan eli välille 0,5–2,5 m/s ja ovat näin keskiarvoa pienempiä.

Mediaani 1,5 m/s kuvaa keskimääräistä muuttujan arvoa paremmin.

Huomautus: Keskiluvun tulisi olla edustava arvo, joka on jossain mielessä lähellä joukon lukuja. Keskiarvo ja mediaani ovat molemmat ”edustavia arvoja”, mutta eri tavoin. Toisinaan tarvitaan huolellista harkintaa kumpi keskiluku, keskiarvo vai mediaani, kuvaa jakauman keskimääräistä arvoa paremmin. Tähän asiaan palataan kappaleessa 2.3.

111.

Luokitellun aineiston tunnusluvut määritetään luokkakeskusten ja frekvenssien perusteella. Kirjoitetaan todelliset luokkarajat sekä luokkia vastaavat frekvenssit laskin- tai matematiikkaohjelmiston taulukkosovellukseen.

Määritetään luokkakeskukset. Määritetään tunnusluvut tilastotoiminnoilla.

Toimipiste A			
tod.alaraja	tod.yläraja	luokkakeskus	f
0	5.5	2.75	10
5.5	11.5	8.5	26
11.5	17.5	14.5	8
17.5	23.5	20.5	35
23.5	29.5	26.5	5

Keskiarvo: $\bar{x}_A =$
14,45 ... $\approx$ 14,5 (min)

Keskihajonta: $\sigma_A =$
7,15 ... $\approx$ 7,2 (min)

Toimipiste B			
tod.alaraja	tod.yläraja	luokkakeskus	f
0	5.5	2.75	12
5.5	11.5	8.5	15
11.5	17.5	14.5	30
17.5	23.5	20.5	20
23.5	29.5	26.5	6

Keskiarvo: $\bar{x}_B =$
14,03 ... $\approx$ 14,0 (min)

Keskihajonta: $\sigma_B =$
6,73 ... $\approx$ 6,7 (min)

a) Keskiarvoista nähdään, että keskimääräinen jonotusaika oli hieman lyhyempi toimipisteessä B (14,0 min < 14,5 min).

b) Keskihajonnoista nähdään, että jonotusaika jakautui tasaisemmin toimipisteessä B (6,7 min < 7,2 min).

c) Lasketaan arvon 29 (min) normitettu arvo toimipisteen A jakaumassa. Käytetään laskussa tunnuslukujen tarkkoja arvoja eli kopioidaan ne laskimen laskukenttään tai tehdään lasku taulukossa soluviittauksilla.

$$Z_{29} = \frac{29 - \bar{x}_A}{\sigma_A} = \frac{29 - 14,45 \dots}{7,15 \dots} = 2,033 \dots$$

Huomaa: Jos käytät matematiikkaohjelmistoa, voit kopioida tunnuslukujen arvot tilastojen yhteenvetotaulukosta laskentataulukon tyhjiin soluihin. Normituslasku voidaan tämän jälkeen tehdä laskentataulukossa soluviittauksilla.

Normitetun arvon itseisarvo on $|Z_{29}| = 2,033 \dots > 2$, joten jonotusaika 29 min poikkeaa merkitsevästi keskiarvosta.

112.

a) Luokitellun aineiston tunnusluvut määritetään luokkakeskusten ja frekvenssien perusteella. Kirjoitetaan todelliset luokkarajat sekä luokkia vastaavat frekvenssit laskin- tai matematiikkaohjelmiston taulukkosovellukseen. Määritetään tunnusluvut tilastotoiminnoilla.

Paino (g)				
tod.alaraja	tod.yläaraja	luokkakeskus	f	
204.5	254.5	229.5	3	Keskiarvo: $\bar{x} =$
254.5	304.5	279.5	6	331,42 ... ≈ 331 (g)
304.5	354.5	329.5	6	Keskihajonta: $\sigma =$
354.5	404.5	379.5	9	57,98 ... ≈ 58 (g)
404.5	454.5	429.5	2	

b) Lasketaan arvon 452 (g) normitettu arvo. Käytetään laskussa tunnuslukujen tarkkoja arvoja eli kopioidaan ne laskimen laskukenttään tai tehdään lasku taulukossa soluviittauksilla.

$$Z_{452} = \frac{452 - \bar{x}}{\sigma} = \frac{452 - 331,42 \dots}{57,98 \dots} = 2,079 \dots$$

Huomaa: Jos käytät matematiikkaohjelmistoa, voit kopioida tunnuslukujen arvot tilastojen yhteenvetotaulukosta laskentataulukkoon, tyhjiin soluihin. Normituslasku voidaan tämän jälkeen tehdä laskentataulukossa soluviittauksilla.

Normitetun arvon itseisarvo on $|Z_{452}| = 2,079 \dots > 2$, joten paino 452 g poikkeaa merkitsevästi keskiarvosta.

c) Tasan yhden keskihajonnan päässä keskiarvosta vasemmalle on kurkku, jonka massa on

$$\bar{x} - \sigma = 331,42 \dots - 57,98 \dots = 273,44 \dots \approx 273 \text{ (g)}.$$

Tasan yhden keskihajonnan päässä keskiarvosta oikealle on kurkku, jonka massa on

$$\bar{x} + \sigma = 331,42 \dots + 57,98 \dots = 389,40 \dots \approx 389 \text{ (g)}.$$

Korkeintaan yhden keskihajonnan verran keskiarvosta ovat massat 273 g – 389 g.

113.

a) Määritetään pinta-alojen keskiarvo ja keskihajonta.

- keskiarvo on $\bar{x} = 138,86 \dots \approx 139 \text{ (m}^2\text{)}$
- keskihajonta on $\sigma = 50,09 \dots \approx 50 \text{ (m}^2\text{)}$

b) Lasketaan arvon 266 (m²) normitettu arvo.

$$Z_{266} = \frac{266 - \bar{x}}{\sigma} = \frac{266 - 138,86 \dots}{50,09 \dots} = 2,537 \dots$$

Normitetun arvon itseisarvo on $|Z_{266}| = 2,53 \dots > 2$, joten pinta-ala 266 m² poikkeaa merkitsevästi keskiarvosta.

c) Tasan yhden keskihajonnan päässä keskiarvosta vasemmalle on näyttelykohde, jonka pinta-ala on

$$\bar{x} - \sigma = 138,86 \dots - 50,09 \dots = 88,77 \dots \approx 89 \text{ (m}^2\text{)}.$$

Tasan yhden keskihajonnan päässä keskiarvosta oikealle on näyttelykohde, jonka pinta-ala on

$$\bar{x} + \sigma = 138,86 \dots + 50,09 \dots = 188,96 \dots \approx 189 \text{ (m}^2\text{)}.$$

Yhden keskihajonnan verran näyttelykohteiden keskikoosta poikkeavat pinta-alat ovat 89 m² ja 189 m².

114.

Luokitellun aineiston tunnusluvut lasketaan luokkakeskusten ja frekvenssien perusteella. Lasketaan luokkakeskukset laskin- tai matematiikkaohjelmiston taulukkosovelluksessa todellisten luokkarajojen keskiarvona. Määritetään tunnusluvut tilastotoiminnoilla.

Aika (min)			
tod.alaraja	tod.yläraja	luokkakeskus	f
59.5	104.5	82	3
104.5	149.5	127	2
149.5	194.5	172	5
194.5	239.5	217	8
239.5	284.5	262	4
284.5	329.5	307	3
329.5	374.5	352	2

Keskiarvo: $\bar{x} =$
213,66 ... $\approx$ 213,7 (min)

Keskihajonta: $\sigma =$
74,42 ... $\approx$ 74,4 (min)

b) Jyri teki koetta 3 h = 3 · 60 min = 180 min.

Lasketaan arvon 180 (min) normitettu arvo.

$$Z_{180} = \frac{180 - \bar{x}}{\sigma} = \frac{180 - 213,66 \dots}{74,42 \dots} = -0,452 \dots$$

Normitetun arvon itseisarvo on $|Z_{180}| = 0,45 \dots < 2$, joten Jyrin kokeeseen käyttämä aika ei poikennut merkittävästi keskiarvosta.

c) Keskihajonta ilmaisee koeaikojen tasaisuutta.

- Lukiossa A kokeeseen käytetyn ajan keskihajonta oli $\sigma \approx 74,4$ min.
- Lukiossa B kokeeseen käytetyn ajan keskihajonta oli $\sigma \approx 75,1$ min.

Keskihajonta oli hieman pienempää lukiossa A ($74,4 < 75,1$), joten lukiossa A kokeeseen käytetyssä ajassa oli vähemmän hajontaa eli kokeeseen käytetty aika jakautui tasaisemmin.

115.

a) Luetaan kuvaajasta pylväiden reunojen kohdalta luokkien todelliset ala- ja ylärajat. Luokkakeskukset voidaan lukea frekvenssimonikulmion frekvenssipisteistä tai laskea todellisten luokkarajojen keskiarvoina. Luokkia vastaavat frekvenssit luetaan pystyakselilta pylvään korkeuden kohdalta.

tod.alaraja (€/m²)	tod.ylärajaja (€/m²)	luokkakeskus (€/m²)	f
5700	6000	5850	7
6000	6300	6150	9
6300	6600	6450	5
6600	6900	6750	3
6900	7200	7050	1

b) Tunnusluvut määritetään luokkakeskusten ja frekvenssien perusteella tilastotoiminnoilla.

Neliöhinnan keskiarvoksi saadaan $\bar{x} = 6234 \text{ €/m}^2$.

Kyseessä on kokonaisaineisto, joten keskihajonnaksi saadaan $\sigma = 334,28 \dots \approx 334 \text{ €/m}^2$.

b) Neliöhinnan mediaaniksi saadaan $M_d = 6150 \text{ €/m}^2$.

c) Neliöhinnan keskiarvoksi saatiin 6234 €/m^2 kun mediaani 6150 €/m^2 on sitä hivenen pienempi. Moodiluokan luokkakeskus on $Mo = 6150 \text{ €/m}^2$ eli yhtä suuri kuin mediaani.

Frekvenssijakauman kuvaajasta nähdään, että jakauma on vasemmalle painottunut eli pienet muuttujan arvot ovat yleisimpiä. Jakauman oikealla oleva ”häntä” eli suuret muuttujan arvot, joita on vähemmän, siirtää keskiarvoa mediaanista oikealle.

Tällaisessa vinossa jakaumassa keskiarvo on yleensä huono tunnusluku kuvaamaan keskimääräistä muuttujan arvoa. Mediaani ja moodi kuvaavat keskimääräistä muuttujan arvoa paremmin.

Asuntojen keskimääräinen neliöhinta on 6150 €/m^2 (mediaanihinta).

Huomautus: Toisinaan tarvitaan huolellista harkintaa kumpi keskiluku, keskiarvo vai mediaani, kuvaa keskimääräistä arvoa asianmukaisemmin. Tässä aineistossa keskiarvo ja mediaani arvioitiin luokitellusta aineistosta eikä ero niiden välillä ole suuri.

Yleensä ”keskimääräisellä arvolla” tarkoitetaan (aritmeettista) keskiarvoa. Jos jakauma on hyvin vino tai siinä on keskiarvosta merkitsevästi poikkeavia arvoja, keskiarvon käyttöä tunnuslukuna tulisi kuitenkin välttää. Tätä asiaa pohditaan tarkemmin kappaleessa 2.3.

116.

a) Kirjoitetaan tai kopioidaan pinta-alat laskin- tai matematiikkaohjelmiston taulukkosovellukseen ja määritetään tunnusluvut tilastotoiminnoilla.

- keskiarvo: $\bar{x} = 50,533 \dots \approx 50,5$ (ha)
- keskihajonta: $\sigma = 23,773 \dots \approx 23,8$ (ha)
- mediaani: $Md = 50,45 \approx 50,5$ (ha)

b) Lasketaan, kuinka monen keskihajonnan σ päässä arvo 60 (ha) on keskiarvosta. Käytetään laskussa tunnuslukujen tarkkoja arvoja, eli tehdään lasku laskentataulukossa soluviittauksilla tai kopioidaan tunnuslukujen arvot laskimen laskukenttään.

$$\frac{60 - \bar{x}}{\sigma} = \frac{60 - 50,533 \dots}{23,773 \dots} = 0,398 \dots \approx 0,4$$

Ilmoitettu arvo on 0,4 keskihajonnan päässä keskiarvosta.

Ohje: laskun tekeminen laskentataulukossa.

Pinta-alat on syötetty sarakkeeseen A.

	A	B
1	33.4	50.5333333333
2	45	23.7737296658
3	67.5	
4	23.6	0.3981986335
5	46.8	
6	35.6	
7	56.7	
8	12.5	

Soluun B1 on määritetty keskiarvo ja soluun B2 keskihajonta.

Soluun B4 on kirjoitettu laskukaava:
 $= (60 - B1) / B2$

117.

a) Mediaani arvioidaan jakauman summakäyrästä (kappale 2.1). Summakäyrän piirtämiseen tarvitaan frekvenssipisteet, jotka määritellään luokkien todellisten luokkarajojen perusteella. Mukana on nyt avoimia luokkia (alin luokka ja ylin luokka, toinen luokkaraja puuttuu), joten kaikkia frekvenssipisteitä ei voida määrittää.

Arvioidaan mediaani frekvenssijakauman perusteella. Laaditaan frekvenssijakauma taulukkosovelluksessa.

	A	B	C	D
1	Ikä (vuotta)	f	sf	sf %
2	- 19	839	839	1.587
3	20 -24	7254	8093	15.307
4	25 - 29	15576	23669	44.768
5	30 - 34	17583	41252	78.025
6	35 - 39	9298	50550	95.612
7	40 -	2320	52870	100

Soluun D2 on kirjoitettu laskukaava:
 $=C2/52870*100$
Kaava on kopioitu sarakkeessa D alaspäin.

Viimeisen luokan summafrekvenssistä (solu C7) nähdään, että havaintoja on yhteensä 52 870.

Mediaaniluokka (eli luokka, johon mediaani kuuluu) on se luokka, jonka suhteellinen summafrekvenssi (sf%) ensimmäisen kerran ylittää arvon 50 %. Frekvenssijakauman perusteella mediaaniluokka on 30–34 vuotta.

Luokitellun aineiston mediaanina voidaan käyttää mediaaniluokan luokkakeskusta, joka on mediaaniluokan todellisten luokkarajojen 30 ja 35 avulla laskettuna

$$\frac{30 + 35}{2} = 32,5 \text{ (vuotta).}$$

Synnyttäjien mediaani-ikä on 32,5 vuotta.

b) Luokan 30–34 vuotta frekvenssi on suurin (17 583), joten moodiluokka on 30–34 vuotta. Synnyttäjien yleisin ikäluokka on 30–34 vuotta.

Tyyppiarvona voidaan pitää moodiluokan luokkakeskusta, eli $M_o = 32,5$ vuotta.

c) Luokitellusta aineistosta keskiarvon laskemiseen käytetään luokkakeskuksia. Alimmalla ikäluokalla (19-vuotiaat ja sitä nuoremmat synnyttäjät) ei ole alarajaa ja ylimmällä luokalla (40-vuotta täyttäneet synnyttäjät) ei ole ylärajaa, joten näiden luokkien luokkakeskusta eli luokkarajojen keskiarvoa ei voida laskea. Tästä syystä keskiarvon määrittäminen ei ole mahdollista pelkän taulukon tietojen perusteella.

d) Määritetään luokkien luokkakeskukset laskimen tai matematiikkaohjelmiston taulukkosovelluksessa. Määritetään keskiarvo tilastotoiminnoilla.

Luokkakeskukset
määritetään todellisten
luokkarajojen avulla,
niiden keskiarvona.

	A	B	C	D	E
1	Ikä (vuotta)	tod. alaraja	tod. yläraja	luokkakeskus	f
2	- 19				839
3	20 -24	20	25	22.5	7254
4	25 - 29	25	30	27.5	15576
5	30 - 34	30	35	32.5	17583
6	35 - 39	35	40	37.5	9298
7	40 -				2320

Keskiarvo määritetään
luokkakeskusten ja
luokkia vastaavien
frekvenssien perusteella.

Synnyttäjien keski-ikäksi saadaan $\bar{x} = 30,409 \dots \approx 30,4$ (vuotta).

Lasketaan, kuinka monta prosenttia tulos poikkeaa tunnetusta keski-ikästä 30,8 vuotta.

$$\frac{30,409 \dots}{30,8} = 0,987 \dots \approx 99 \%$$

Tulos on $100 \% - 99 \% = 1 \%$ tunnettua keski-ikää pienempi.

118.

a) Kirjoitetaan tai kopioidaan iät laskin- tai matematiikkaohjelmiston taulukkosovellukseen ja määritetään tunnusluvut tilastotoiminnoilla. Kyseessä on kokonaisaineisto, joten käytetään perusjoukon keskihajontaa σ .

Keskiarvo on $\bar{x} = 54,97 \dots \approx 55,0$ (vuotta).

Keskihajonta on $\sigma = 6,49 \dots \approx 6,5$ (vuotta)

b) Lasketaan arvon 70 normitettu arvo.

$$Z_{70} = \frac{70 - \bar{x}}{\sigma} = \frac{70 - 54,97 \dots}{6,49 \dots} = 2,311 \dots$$

Normitetun arvon itseisarvo on $|Z_{70}| = 2,311 \dots > 2$, joten Trumpin ikä 70 poikkeaa merkitsevästi keskiarvosta.

Ohje: Laskun tekeminen laskentataulukossa:

Iät on syötetty sarakkeeseen A.

	A	B
1	57	54.9778
2	61	6.4995
3	57	
4	57	2.3113
5	58	
6	57	
7	61	
8	54	

Soluun B1 on määritetty keskiarvo ja soluun B2 keskihajonta.

Soluun B4 on kirjoitettu laskukaava:
 $= (70 - B1) / B2$

c) Aineiston pienin arvo on 42 ja suurin arvo on 70. Vaihteluvälin pituus on $70 - 42 = 28$. Luokkien lukumääräksi on valittu kuusi.

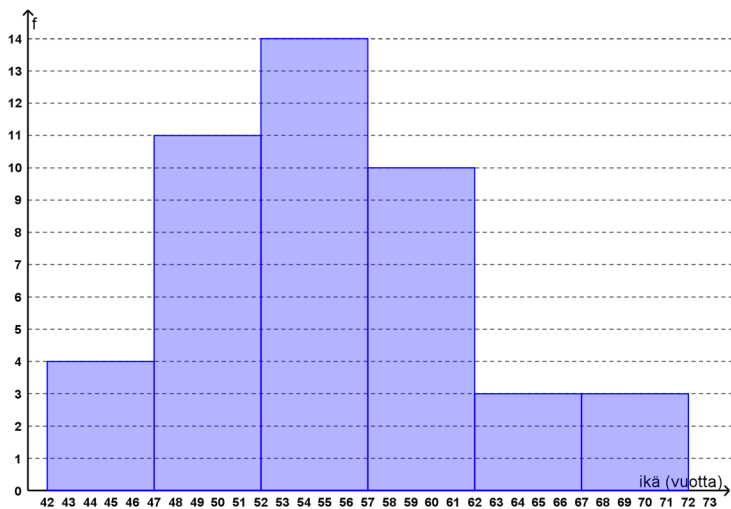
$$\frac{28}{6} = 4,66 \dots \approx 5$$

Valitaan luokkaväliksi 5 (vuotta). Valitaan ensimmäisen luokan alarajaksi aineiston pienin arvo 42. Ensimmäisen luokan todellinen alaraja on 42 (vuotta).

Luokittelu kannattaa tehdä matematiikkaohjelmiston taulukkosovelluksessa. Kirjoitetaan iät laskentataulukkoon ja tehdään yhden muuttujan analyysi. Asetetaan luokat käsin:

- valitaan luokituksen alku: 42
- valitaan luokkavälin pituus: 5

Kuvataan jakaumaa histogrammilla.



Ikä (vuotta)	Frekvenssi
42-46	4
47-51	11
52-56	14
57-61	10
62-66	3
67-71	3

d) Luokitellun aineiston tunnusluvut lasketaan luokkakeskusten ja luokkia vastaavien frekvenssien perusteella.

tod.alaraja	tod.yläraja	luokkakeskus	f
42	47	44.5	4
47	52	49.5	11
52	57	54.5	14
57	62	59.5	10
62	67	64.5	3
67	72	69.5	3

Luokitellun aineiston

keskiarvo on $\bar{x} = 55,16 \dots \approx 55,2$ (vuotta).

keskihajonta on $\sigma = 6,46 \dots \approx 6,5$ (vuotta)

Kun aineisto luokitellaan, tieto yksittäisistä havainnoista menetetään. Luokitellussa aineistossa kunkin luokan luokkakeskus edustaa kaikkia kyseisen luokan havaintoja. Siis esimerkiksi ensimmäisen luokan 42–46 (vuotta) luokkakeskus 44,5 (vuotta) edustaa kaikkia kyseiseen luokkaan kuuluvaa neljää havaintoa (havainnot ovat: 42, 43, 46 ja 46).

Tunnuslukua laskettaessa yksittäisten havaintojen arvot korvataan kyseisen luokan luokkakeskuksella.

Tästä syystä a-kohdassa määritetyt tunnuslukujen arvot (pyöristettynä: 55,0 vuotta ja 6,5 vuotta) poikkeavat d-kohdassa lasketuista arvoista (pyöristettynä: 55,2 vuotta ja 6,5 vuotta). Ero ei kuitenkaan ole suuri.

2.3 Jakaumien kuvailu ja vertailu

119.

Jakauma I on yksihiippuinen ja oikealle vino. Suurin osa havainnoista on vasemmassa laidassa (eli pienet arvot ovat yleisimpiä). Matala "häntä" vetää keskiarvoa mediaanista katsottuna oikealle. Keskiarvo voi olla harhaanjohtava, kun muuttujan jakauma on vino. Keskiarvo ja keskihajonta eivät sovi tämän jakauman kuvailuun.

Jakauma II on yksihiippuinen ja huippu on jakauman keskellä. Jakauma on varsin symmetrinen, joten keskiarvo on jakauman keskellä lähellä mediaania, ja suurin osa havainnoista on lähellä keskiarvoa. Keskiarvo ja keskihajonta sopivat tämä jakauman kuvailuun.

Jakaumassa III on tyhjä luokka 50–55 ja siksi poikkeavia havaintoja viimeisessä luokassa 55–60. Keskiarvon (ja keskihajonnan) käyttöä tunnuslukuina tulisi välttää, jos aineistossa on poikkeavia havaintoja.

120.

Jakaumassa III on korkea huippu jakauman keskellä moodiluokan 40–45 (kg) kohdalla. Jakauma on symmetrinen ja suurin osa havainnoista on jakauman keskellä eli havainnot ovat keskittyneet keskiarvon läheisyyteen: keskiarvosta poikkeavia havaintoja on vähän. Jakaumassa III keskihajonta on siis pieni.

Päätellään, että jakauman III tunnusluvut ovat B: $\bar{x} \approx 44$ (kg) ja $\sigma \approx 3,3$ (kg)

Myös **jakaumassa I** on huippu moodiluokan 45–50 (kg) kohdalla ja arvot ovat keskittyneet huipun läheisyyteen. Keskiarvosta poikkeavia arvoja on kuitenkin enemmän kuin jakaumassa III eli hajonta on suurempaa.

Päätellään, että jakauman I tunnusluvut ovat A: $\bar{x} \approx 46$ (kg) ja $\sigma \approx 6,2$ (kg)

Jakaumassa II ei ole nähtävissä yhtä huippua ja arvoissa esiintyy suurempaa hajontaa.

Jakauman II tunnusluvut ovat C: $\bar{x} \approx 46$ (kg) ja $\sigma \approx 8,3$ (kg)

Vastaukseksi saadaan: I – A, II – C, III – B

121.

Ratkaistaan tehtävä laskentataulukossa.

1) Havaintoarvot on kirjoitettu sarakkeeseen A.

3) Poikkeamat keskiarvosta on laskettu sarakkeeseen B.

	A	B
1	x	poikkeama
2	63	11
3	89	37
4	44	-8
5	58	6
6	32	-20
7	26	-26
8	52	0

3) Soluun B2 on kirjoitettu laskukaava: =A2 - 52
Kaava on kopioitu sarakkeessa B alaspäin.

Soluun B8 on laskettu poikkeamien summa, eli yläpuolella olevien solujen summa. Summa on 0.

2) Soluun A8 on laskettu havaintoarvojen keskiarvo.

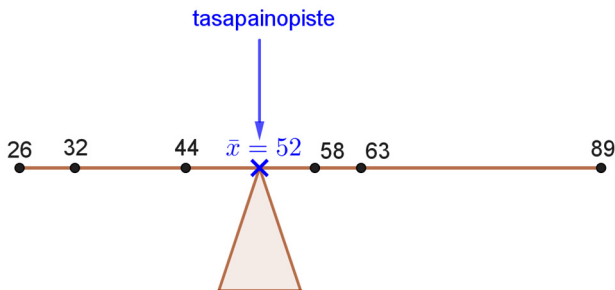
a) Havaintoarvojen keskiarvo on $\bar{x} = 52$.

b) Poikkeamien summa on 0.

c) Keskiarvo on luku, joka "tasapainottaa" havaintojen ja keskiarvon erotukset. Poikkeamat keskiarvosta summautuvat nolllaksi.

Havainto voidaan ilmaista myös seuraavasti: kun valitaan keskiarvoa suuremmat luvut, lasketaan kunkin luvun poikkeama keskiarvosta ja summataan poikkeamat, tulos on täsmälleen sama kuin jos tehtäisiin sama lasku keskiarvoa pienemmille luvuille.

Keskiarvo voidaan ajatella myös seuraavasti: joukko kilogramman painoja on sijoitettu eri kohtiin (painottoman) keinulaudan päälle. Painojen etäisyydet laudan päästä edustavat havaintoarvoja: pienet arvot ovat vasemmalla ja suuret oikealla. Keskiarvo on laudan vasemmasta päästä mitattu etäisyys tukipisteelle (tasapainopisteelle), jonka suhteen lauta on tasapainossa.



122.

Kirjoitetaan kuukausipalkat esimerkiksi laskinohjelmiston taulukkosovellukseen. Määritetään jakauman tunnusluvut tilastotoiminnoilla.

a) Kuukausipalkan

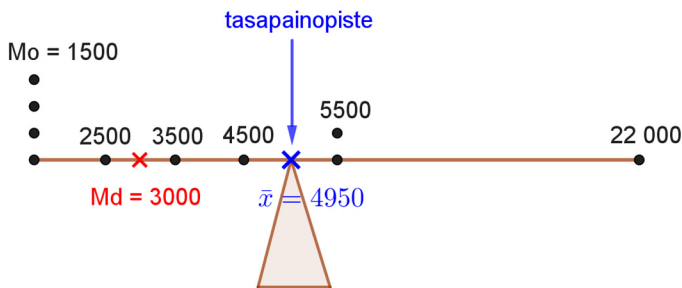
- keskiarvo on $\bar{x} = 4950$ €/kk
- mediaani on $M_d = 3000$ €/kk
- tyyppiarvo eli moodi on $M_o = 1500$ €/kk

Pääjohtajan väite keskipalkasta perustuu siis (aritmeettiseen) keskiarvoon, työntekijöiden väite moodiin ja sovintoneuvottelijan väite mediaaniin.

b) Kukin keskiluku (moodi, mediaani ja keskiarvo) kuvaa palkkajakauman keskikohtaa eli keskipalkkaa eri tavalla.

- **Moodi** ilmaisee, että kuukausipalkka 1500 €/kk esiintyy palkkajakaumassa useimmin: neljällä työntekijällä kymmenestä on tämä palkka.
- **Mediaani** ilmaisee, että puolet työntekijöistä ansaitsee korkeintaan 3000 €/kk ja puolet vähintään tämän verran.
- **Keskiarvo** ilmaisee arvon, jonka suhteen palkkojen poikkeamat summautuvat nolnaan (jakauman tasapainopisteen). Palkkajakaumassa on yksi muita selvästi suurempi palkka: 22 000 €/kk. Tämä yksi arvo kasvattaa keskiarvoa mediaaniin nähden.

Ei siis voida sanoa, että joku väitteistä olisi toista oikeampi. Usein keskimääräisestä arvosta puhuttaessa tarkoitetaan keskiarvoa. Keskiarvon herkkyyys poikkeaville arvoille on kuitenkin yksi syy siihen, että toisinaan jakauman keskikohdan ilmaisemiseen valitaan mediaani.



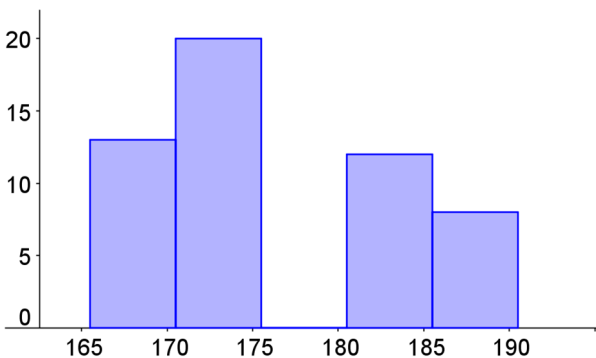
123.

Aineistot on luokiteltu tasavälisesti. Ensimmäisen luokan todellinen alaraja on 165,5 cm ja luokkavälin pituus $171 - 166 = 5$ (cm).

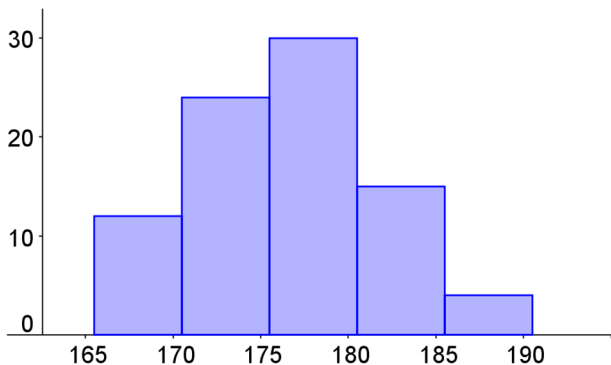
Esimerkiksi matematiikkaohjelmiston taulukkosovelluksessa on työväline luokitellun aineiston frekvenssijakauman kuvaamiseen. Kirjoitetaan frekvenssit laskentataulukkoon ja tehdään yhden muuttujan analyysi. Valitaan asetus "Luokat ja frekvenssi". Asetetaan:

- ensimmäisen luokan todellinen alaraja 165,5
- luokkavälin leveys 5

Ryhmän A pituusjakauma:



Ryhmän *B* pituusjakauma:



b) Luokitellusta aineistosta tunnusluvut arvioidaan luokkakeskusten perusteella. Keskilukujen (keskiarvo ja mediaani) arvioimista varten luokkakeskukset lasketaan laskentataulukossa. Määritetään tunnusluvut tilastotoiminnoilla luokkakeskusten ja niitä vastaavien frekvenssien perusteella.

Jakaumassa *A* keskiluvut ovat: $\bar{x} = 176,30 \dots \approx 176$ (cm) ja
 $Md = 173$ (cm)

Jakaumassa *B* keskiluvut ovat: $\bar{x} = 176,52 \dots \approx 177$ (cm) ja
 $Md = 178$ (cm)

Jakaumassa *B* keskiarvo ja mediaani ovat varsin lähellä toisiaan. Jakaumasta piirretyn histogrammin perusteella jakaumassa on keskilukujen kohdalla huippu ja jakauma on varsin symmetrinen keskiarvon suhteen: havainnot ovat keskittyneet keskilukujen ympärille. Keskipituutta voidaan kuvata yhtä hyvin keskiarvolla kuin mediaanilla.

Ryhmässä B keskipituus on $\bar{x} \approx 177$ cm.

Jakaumassa A keskiarvo ja mediaani poikkeavat enemmän toisistaan. Jakaumasta piirretyn histogrammin perusteella ryhmän pituudet ovat keskittyneet jakauman reunoille: lyhyitä ja pitkiä pituuksia esiintyy melko tasaisesti, mutta keskiarvon läheisyyteen sijoittuvat pituudet puuttuvat. Luokka 176–180 (cm), johon keskiarvo sijoittuu, on tyhjä. Tällaisessa jakaumassa mediaani kuvaa keskipituutta keskiarvoa paremmin.

Ryhmässä A keskipituus on $M_d = 173$ cm.

c) Jakaumassa A pituudet vaihtelevat välillä 166 cm – 190 cm. Puolet ryhmäläisistä on pituudeltaan alle 173 cm ja puolet yli tämän arvon. Pituuksien keskiarvo on 176 cm ja keskihajonta 7,2 cm. Keskiarvon mukaiset pituudet puuttuvat jakaumasta.

d) Jakaumassa B pituudet vaihtelevat välillä 166 cm – 190 cm. Pituuksien keskiarvo on 177 cm ja keskihajonta 5,3 cm. Pituudet jakautuvat melko symmetrisesti keskiarvon ympärille.

124.

a) Kirjoitetaan frekvenssit esimerkiksi matematiikkaohjelmiston taulukkosovellukseen. Lasketaan frekvenssien summa.

Yrityksessä A on 463 työntekijää.

Yrityksessä B on 374 työntekijää.

b) Luokitellusta aineistosta tunnusluvut arvioidaan luokkakeskusten perusteella. Lasketaan luokkakeskukset laskentataulukkoon todellisten luokkarajojen avulla. Määritetään tunnusluvut tilastotoiminnoilla luokkakeskusten ja niitä vastaavien frekvenssien perusteella.

Tod.alaraja	Tod.yläraja	Luokkakeskus	f (yritys A)	f (yritys B)
21	26	23.5	15	75
26	31	28.5	35	90
31	36	33.5	51	63
36	41	38.5	100	51
41	46	43.5	97	34
46	51	48.5	90	20
51	56	53.5	48	22
56	61	58.5	17	15
61	66	63.5	10	4

↑ ↑

Tunnusluvut määritetään luokkakeskusten ja luokkia vastaavien frekvenssien perusteella.

Yrityksessä A

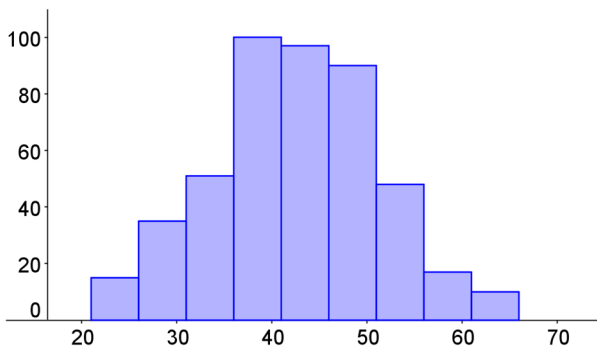
- keskiarvo on $\bar{x} = 42,52 \dots \approx 42,5$ (vuotta)
- keskihajonta on $\sigma = 8,80 \dots \approx 8,8$ (vuotta)
- mediaani on $Md = 43,5$ (vuotta)

Yrityksessä B

- keskiarvo on $\bar{x} = 35,18 \dots \approx 35,2$ (vuotta)
- keskihajonta on $\sigma = 10,29 \dots \approx 10,3$ (vuotta)
- mediaani on $Md = 33,5$ (vuotta)

c) Havainnollistetaan molempia ikäjakaumia histogrammilla.

Yrityksen A ikäjakauma:



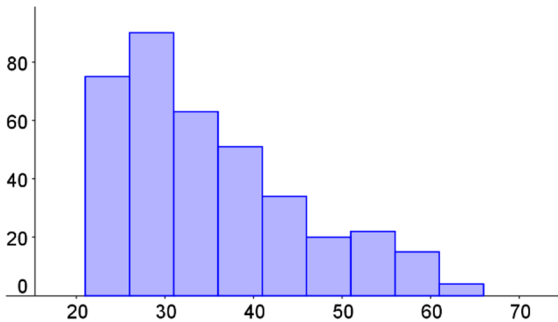
Tunnusluvut:

- $\bar{x} = 42,52 \dots \approx 43$
- $\sigma = 8,80 \dots \approx 8,8$
- $Q_1 = 38,5 \approx 39$
- $Md = 43,5 \approx 44$
- $Q_3 = 48,5 \approx 49$
- vaihteluväli: 21–65
- kvartiiliväli: 39–49

Ikäjakauma on varsin symmetrinen ja arvot ovat keskittyneet jakauman keskelle. Tunnusluvuista keskiarvo ja keskihajonta sopivat kuvaamaan jakaumaa. Mediaani ja muut kvartiilit antavat lisätietoa jakaumasta.

Yrityksessä A työntekijöiden ikä on välillä 21–65 vuotta. Iän keskiarvo on 43 vuotta ja keskihajonta 8,8 vuotta. Keskimäinen 50 % työntekijöistä on iältään 39–49 vuotta.

Yrityksen B ikäjakauma:



Tunnusluvut:

- $\bar{x} = 35,18 \dots \approx 35$
- $\sigma = 10,29 \dots \approx 10$
- $Q_1 = 28,5 \approx 29$
- $Md = 33,5 \approx 34$
- $Q_3 = 43,5 \approx 44$
- vaihteluväli: 21–65
- kvartiiliväli: 29–44

Ikäjakauma on oikealle vino ja arvot ovat keskittyneet vasemmalle (pieniin arvoihin). Keskiarvoa tulisi välttää tunnuslukuna, kun aineisto on vino, sillä se voi olla harhaanjohtava. Paremmin jakaumaa kuvaavat mediaani ja muut kvartiilit.

Yrityksessä B työntekijöiden ikä on välillä 21–65 vuotta. Iän mediaani on 34 vuotta, eli puolet työntekijöistä on iältään alle 34-vuotiaita. Iältään keskimäinen 50 % työntekijöistä on 29–44 vuotiaita.

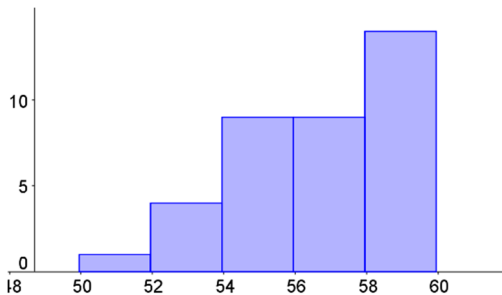
125.

Aineistot on luokiteltu tasavälisesti. Ensimmäisen luokan todellinen alaraja on 49,95 kg/hl ja luokkavälin pituus $52 - 50 = 2$ (kg/hl).

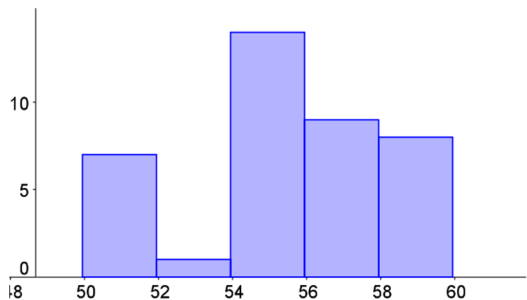
a) Luokitellun aineiston histogrammi voidaan piirtää esimerkiksi matematiikkaohjelmiston taulukkosovelluksessa. Kirjoitetaan frekvenssit laskentataulukkoon ja tehdään yhden muuttujan analyysi. Valitaan asetus "Luokat ja frekvenssi". Asetetaan:

- ensimmäisen luokan todellinen alaraja 49,95
- luokkavälin leveys 2

Hämeen painojakauma:



Etelä-Pohjanmaan painojakauma:



b) Tunnusluvut määritetään luokkakeskusten ja frekvenssien perusteella. Lasketaan luokkakeskukset ja määritetään tunnusluvut tilastotoiminnoilla.

Hämeessä:

- keskiarvo on $\bar{x} = 56,62 \dots \approx 56,6$ (kg/hl)
- mediaani on $M_d = 56,95 \approx 57,0$ (kg/hl)

Etelä-Pohjanmaalla:

- keskiarvo on $\bar{x} = 55,46 \dots \approx 55,5$ (kg/hl)
- mediaani on $M_d = 54,95 \approx 55,0$ (kg/hl)

Frekvenssijakauman perusteella Hämeen jakauma on vasemmalle vino. Arvot ovat keskittyneet oikealle (eli suuriin arvoihin). Keskiarvoa tuli välttää tunnuslukuna, kun aineisto on vino. Paremmin jakauman keskikohtaa kuvaa mediaani.

Hämeessä keskimääräinen hehtolitrainen paino on 57,0 kg/hl.

Myös Etelä-Pohjanmaan jakauma ei ole symmetrinen, mutta ei myöskään selvästi vino. Poikkeavia arvoja ei ole. Jakauman keskikohtaa voidaan siis kuvata keskiarvolla.

Keskimääräinen hehtolitrainen paino on 55,5 kg/hl.

c) Koko maassa $\bar{x} = 57,1$ (kg/hl) ja $\sigma = 3,9$ (kg/hl).

Normitetaan Etelä-Pohjanmaan alueen keskimääräinen hehtolitrainen paino 55 (kg/hl).

$$Z = \frac{55 - \bar{x}}{\sigma} = \frac{55 - 57,1}{3,9} = \frac{-2,1}{3,9} = -0,538 \dots$$

Normitetun arvon itseisarvo on $|Z| = 0,538 \dots < 2$, joten arvo ei poikkea merkitsevästi koko maan keskiarvosta.

126.

Kirjoitetaan tai kopioidaan havainnot matematiikkaohjelmiston taulukkosovellukseen. Tehdään yhden muuttujan analyysi.

a) Hirvitiheyden

- keskiarvo on $\bar{x} = 3,169 \dots \approx 3,2$ (hirveä/1000 ha)
- keskihajonta on $\sigma = 0,869 \dots \approx 0,9$ (hirveä/1000 ha).

b) Frekvenssijakaumaa varten aineisto on luokiteltava. Luokitteluun ei ole yhtä oikeaa tapaa. Jos valitaan suurempi määrä luokkia, frekvenssitaulukko antaa tarkempaa tietoa. Jos valitaan pienempi määrä luokkia, frekvenssitaulukko on helppolukuisempi.

Ohjelmisto asettaa luokittelun alkamaan aineiston minimiarvosta ja päättymään aineiston maksimiarvoon, ja asettaa luokkien lukumääräksi viisi. Tämän valinnan mukaisessa luokittelussa luokkarajoiksi tulee ”epätasalukuja”.

Muutetaan luokittelua asettamalla luokat käsin. Sopivan luokittelun valinnassa voidaan edetä esimerkiksi seuraavalla, kappaleessa 2.1 esitetyllä tavalla:

Tilastojen yhteenvetotaulukosta nähdään aineiston pienin ja suurin arvo.

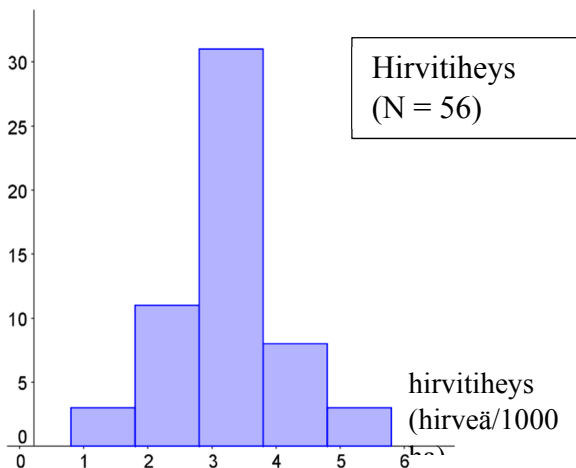
$$\text{min} = 0,8 \text{ ja } \text{max} = 5,4$$

Vaihteluvälin pituus on $5,4 - 0,8 = 4,6$ Sopiva luokkien lukumäärä on esimerkiksi **viisi**.

$$\frac{4,6}{5} = 0,92 \approx 1.$$

Valitaan:

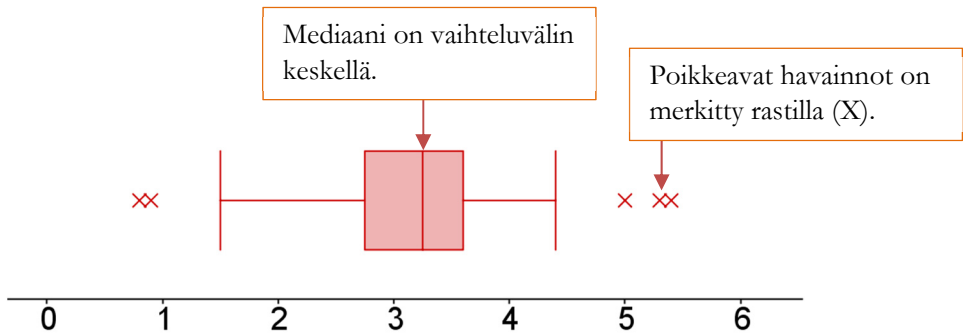
- ensimmäisen luokan alarajaksi aineiston minimiarvo **0,8**
- luokkavälin leveydeksi **1**.



Hirvitiheys (hirveä/1000 ha)	<i>f</i>
0,8-1,7	3
1,8-2,7	11
2,8-3,7	31
3,8-4,7	8
4,8-5,7	3

Hirvitiheys vaihtelee Suomessa välillä 0,8–5,4 hirveä/1000 ha. Jakauma on symmetrinen. Pienet ja suuret arvot ovat kuitenkin harvinaisia. Yleisimmin tiheys on 2–4 hirveä/1000 ha.

c) Laatikkokuva antaa hyvän yleiskuvan hirvitiheyden jakaumasta ja paljastaa poikkeavat havainnot.



Havainnoissa on sekä poikkeavan pieniä että poikkeavan suuria havaintoja.

d) Hirvitiheyden jakauman tunnusluvut nähdään tilastojen yhteenvetotaulukosta:

$$\bar{x} = 3,16 \dots \approx 3,2$$

$$\sigma = 0,86 \dots \approx 0,9$$

$$\min = 0,8$$

$$Q_1 = 2,75$$

$$Md = 3,25$$

$$Q_3 = 3,6$$

$$\max = 5,4$$

Hirvitiheys vaihtelee Suomessa alueittain välillä 0,8–5,4 hirveä/1000 ha. Vaihteluväli on suuri, mutta joukossa on poikkeavan pieniä ja suuria arvoja. Suomessa on siis alueita, joilla hirvitiheys on poikkeuksellisen pieni tai suuri.

Suurin osa havainnoista sijoittuu vaihteluvälin keskivaiheille. Puolessa alueista tiheys on välillä 2,75–3,6 hirveä/1000 ha, ja keskimääräinen hirvitiheys on noin 3 hirveä/1000 ha. Vaikka jakaumassa esiintyy poikkeavia havaintoja, ne ovat jakauman molemmissa äärilaidoissa ja tasapainottavat toistensa vaikutusta keskiarvoon. Tästä syystä mediaani ja keskiarvo ovat lähellä toisiaan.

127.

a) Määritetään tunnusluvut ohjelmiston tilastotoiminnoilla.

Alueen A jakaumassa

- keskiarvo on $\bar{x} = 51,57 \approx 51,6$ (ha)
- keskihajonta on $\sigma = 25,05 \dots \approx 25,1$ (ha)

Alueen B jakaumassa

- keskiarvo on $\bar{x} = 56,88 \dots \approx 56,9$ (ha)
- keskihajonta on $\sigma = 19,37 \dots \approx 19,4$ (ha)

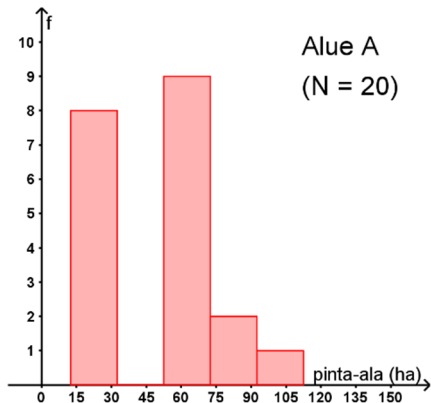
Metsätilat ovat keskimäärin suurempia alueella B ($56,9 > 51,6$).

Suurempaa hajontaa on alueen A pinta-aloissa ($25,1 > 19,4$).

b) Luokittelu voidaan tehdä esimerkiksi matematiikkaohjelmiston taulukkosovelluksessa.

Alueen A pienin arvo on $\min = 12,5$. Asetetaan luokat käsin:

- luokittelun alku 12,5
- luokkavälin leveys 20

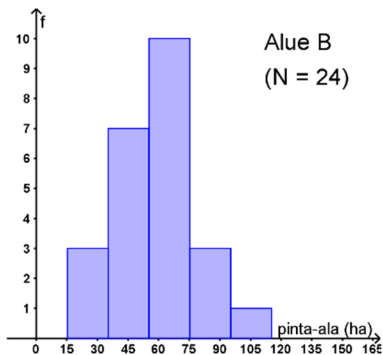


Frekvenssitaulukossa luokkarajat voidaan esittää pyöristettyinä lukuina.

Pinta-ala (ha)	<i>f</i>
13-32	8
33-52	0
53-72	9
73-92	2
93-112	1

Alueen B pienin arvo on min = 15,3. Asetetaan luokat käsin:

- luokittelun alku 15,3
- luokkavälin leveys 20



Frekvenssitaulukossa edellisen luokan **yläraja** on eri luku kuin seuraavan luokan **alaraja**.

Pinta-ala (ha)	<i>f</i>
15,3-35,2	3
35,3-55,2	7
55,3-75,2	10
75,3-95,2	3
95,3-115,3	1

Alueen A jakaumassa on pieniä havaintoja ensimmäisessä luokassa 13–32 (ha), mutta toinen luokka on tyhjä ($f = 0$). Keskiarvo 51,6 ha sijoittuu tyhjään luokkaan. Keskiarvon suuruisia havaintoja ei siis ole. Suuri osa havainnoista on moodiluokassa 53–72 (ha), keskiarvoa suurempia.

Alueen B jakaumassa suuri osa havainnoista on moodiluokassa 55,3–75,2 (ha). Jakauma on suhteellisen symmetrinen ja suuri osa havainnoista on keskiarvon 56,9 ha lähellä.

c) Laatikkokuvaaja havainnollistaa viiden tunnusluvun yhteenvetoa.

Alue A:

min = 12,5

$Q_1 = 27,05 \approx 27$

Md = 58,2 $\approx$ 59

$Q_3 = 68,15 \approx 68$

max = 101,3

Alue B:

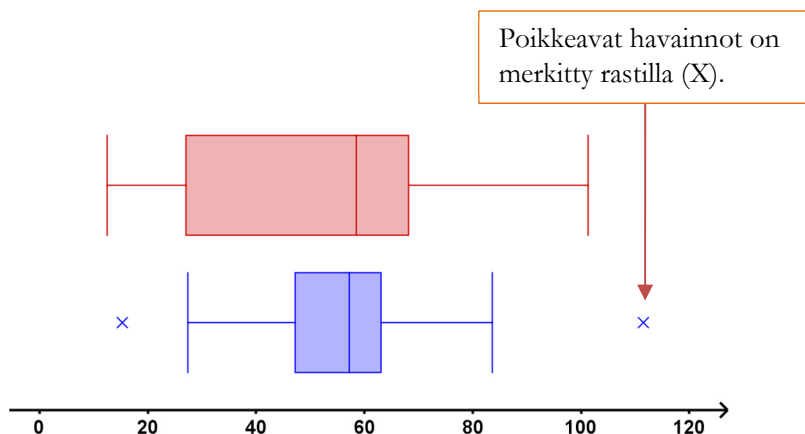
min = 15,3

$Q_1 = 47,2 \approx 47$

Md = 57,2 $\approx$ 57

$Q_3 = 63,05 \approx 63$

max = 111,5



Jakaumassa *A* pinta-alan vaihteluväli on 12,5 ha – 101,3 ha. Mediaani 59 ha on likimain vaihteluvälin keskellä, selvästi keskiarvoa 51,6 ha suurempi. Poikkeavia havaintoja ei ole. Pinta-alaltaan keskimmäinen 50 % sijoittuu välille 27–68 ha.

Jakaumassa *B* pinta-alan vaihteluväli on 15,3 ha – 111,5 ha. Aineistossa on pienin arvo 15,3 sekä suurin arvo 111,5 ovat poikkeavia havaintoja. Mediaani on 57 ha on lähellä jakauman keskiarvoa 56,9 ha. Jakauma on varsin symmetrinen, ja vaikka aineistossa on poikkeavia havaintoja, nämä ovat jakauman molemmissa äärilaidoissa tasapainottaen toistensa vaikutuksia keskiarvoon.

Alueen *A* mediaani on suurempi kuin vastaava tunnusluku alueella *B* (59 > 57) vaikka keskiarvoissa järjestys oli toinen (56,9 > 51,6). Tämä selittyy sillä, että alueen *A* jakauma ei ole symmetrinen vaan jakaumassa on huippu ensimmäisen luokan kohdalla. Jos poikkeavat havainnot jätetään huomioimatta, alueen *B* pinta-alojen vaihteluväli on kapeampi kuin alueen *A* vaihteluväli. Pinta-alaltaan keskimmäinen 50 % sijoittuu välille 47 ha – 63 ha, eli kapeammalle välille kuin jakaumassa *A*.

d) Normitetaan alueen B pienin arvo $\min = 15,3$.

$$Z = \frac{15,3 - \bar{x}}{\sigma} = \frac{15,3 - 56,88 \dots}{19,37 \dots} = -2,145 \dots$$

Pienin arvo poikkeaa siis noin 2,1 keskihajontaa keskiarvosta.

Normitetun arvon itseisarvo on $|Z| = 2,145 \dots > 2$, joten pienin arvo poikkeaa merkitsevästi keskiarvosta.

Normitetaan B suurin arvo $\max = 111,5$.

$$Z = \frac{111,5 - \bar{x}}{\sigma} = \frac{111,5 - 56,88 \dots}{19,37 \dots} = 2,818 \dots$$

Suurin arvo poikkeaa siis noin 2,8 keskihajontaa keskiarvosta.

Normitetun arvon itseisarvo on $|Z| = 2,818 \dots > 2$, joten suurin arvo poikkeaa merkitsevästi keskiarvosta.

128.

Kirjoitetaan tai kopioidaan pinta-alat matematiikkaohjelmiston taulukkosovellukseen. Tehdään yhden muuttujan analyysi.

a) EU-jäsenmaiden pinta-alan

- keskiarvo on $\bar{x} = 157,29 \dots \approx 160$ (/1000 km²)
- keskihajonta on $\sigma = 167,73 \dots \approx 170$ (/1000 km²).

Eli keskiarvo on 160 000 km² ja mediaani 170 000 km².

b) Aineiston luokitteluun ei ole yhtä oikeaa tapaa. Jos valitaan suurempi määrä luokkia, frekvenssitaulukko antaa tarkempaa tietoa. Jos valitaan pienempi määrä luokkia, frekvenssitaulukko on helppolukuisempi. Ohjelmisto asettaa luokittelun alkamaan aineiston minimiarvosta ja päättymään aineiston maksimiarvoon, ja asettaa luokkien lukumääräksi viisi. Tämän valinnan mukaisessa luokittelussa luokkarajoiksi tulee ”epätasalukuja”.

Muutetaan luokittelua asettamalla luokat käsin. Sopivan luokittelun valinnassa voidaan edetä esimerkiksi seuraavalla, kappaleessa 2.1 esitetyllä tavalla:

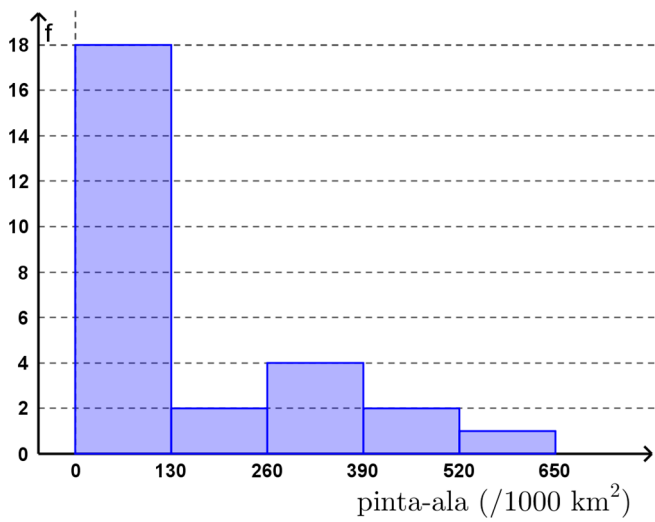
Tilastojen yhteenvetotaulukosta nähdään aineiston pienin ja suurin arvo.

$$\text{min} = 0,3 \text{ ja } \text{max} = 633 \text{ (/1000 km}^2\text{)}$$

Vaihteluvälin pituus on $633 - 0,3 = 632,7$ (/1000 km²). Valitaan luokkien lukumääräksi **viisi** luokkaa.

$$\frac{632,7}{5} = 126,54 \approx 130.$$

Valitaan ensimmäisen luokan alarajaksi 0 ja luokkavälin leveydeksi 130.

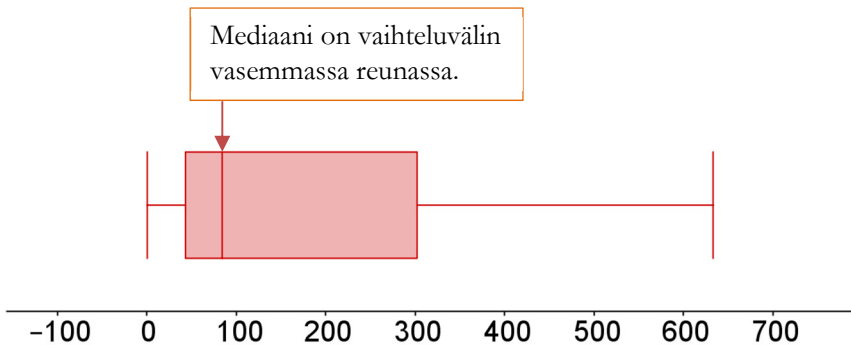


Pinta-ala (/1000 km ²)	<i>f</i>
0–129	18
130–259	2
260–389	4
390–519	2
520–650	1

Kuvaillaan jakaumaa sanallisesti:

EU-jäsenmaiden pinta-alat vaihtelevat välillä $300 \text{ km}^2 - 633\,000 \text{ km}^2$, eli vaihteluvälin pituus on suuri. Pinta-alojen jakauma on painottunut voimakkaasti vasemmalle, eli jakauma on oikealle vino: pienet muuttujan arvot ovat paljon yleisimpiä kuin suuret arvot.

Laatikkokuva, joka on luokittelusta riippumaton, antaa myös hyvän yleiskuvan pinta-alojen jakaumasta: vaihteluväli on suuri, mutta arvot ovat painottuneet vasempaan reunaan. Mediaani on vaihteluvälin vasemmassa reunassa eli pienin 50 % pinta-aloista sijoittuu paljon kapeammalle välille kuin suurin 50 % pinta-aloista.



c) Mediaani on $M_d = 84\,000 \text{ km}^2$ kun taas keskiarvo on $\bar{x} \approx 160\,000 \text{ km}^2$. Keskiluvut poikkeavat toisistaan, koska pinta-alan jakauma on oikealle vino: arvot ovat painottuneet voimakkaasti vasemmalle (pitkä matala "häntä" oikealle). Yksittäiset suuret arvot "vetävät" keskiarvoa mediaanista katsottuna oikealle.

d) Vinossa jakaumassa mediaani on yleensä keskiarvoa parempi tunnusluku kuvaamaan "keskimääräistä" arvoa.

Siis mediaani $M_d = 84\,000 \text{ km}^2$ kuvaa EU-jäsenmaisen keskimääräistä pinta-alaa keskiarvoa paremmin.

129.

a) Kirjoitetaan tai kopioidaan juoksuajat laskentataulukkaan ja määritetään tunnusluvut tilastotoiminnoilla.

- keskiarvo on $\bar{x} = 293,68 \dots \approx 294$ (min)
- keskihajonta on $\sigma = 43,22 \dots \approx 43,2$ (min)

b) Luokittelu voidaan tehdä esimerkiksi matematiikkaohjelmiston taulukkosovelluksessa.

Ohjelmisto asettaa luokittelun alkamaan aineiston minimiarvosta ja päättymään aineiston maksimiin ja asettaa luokkien lukumääräksi viisi. Tämän valinnan mukaisessa luokittelussa luokkarajoiksi tulee ”epätasalukuja”.

Muutetaan luokittelua asettamalla luokat käsin. Edetään luokittelussa kappaleessa 2.1 esitetyllä tavalla.

Aineiston pienin ja suurin arvo nähdään tilastojen yhteenvetotaulukosta.

min = 201 (min) ja max = 408 (max)

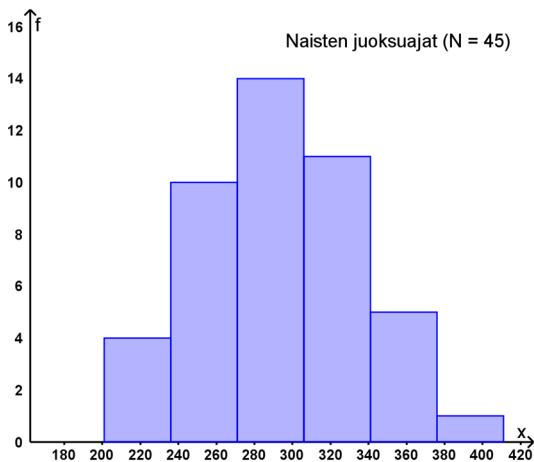
Vaihteluvälin pituus on $408 - 201 = 207$. Sopiva luokkien lukumäärä on **esimerkiksi kuusi**. (Huom: Vastauksissa luokkien lukumääräksi on valittu 7, eli eri luokittelu.)

Luokkavälin pituudeksi saadaan

$$\frac{207}{6} = 34,5 \approx 35 \text{ (min)}.$$

Asetetaan siis:

- luokittelun alku **201**
- luokkavälin leveys **35**



Aika (min)	Frekvenssi
201-235	4
236-270	10
271-305	14
306-340	11
341-375	5
376-410	1

b) Tunnusluvut nähdään tilastojen yhteenvetotaulukosta.

$$\bar{x} = 293,68 \dots \approx 294 \text{ (min)}$$

$$\sigma = 43,22 \dots \approx 43,2 \text{ (min)}$$

$$\text{min} = 201 \text{ (min)}$$

$$Q_1 = 263,5 \approx 264 \text{ (min)}$$

$$M_d = 294 \text{ (min)}$$

$$Q_3 = 327,5 \approx 328$$

$$\text{max} = 408 \text{ (min)}$$

$$\text{vaihteluväli: } 201\text{--}408 \text{ (min)}$$

$$\text{kvartiiliväli: } 264\text{--}328 \text{ (min)}$$

Naisten juoksuaika ikäryhmässä 12–18-vuotiaat vaihtelee välillä 201 min – 408 min. Keskimäinen 50 % juoksuaajoista oli välillä 264 min – 328 min. Juoksuaikojen jakauma on varsin symmetrinen ja suurin osa havainnoista on keskittynyt keskiarvon 294 min.

c) Muutetaan nuorimman juoksijan juoksuaika minuuteiksi. Merkintä 5:56:42 tarkoittaa, että juoksuaika oli 5 h 56 min 42 s.

$$5 \cdot 60 \text{ min} + 56 \text{ min} + \frac{42}{60} \text{ min} = 356,7 \text{ min}$$

Juoksuaika on suurempi kuin yläkvartiili $Q_3 \approx 328$ (min), joten aika kuuluu jakauman ylimpään 4. neljännekseen. Nuorin juoksija siis kuului hitaimpaan (viimeisenä maaliin tulleeseen) neljännekseen juoksijoista.

Normitetaan aika 356,7 min. Juoksuaikojen keskiarvo oli $\bar{x} = 293,68 \dots$ ja keskihajonta $\sigma = 43,22 \dots$

$$Z = \frac{356,7 - \bar{x}}{\sigma} = \frac{356,7 - 293,68 \dots}{43,22 \dots} = 1,457 \dots$$

Normitetun arvon itseisarvo on $|Z| = 1,457 \dots < 2$, joten nuorimman juoksijan juoksuaika ei poikennut merkitsevästi ikäluokan keskiarvosta.

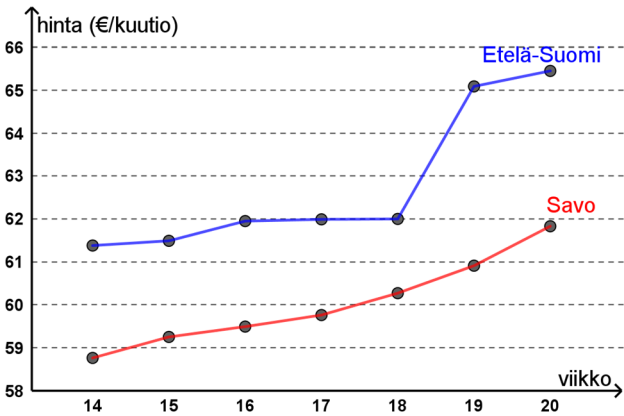
130.

a) Aikasarjan tarkastelu kannattaa aina aloittaa viivadiagrammin piirtämisellä.

Kirjoitetaan x -arvot eli viikkonumerot laskentataulukkoon (sarake A). Poimitaan aineistosta kuusitukkia (KuT) koskevat havainnot Etelä-Suomessa (sarake B) ja Savossa (sarake C).

Piirretään viivadiagrammit.

	A	B	C
1	Kuusitukki	(€/kuutio)	
2	Viikkonro	Etelä-Suomi	Savo
3	20	65.45	61.83
4	19	65.09	60.91
5	18	62	60.27
6	17	61.99	59.76
7	16	61.95	59.49
8	15	61.49	59.25
9	14	61.38	58.76



Kuvaajasta nähdään (taulukoituja arvoja havainnollisemmin), että Etelä-Suomessa kuusitukin kuutiohintaa on koko ajanjakson ajan ollut Savon hintaa korkeampi.

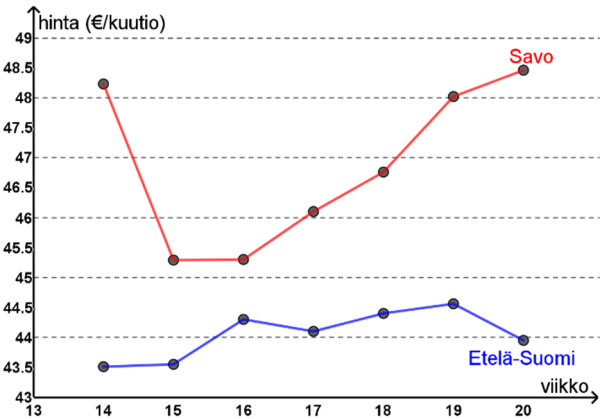
Molemmilla alueilla kuusitukin hinta oli nousussa koko ajanjakson ajan. Savossa hinnannousu oli tasaista (lineaarista). Etelä-Suomen hintojen nousu oli viikoilla 14–18 vähäistä, mutta viikolla 19 hinta nousi suhteellisesti eniten.

$$\frac{65,09}{62} = 1,0498 \dots \approx 105,0 \% \text{ eli hinta nousi } 5,0 \%.$$

b) Kirjoitetaan x-arvot eli viikkonumerot laskentataulukkoon (sarake A). Poimitaan aineistosta koivutukkia (KoT) koskevat havainnot Etelä-Suomessa (sarake B) ja Savossa (sarake C).

Piirretään molempien alueiden hintakehitystä kuvaavat viivadiagrammit samaan kuvaan.

	A	B	C
1	Koivutukki	(€/kuutio)	
2	Viikkonro	Etelä-Suomi	Savo
3	20	43.95	48.46
4	19	44.56	48.02
5	18	44.4	46.76
6	17	44.1	46.1
7	16	44.3	45.3
8	15	43.55	45.29
9	14	43.51	48.23



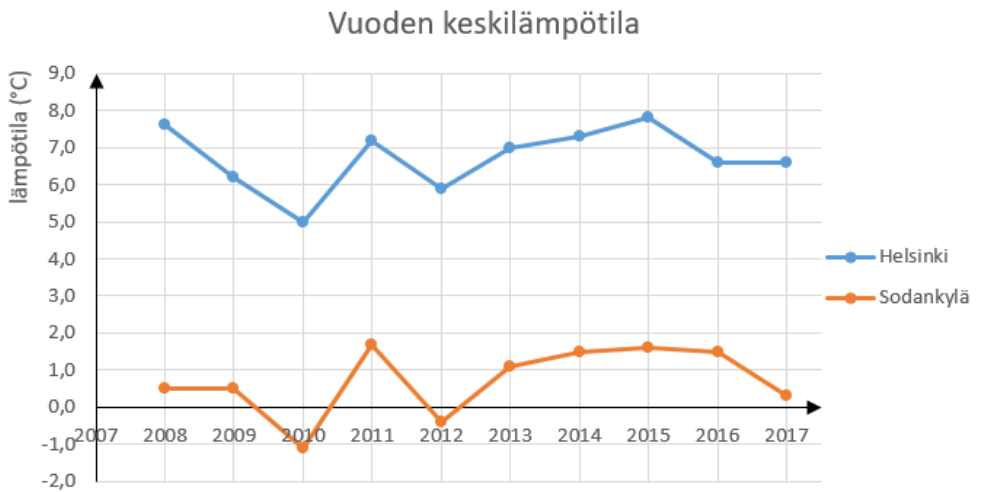
Alueellisten hintojen vertailuun sopivia tunnuslukuja ovat vaihteluväli, ajanjakson keskiarvo sekä keskihajonta. Määritetään tunnusluvut tilastotoiminnoilla.

	Etelä-Suomi	Savo
pienin arvo	min = 43,51	min = 45,29
suurin arvo	max = 44,56	max = 48,46
vaihteluväli	43,51–44,56	45,29–48,46
keskiarvo	$\bar{x}$ = 44,05 ... $\approx$ 44,10	$\bar{x}$ = 46,88 ... $\approx$ 46,90
keskihajonta	σ = 0,377 ... $\approx$ 0,4	σ = 1,268 ... $\approx$ 1,2

Etelä-Suomessa koivutukin hinta oli koko ajanjakson ajan Savon hintaa matalampi. Etelä-Suomessa hinta vaihteli välillä 43,51–44,56 €/kuutio kun Savossa vaihteluväli oli hieman suurempi, 45,29–48,46 €/kuutio. Etelä-Suomessa ajanjakson keskihinta oli 44,10 €/kuutio kun Savossa keskiarvo oli suurempi, 46,90 €/kuutio. Etelä-Suomessa hintojen keskihajonta oli pientä (0,4 €/kuutio) kun taas Savon hinnoissa esiintyi suurempaa hajontaa (1,2 €/kuutio).

131.

a) Piirretään molempien kaupunkien vuoden keskilämpötiloista viivadiagrammi samaan kuvaajaan esimerkiksi taulukkolaskentaohjelmalla.



b) Määritetään Helsingissä tehtyjen havaintojen pienin ja suurin arvo.

$$\min = 5,0 \text{ (}^{\circ}\text{C)}$$

$$\max = 7,8 \text{ (}^{\circ}\text{C)}$$

Vaihteluvälin pituus oli $7,8 \text{ }^{\circ}\text{C} - 5,0 \text{ }^{\circ}\text{C} = 2,8 \text{ }^{\circ}\text{C}$.

Määritetään Rovaniemessä tehtyjen havaintojen pienin ja suurin arvo.

$$\min = -1,1 \text{ (}^{\circ}\text{C)}$$

$$\max = 1,7 \text{ (}^{\circ}\text{C)}$$

Vaihteluvälin pituus oli $1,7 \text{ }^{\circ}\text{C} - (-1,1 \text{ }^{\circ}\text{C}) = 2,8 \text{ }^{\circ}\text{C}$.

Vaihteluvälin pituus on kaupungeissa yhtä suuri. Kaupunkien keskilämpötiloissa esiintyi siis tämän perusteella yhtä suurta vaihtelua.

c) Verrataan Helsingin vuoden 2015 keskilämpötilaa $7,8 \text{ }^{\circ}\text{C}$ vuoden 2014 arvoon $7,3 \text{ }^{\circ}\text{C}$.

$$\frac{7,8 \text{ }^{\circ}\text{C}}{7,3 \text{ }^{\circ}\text{C}} = 1,0684 \dots \approx 106,8 \%$$

Helsingissä vuoden 2015 keskilämpötila nousi $106,8 \% - 100 \% = 6,8 \%$ edellisvuoteen verrattuna.

Verrataan Rovaniemen vuoden 2015 keskilämpötilaa $1,6 \text{ }^{\circ}\text{C}$ vuoden 2014 arvoon $1,5 \text{ }^{\circ}\text{C}$.

$$\frac{1,6 \text{ }^{\circ}\text{C}}{1,5 \text{ }^{\circ}\text{C}} = 1,0666 \dots \approx 106,7 \%$$

Rovaniemellä vuoden 2015 keskilämpötila nousi $106,7 \% - 100 \% = 6,7 \%$ edellisvuoteen verrattuna.

d) Helsingin keskilämpötila on koko ajanjakson ajan Sodankylän keskilämpötilaa suurempi. Keskilämpötila on kehittynyt molemmissa kaupungeissa samansuuntaisesti. Ainoastaan vuonna 2009 Helsingin keskilämpötila laski edellisvuoteen verrattuna, kun Sodankylässä ei tapahtunut muutosta edellisvuoteen verrattuna. Toisaalta, vuonna 2017 Sodankylän keskilämpötila laski edellisvuoteen verrattuna.

$$\frac{0,3}{1,5} = 0,2 = 20 \%$$

Sodankylän keskilämpötila laski $100 \% - 20 \% = 80 \%$, kun Helsingissä vuoden 2017 keskilämpötila pysyi edellisvuoden tasolla.

132.

Kirjoitetaan vuosiluvut ja niitä vastaavat liikevoitot esimerkiksi matematiikkaohjelmiston taulukkosovellukseen.

a) Määritetään liikevoittojen keskiarvo ja keskihajonta tilastotoiminnoilla tai laskentataulukossa soluviittauksilla.

- keskiarvo on $\bar{x} = 2,61$ (milj. €)
- keskihajonta on $\sigma = 0,2782 \dots \approx 0,278$ (milj. €)

b) Lasketaan vuoden 2013 liikevoiton 2,15 (milj. €) normitettu arvo.

$$Z = \frac{2,15 - \bar{x}}{\sigma} = \frac{2,15 - 2,61}{0,278 \dots} = -1,653 \dots$$

Normitetun arvon itseisarvo on $|Z| = 1,653 \dots < 2$, joten vuoden 2013 arvo ei poikkea merkitsevästi keskiarvosta.

c) Muodostetaan indeksisarja laskentataulukossa. Perusarvoksi on valittu vuoden 2013 arvo, joten muiden vuosien arvoja verrataan lukuun 2,15.

	A	B	C
1	Vuosi	Liikevoitto (milj.€)	Indeksi (2013=100)
2	2013	2.15	100
3	2014	2.45	113.9535
4	2015	2.9	134.8837
5	2016	2.85	132.5581
6	2017	2.7	125.5814

Soluun C2 on kirjoitettu laskukaava:

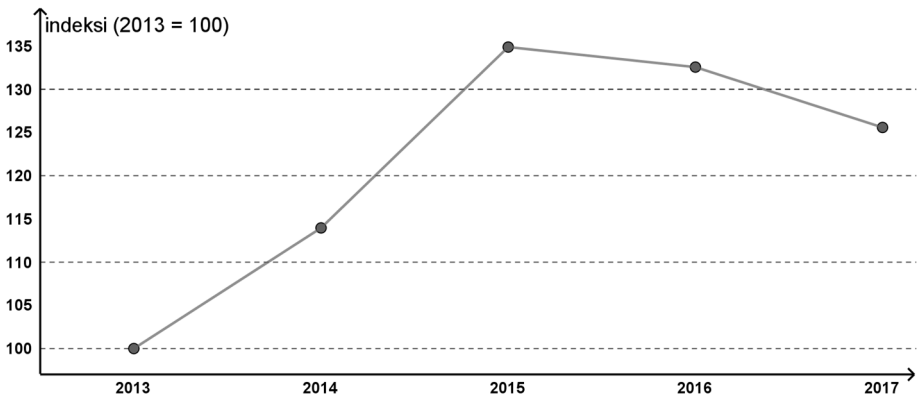
=B2/2.15*100

Kaava on kopioitu sarakkeessa C alaspäin.

Indeksin arvo ilmaistaan pyöristettynä prosenttilukuna ilman prosenttimerkkiä.

Vuosi	Indeksi (2013 = 100)
2013	100,0
2014	114,0
2015	134,9
2016	132,6
2017	125,6

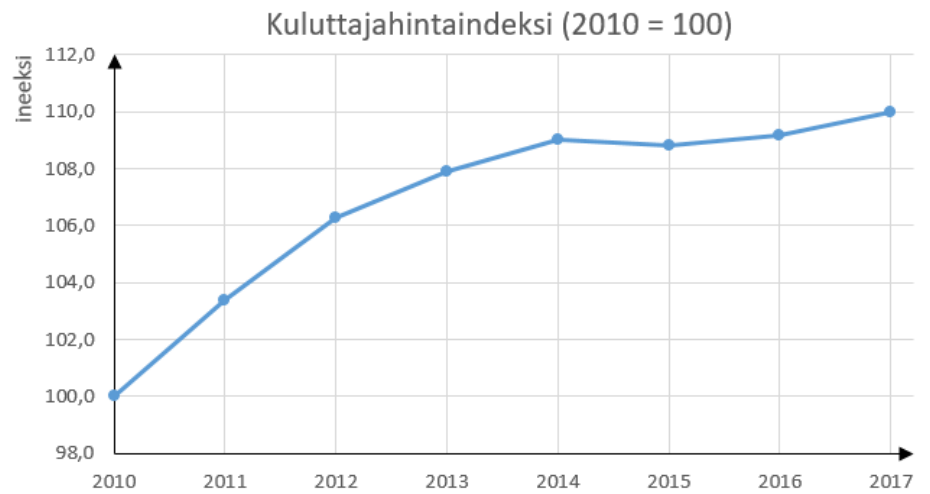
Kuvataan indeksisarjaa viivadiagrammilla.



Yrityksen liikevoitto (milj.€ /vuosi) nousi aikavälillä 2013–2015 yhteensä noin 35 %. Tämän jälkeen liikevoitto laski vuoteen 2017 saakka. Lasku oli kuitenkin vähäisempää, ja vuoden 2017 liikevoitto oli noin 26 % suurempi kuin ajanjakson alussa vuonna 2013.

133.

a) Kirjoitetaan vuosiluvut ja niitä vastaavat indeksipisteet laskentataulukkoon. Piirretään viivadiagrammi.



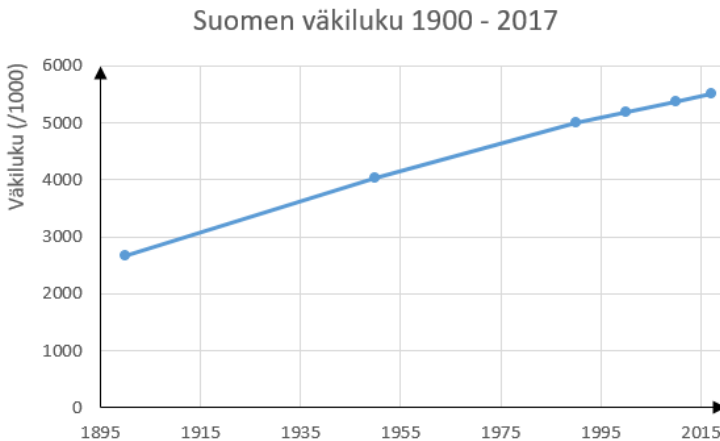
Vuosi	KHI (2010 = 100)
2010	100,0
2011	103,4
2012	106,3
2013	107,9
2014	109,0
2015	108,8
2016	109,2
2017	110,0

b) Indeksisarja kuvaa prosentuaalista muutosta perusajankohtaan verrattuna. Vuoden 2017 indeksipiste on 110,0 eli hintataso oli kasvanut $110,0 - 100 = 10$ prosenttia vuodesta 2010.

134.

a) Kirjoitetaan vuosiluvut ja niitä vastaavat muuttujan arvot laskentataulukkaan.

Piirretään viivadiagrammi.



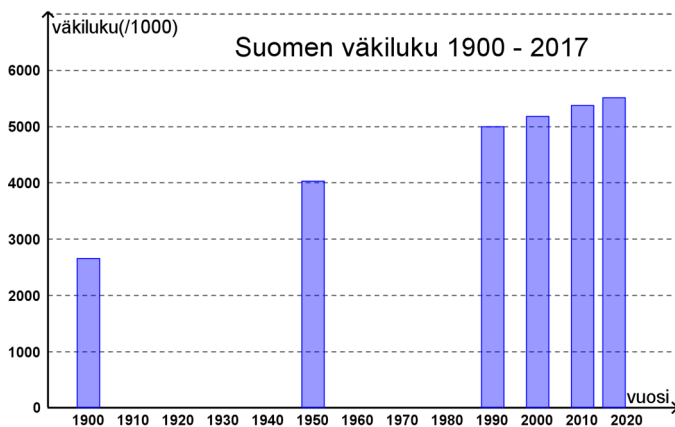
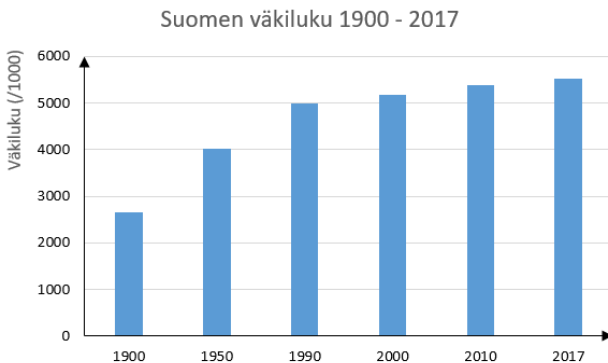
b) Väkiluku on ilmaistu tuhansina henkilöinä (eli /1000). Vuoden 2017 lopussa väkiluku on

$$5513 \cdot 1000 = 5\,513\,000$$

(eli noin 5,5 miljoonaa).

c) Taulukossa muuttujan arvojen yksikkö on ilmaistu merkinnällä /1000 eli taulukon lukuarvot ilmaisevat tuhansia ihmisiä.

Mittausajankohdat eivät ole tasaisin välein. Kahden ensimmäisen mittausajankohdan välillä on 50 vuotta, seuraava väli on 40 vuotta, sitten 10, 10 ja 7 vuotta. Jos väestön määrää kuvataan pylväsdiagrammilla, kuvaaja voi käytetystä ohjelmistosta riippuen joko huomioida epätasaiset mittausajankohdat (kuva oikealla) tai piirtää pylväät tasavälisesti (kuva vasemmalla), mikä saattaa antaa tilastosta harhaanjohtavan kuvan.



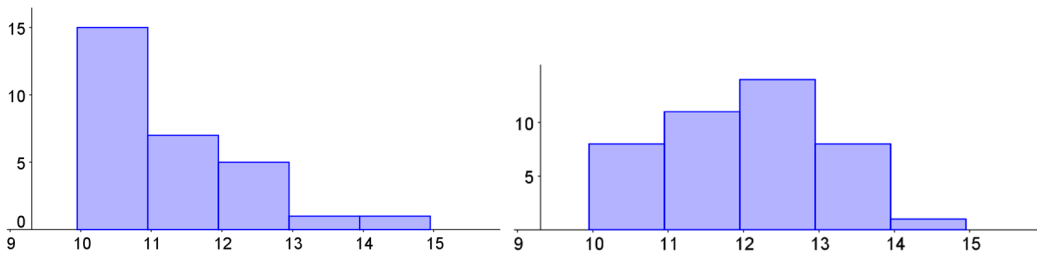
135.

a) Luokitellun aineiston histogrammi voidaan piirtää esimerkiksi matematiikkaohjelmiston taulukkosovelluksessa.

Kirjoitetaan frekvenssit laskentataulukkoon ja tehdään yhden muuttujan analyysi. Valitaan asetus "Luokat ja frekvenssi". Syötetään:

- ensimmäisen luokan todellinen alaraja 9,95
- luokan leveys 1

Uudenmaan jakauma ($N = 29$): Etelä-Pohjanmaan jakauma ($N = 42$):



b) Luokitellun aineiston tunnusluvut määritetään luokkakeskusten avulla. Lasketaan luokkakeskukset laskentataulukossa.

	A	B	C	D	E
1	Tod.alaraja	Tod.yläraja	Luokkakeskus	f (Uusimaa)	f (Et.-Pohj.)
2	9.95	10.95	10.45	15	8
3	10.95	11.95	11.45	7	11
4	11.95	12.95	12.45	5	14
5	12.95	13.95	13.45	1	8
6	13.95	14.95	14.45	1	1

Määritetään tunnusluvut luokkakeskusten ja niitä vastaavien frekvenssien avulla tilastotoiminnoilla.

	Uusimaa	Etelä-Pohjanmaa
keskiarvo:	$\bar{x} = 11,27 \dots \approx 11,3 \text{ (\%)}$	$\bar{x} = 12,04 \dots \approx 12,0 \text{ (\%)}$
keskihajonta:	$\sigma = 1,05 \dots \approx 1,1 \text{ (\%)}$	$\sigma = 1,07 \dots \approx 1,1 \text{ (\%)}$
mediaani:	Med = 10,45 $\approx$ 10,5 (%)	Med = 12,45 $\approx$ 12,5 (%)

c) Uudeltamaalta kerättiin yhteensä 29 näytettä. Kauran valkuaisainepitoisuus näytteissä vaihteli välillä 10 % – 14,9 %. Suurimmassa osassa näytteitä (yhteensä 15 näytettä) pitoisuus oli välillä 10 % – 10,9 % ja yli 13 %:n pitoisuuksia tavattiin vain kahdessa näytteessä.

Uudenmaan jakauma on oikealle vino ja arvot ovat keskittyneet vasemmalle (pieniin arvoihin). Keskiarvoa (ja siis myös keskihajontaa) tulisi välttää tunnuslukuna, kun aineisto on vino. Paremmin jakauman keskikohtaa kuvaa mediaani.

Uudellamaalla kauran valkuaispitoisuus oli keskimäärin 10,5 %.

d) Etelä-Pohjanmaalta kerättiin yhteensä 42 näytettä. Kauran valkuaisainepitoisuus näytteissä vaihteli välillä 10 % – 14,9 %. Suurimmassa osassa näytteitä (yhteensä 11 + 14 = 25 näytettä) pitoisuus oli välillä 11 % – 12,9 %.

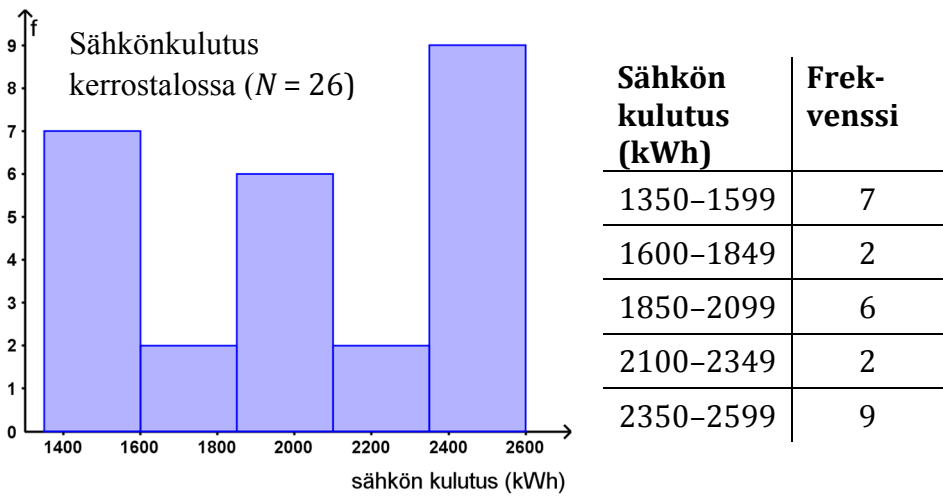
Etelä-Pohjanmaan jakauma on likimain symmetrinen ja havainnot ovat keskittyneet vaihteluvälin keskipaikoille. Mediaani (12,5 %) ja keskiarvo (12,0 %) ovat lähellä toisiaan.

Etelä-Pohjanmaalla kauran valkuaisainepitoisuus oli keskimäärin 12,0 % ja arvoissa esiintyvä keskihajonta oli 1,1 %-yksikköä. Mediaanin perusteella voidaan sanoa, että puolessa näytteistä valkuaispitoisuus oli alle 12,5 % ja puolessa yli 12,5 %.

136.

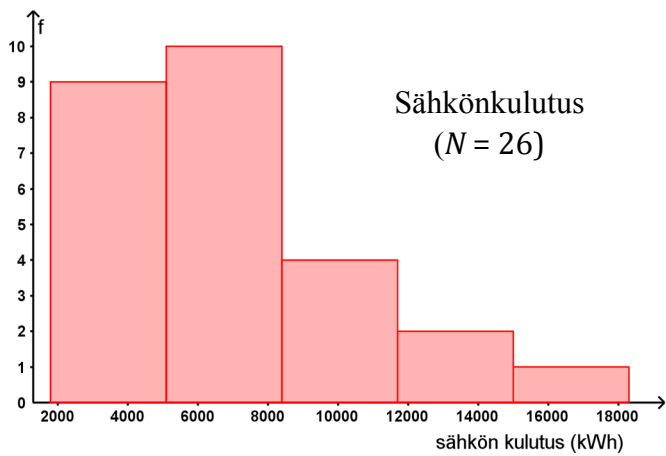
a) Määritetään aineistojen pienimmät arvot tilastotoiminnoilla.

Kysely 1: aineiston pienin arvo on min = 1350 (kWh). Asetetaan luokitus alkamaan minimiarvosta ja luokan leveydeksi 250 (kWh).



Histogrammin muoto riippuu aina valitusta luokittelusta. Jakaumassa ei ole havaittavissa selkeää huippua vaan arvot ovat melko tasaisesti jakautuneet koko vaihteluvälille 1350–2600 (kWh).

Kysely 2: aineiston pienin arvo on min = 1794 (kWh). Asetetaan luokitus alkamaan minimiarvosta ja luokan leveydeksi 3300 (kWh).



Sähkön kulutus (kWh)	Frek- venssi
1794–5093	9
5094–8393	10
8394–11 693	4
11 694–14 993	2
14 994–18 294	1

Jakauma on oikealle vino. Vaihteluvälin 1794–18 200 (kWh) pienet arvot ovat yleisimpiä kuin suuret arvot.

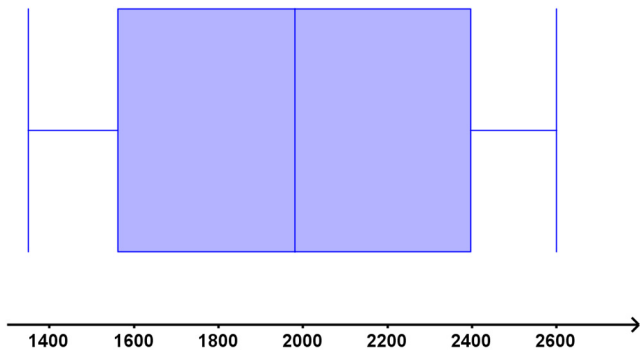
b) Määritetään tunnusluvut ohjelmiston tilastotoiminnoilla.

	Kysely 1 (kWh)	Kysely 2 (kWh)
keskiarvo:	$\bar{x} = 1986,73 \dots \approx 1987$	$\bar{x} = 7137,15 \dots \approx 7137$
keskihajonta:	$s = 422,96 \dots \approx 422$	$s = 3936,73 \dots \approx 3937$
mediaani:	$Md = 1981$	$Md = 6592,5 \approx 6593$

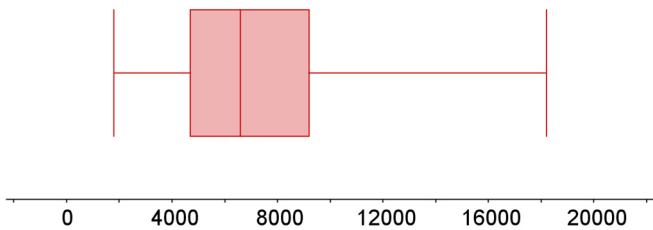
c) Kyselyssä 1 keskiarvo (1987 kWh) ja mediaani (1981) ovat likimain yhtä suuret. Jakauma on siis symmetrinen. Jakauman symmetrisyys voidaan todeta myös laatikkokuvioista, jossa jokainen neljännes on likimain yhtä leveä, ja mediaani on laatikko-osan keskellä.

Kyselyssä 2 keskiarvo (7137 kWh) on mediaania (6593 kWh) suurempi. Jakauma on oikealle vino, ja matala "häntä" eli pieni määrä suuria arvoja siirtää keskiarvoa mediaanista katsoen oikealle. Jakauman muoto voidaan todeta myös laatikkokuvioista, jossa on pitkä "viiksi" oikealle, ja mediaani on lähempänä laatikko-osan vasenta reunaa. (Lisäksi: suurin arvo poikkeaa merkitsevästi jakauman keskiarvosta.)

Kysely 1: laatikkokuvio (symmetrinen)



Kysely 2: laatikkokuvio (oikealle vino)



137.

a) Piirretään aikasarjasta viivadiagrammi.



b) Voittoajoissa oli selkeä lasku ensimmäisen noin 10 vuoden aikana eli aikavälillä 1972–1983. Tänä aikana voittoaika parani vuoden 1972 ajasta 190,4 min vuoden 1983 aikaan 142,7 min.

$$\frac{142,7}{190,4} = 0,7494 \dots \approx 75 \%$$

Voittoaika parani näiden vuosien aikana $100 \% - 75 \% = 25 \%$.

Tämän jälkeen voittoajoissa esiintynyt vaihtelu on ollut pientä. Vuosina 1986–2006 voittoajoissa esiintynyt vaihtelu oli vähäistä. Vaihteluväli näiden vuosien voittoajoissa oli 140,7 min – 147,2 min.

c) Viimeisen 10 vuoden aikana, eli vuosina 2008–2018, voittoajoissa on ollut enemmän vaihtelua edelliseen 10 vuoteen verrattuna: vaihteluväli näiden vuosien voittoajoissa on 140 min – 159,9 min. Vaihtelu on kuitenkin varsin pientä verrattuna ensimmäisen 10 vuoden aikana tapahtuneeseen laskuun.

d) Etsitään aineiston minimiarvo. Voittoaikojen minimi on $\min = 140,0$ (min).

Viivadiagrammi käy alimmillaan vuosina 2002 ja 2014. Vertaamalla näiden vuosien arvoja havaintoaineistossa nähdään, että voittoaika 140,0 on vuodelta 2014.

Deban ennätys, 140,0 min, on vuodelta 2014.

e) Määritetään aikavälin 2008–2018 voittoaikojen keskiarvo ja keskihajonta.

- keskiarvo on $\bar{x} = 147,336 \dots$ (min)
- keskihajonta on $\sigma = 5,426 \dots$ (min)

Lasketaan Deban voittoajan 140 min normittu arvo.

$$Z = \frac{140 - \bar{x}}{\sigma} = \frac{140 - 147,336 \dots}{5,426 \dots} = -1,351 \dots$$

Normitetun arvon itseisarvo on $|Z| = 1,35 \dots < 2$, joten Deban voittoaika ei poikkea merkitsevästi aikavälin 2008–2018 keskiarvosta.

138.

a) Miesten luokassa oli yhteensä 3509 juoksijaa. Ensimmäinen maaliin tullut juoksi matkan aikaan 132,1 min ja viimeisenä tullut aikaan 436,65 min. Ero voittajan ja viimeisenä maaliin tulleen välillä oli siis 304,55 min eli noin viisi tuntia.

Juoksuaikojen keskiarvo oli 260,5 min keskihajonnan ollessa 55,89 min. Puolet juoksijoista käytti matkaan korkeintaan 256,6 min ja puolet vähintään tämän ajan. Nopein neljäsnes suoritti matkan alle 219,2 minuuttiin. Hitain neljäsnes käytti matkaan vähintään 295,1 min.

b) Aloitetaan vertailu keskiarvoista. Miesten juoksuaikojen keskiarvo on $\bar{x} = 260,5$ min ja naisilla vastaava tunnusluku on $\bar{x} = 285,2$ min. Eroa keskiarvoissa on $285,2 \text{ min} - 260,5 \text{ min} = 24,7 \text{ min} \approx 25 \text{ min}$. Miehet juoksivat siis keskimäärin 25 min naisia nopeammin.

Miesten voittoaika, 132,1 min, on naistenvoittoaikaa, 152,2 min, parempi. Eroa voittoajoissa on noin $152,2 \text{ min} - 132,1 \text{ min} = 20,1 \text{ min} \approx 20 \text{ min}$. Sen sijaan ero viimeisenä maaliin tulleiden juoksuaajoissa on vain $437,7 \text{ min} - 436,7 \text{ min} = 1 \text{ min}$, jälleen miesten hyväksi.

Mediaaniaajoissa ero sukupuolten välillä on miesten eduksi $281,0 \text{ min} - 256,6 \text{ min} = 24,4 \text{ min}$. Ero on siis samaa luokkaa kuin ero keskiarvoissa.

Miesten juoksuajat vaihtelevat välillä 132,1 min – 436,7 min ja vaihteluvälin pituus on siis 304,7 min. Naisten aikojen vaihteluväli on 152,2 min – 437,7 min ja vaihteluvälin pituus 284,9 min, mikä on pienempi kuin miehillä. Miesten ajoissa hajonta on $\sigma = 55,8$ min, kun naisilla vastaava tunnusluku on pienempi, $\sigma = 52,7$ min. Miesten juoksuaajoissa esiintyi siis kaiken kaikkiaan enemmän hajontaa kuin naisilla.

Verrataan eroja hajonnassa tarkemmin, eri neljänneksissä:

	Nopein neljännes $Q_1 - \min$	Keskimmäinen 50 % $Q_3 - Q_2$	Hitain neljännes $\max - Q_3$
Miehet	$219,2 - 132,1 = 87,1$	$295,1 - 219,2 = 75,9$	$436,7 - 296,1 = 140,6$
Naiset	$245,0 - 152,2 = 92,8$	$319,2 - 245,0 = 74,2$	$437,7 - 319,2 = 118,5$

Taulukkoon laskettujen vaihteluvälien pituuksien perusteella voidaan todeta: molemmilla sukupuolilla suurin vaihtelu esiintyy hitaimmin juosseiden neljänneksessä ja kaiken kaikkiaan suurin vaihtelu esiintyy miesten ryhmässä. Lisäksi naisten nopeimmin juosseiden neljänneksessä vaihtelu on hieman suurempaa kuin miesten vastaavassa ryhmässä.

Yhteenvetona: Voidaan todeta, että miesten juoksu aika on keskimäärin noin 25 min naisten aikaa parempi. Voittoaajoissa ero on noin 20 min miesten hyväksi, mutta viimeisenä maaliin tulleiden ajoissa eroa sukupuolten välillä oli vain minuutin verran miesten hyväksi. Miesten juoksuaajoissa hajonta on suurempaa, ja suurin vaihtelu esiintyy hitaimmin juosseiden neljänneksessä.