

4 JOHDATUS TILASTOTIETEeseen

POHDITTAVAA

1. Etsitään nopeimmat elinajanodotteen muutokset Suomessa appletin avulla.

Elinajanodote aleni vuodesta 1802 huomattavasti vuoteen 1803. Vuonna 1802 elinajanodote oli noin 40 vuotta ja vuonna 1803 alle 30 vuotta.

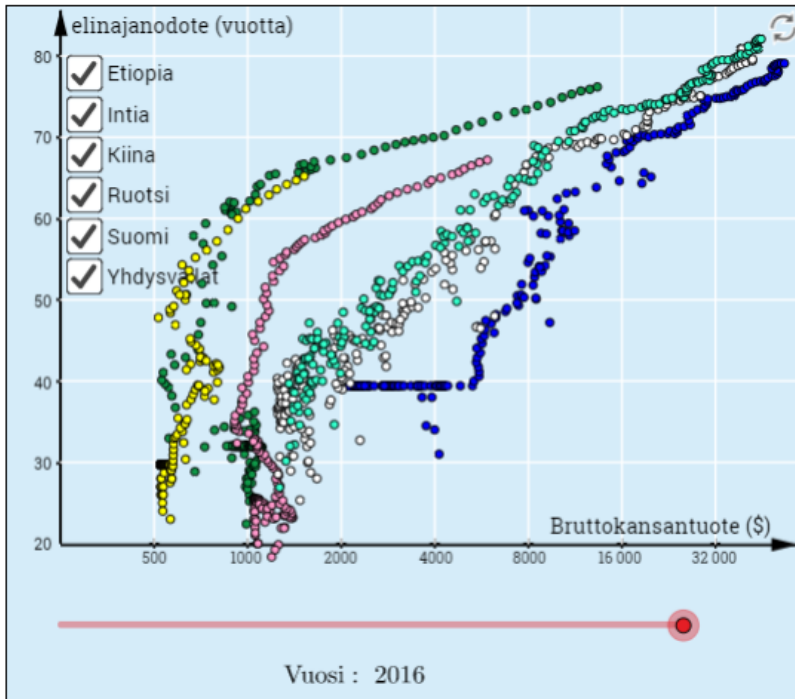
Elinajanodote aleni vuodesta 1864 vuoteen 1867 huomattavasti. Vuonna 1864 elinajanodote oli noin 40 vuotta ja vuonna 1867 noin 25 vuotta.

Sisällissotavuonna 1918 elinajanodote aleni edellisen vuoden noin 46 vuodesta noin 32 vuoteen.

Seuraava merkittävä aleneminen tapahtui sotavuonna 1940. Elinajanodote aleni noin 47 vuoteen. Sodan jälkeen, vuonna 1945, elinajanodote palasi sotaan edeltävälle tasolle eli noin 58 vuoteen.

Vastaus: vuosina 1803, 1864–1867, 1918 ja 1940

2. Appletista nähdään, että bruttokansantuotteen kasvaessa elinajanodote yleensä kasvaa.



Vastaus: Bruttokansantuotteen kasvaessa elinajanodote yleensä kasvaa.

4.1 Tilastojen esittäminen

ALOITA PERUSTEISTA

- 401.** a) Kotimaisia tietokirjoja kustannettiin taulukon mukaan 4637.

Vastaus: 4637 kirjaa

- b) Kotimaisia tietokirjoja kustannettiin 4637 ja kaunokirjallisuuden teoksia 1753, joten kotimaisia kirjoja kustannettiin yhteensä $4637 + 1753 = 6390$.

Vastaus: 6390 kirjaa

- c) Käännöskirjoja kustannettiin $1888 + 98 = 1986$. Kotimaisten kirjojen ja käännöskirjojen määrien ero oli $6390 - 1986 = 4404$.

Vastaus: 4404 kirjaa

- 402.** Luku 674 € on myynnin arvo eli havaintoarvo, joten A ja II kuuluvat yhteen.

Luku 9 on frekvenssi, joten B ja I kuuluvat yhteen.

Lukuväli 1500–1999 € on luokka, joten C ja IV kuuluvat yhteen.

Taulukon tilastomuuttuja on myynti, joten D ja III kuuluvat yhteen.

Vastaus: A: II, B: I, C: IV ja D: III

403. a) Kaavion mukaan 70–79-vuotiaita miehiä oli noin 200 000.

Vastaus: n. 200 000 miestä

b) Kaavion mukaan 60–69-vuotiaita naisia on eniten, joten naisten suurin ikäluokka oli 60–69-vuotiaat.

Vastaus: 60–69-vuotiaat

c) Luokkavälin pituus on peräkkäisten alarajojen erotus eli esimerkiksi $20 \text{ vuotta} - 10 \text{ vuotta} = 10 \text{ vuotta}$.

Vastaus: 10 vuotta

404. Täydennetään taulukkoon frekvenssit pylväiden korkeuksien avulla.

Arvosana	f	$f\%$
4	2	
5	4	
6	3	
7	9	
8	8	
9	6	
10	1	
Yhteensä		

Arvosanoja on yhteensä $2 + 4 + 3 + 9 + 8 + 6 + 1 = 33$.

Täydennetään taulukkoon suhteelliset frekvenssit jakamalla frekvenssit arvosanojen yhteismäärällä.

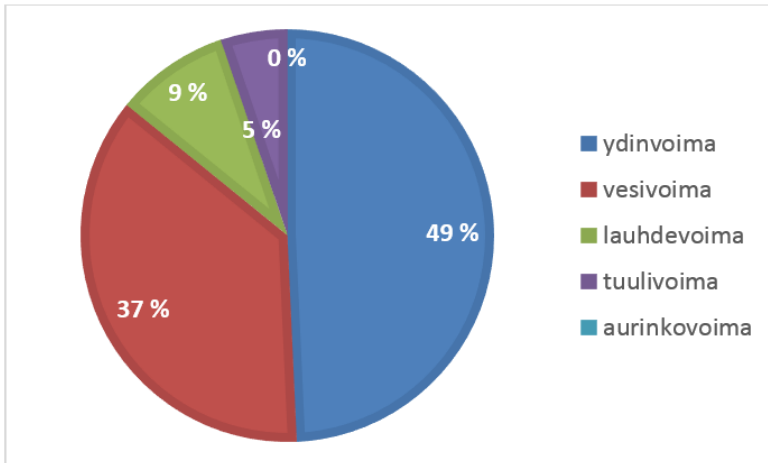
Arvosana	f	$f\%$
4	2	$\frac{2}{33} \approx 6,1\%$
5	4	$\frac{4}{33} \approx 12,1\%$
6	3	$\frac{3}{33} \approx 9,1\%$
7	9	$\frac{9}{33} \approx 27,3\%$
8	8	$\frac{8}{33} \approx 24,2\%$
9	6	$\frac{6}{33} \approx 18,2\%$
10	1	$\frac{1}{33} \approx 3,0\%$
Yhteensä	33	100,0 %

Vastaus:

Arvosana	f	$f\%$
4	2	6,1 %
5	4	12,1 %
6	3	9,1 %
7	9	27,3 %
8	8	24,2 %
9	6	18,2 %
10	1	3,0 %
Yhteensä	33	100,0 %

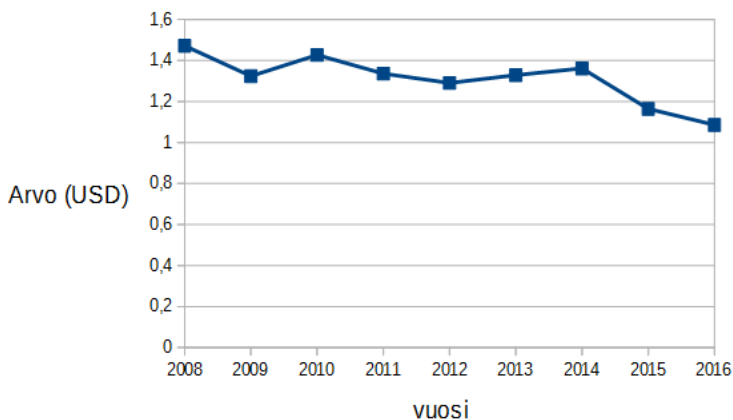
- 405.** Kopioidaan taulukon tiedot taulukkolaskentaohjelmaan ja piirretään ympyräkaavio.

Vastaus:



- 406.** Kopioidaan taulukon tiedot taulukkolaskentaohjelmaan ja piirretään viivakaavio.

Vastaus:



Euron arvo on pienentynyt dollariin verrattuna.

407. a) Luokat on muodostettu 10 metrin tarkkuudella, joten pyöristyssääntöjen mukaan ensimmäisen luokan todellinen alaraja on 1295, koska se on pienin luku, joka pyöristyy luvuksi 1300. Todellinen yläraja on 1795, koska se on pienin luku, joka pyöristyy luvuksi 1800. Täydennetään muiden luokkien todelliset rajat vastaavasti ja luokitellaan tulokset taulukon luokkiin.

Tulos (m)	Todellinen alaraja	Todellinen yläraja	f
1300–1790	1295	1795	2
1800–2290	1795	2295	7
2300–2790	2295	2795	10
2800–3290	2795	3295	5

Suurin frekvenssi on luokalla 2300–2790 m.

Vastaus:

Tulos (m)	Todellinen alaraja	Todellinen yläraja	f
1300–1790	1295	1795	2
1800–2290	1795	2295	7
2300–2790	2295	2795	10
2800–3290	2795	3295	5

2300–2790 m

- b) Luokkavälin pituus on peräkkäisten alarajojen erotus
 $1800 \text{ m} - 1300 \text{ m} = 500 \text{ m}$.

Vastaus: 500 m

- c) Luokitellun aineiston etuna on, että se on selkeämpi ja sen avulla saa helpommin käsityksen minkälaisesta aineistosta on kyse. Haittana on, että luokitellussa aineistossa ei ole niin tarkkaa tietoa kuin alkuperäisessä.

Vastaus: Etuna on selkeys ja kokonaiskuvan hahmottaminen, haittana yksittäisten alkuperäisten arvojen katoaminen.

- 408.** a) Täydennetään taulukkoon summafrekvenssit sf , suhteelliset frekvenssit $f\%$ ja suhteelliset summafrekvenssit $sf\%$.

Hinta (€)	f	sf	$f\%$	$sf\%$
100 000–119 999	1	1	$\frac{1}{8} = 12,5 \%$	$\frac{1}{8} = 12,5 \%$
120 000–139 999	0	$1 + 0 = 1$	$\frac{0}{8} = 0 \%$	$\frac{1}{8} = 12,5 \%$
140 000–159 999	5	$1 + 5 = 6$	$\frac{5}{8} = 62,5 \%$	$\frac{6}{8} = 75 \%$
160 000–179 999	1	$6 + 1 = 7$	$\frac{1}{8} = 12,5 \%$	$\frac{7}{8} = 87,5 \%$
180 000–189 999	1	$7 + 1 = 8$	$\frac{1}{8} = 12,5 \%$	$\frac{8}{8} = 100 \%$

Vastaus:

Hinta (€)	f	sf	$f\%$	$sf\%$
100 000–119 999	1	1	12,5 %	12,5 %
120 000–139 999	0	1	0 %	12,5 %
140 000–159 999	5	6	62,5 %	75 %
160 000–179 999	1	7	12,5 %	87,5 %
180 000–189 999	1	8	12,5 %	100 %

- b) Taulukon perusteella hinta oli alle 160 000 euroa 75 prosentissa yksioista.

Vastaus: 75 %

- c) Taulukon perusteella hinta oli vähintään 140 000 mutta alle 180 000 euroa $62,5 + 12,5 = 75$ prosentissa yksioista.

Vastaus: 75 %

- 409.** a) Kaavion A pystyakselin arvot alkavat nolasta ja yksi ruutu vastaa 500 000:a liittymää. Kaavion B pystyakselin arvot alkavat 2 400 000:sta ja yksi ruutu vastaa 50 000:a liittymää.
- b) Jotta liittymien määrän nousu näyttäisi suhteellisesti suuremmalta kuin se todellisuudessa on tai jotta voidaan nähdä tarkemmin, kuinka monta tuhatta liittymää vuosittain on tullut lisää.

VAHVISTA OSAAMISTA

- 410.** Aikasarjan havainnollistamiseen sopii viivakaavio, joten A ja II kuuluvat yhteen. Aikasarjaa voidaan havainnollistaa myös pylväskaaviolla, joten myös kaavio I sopii aineiston A havainnollistamiseen.

Suhteellisten osuuksien havainnollistamiseen sopii ympyräkaavio, joten B ja III kuuluvat yhteen. Suhteellisia osuuksia voidaan havainnollistaa myös pylväskaaviolla, joten myös kaavio I sopii aineiston B havainnollistamiseen.

Lukumäärien havainnollistamiseen sopii pylväskaavio, joten C ja I kuuluvat yhteen.

Vastaus: A: I ja II, B: I ja III, C: I

- 411.** a) Frekvenssi on tietyn tilastomuuttujan arvon lukumäärä. Jos esimerkiksi 6 opiskelijaa sai kurssista arvosanan 8, arvosanan 8 frekvenssi on 6.
- b) Summafrekvenssi on niiden tilastomuuttujan arvojen lukumäärä, jotka ovat korkeintaan tietyn arvon suuruisia. Esimerkiksi, jos 13 opiskelijaa on korkeintaan 180 cm pitkiä, niin pituuden 180 cm summafrekvenssi on 13.

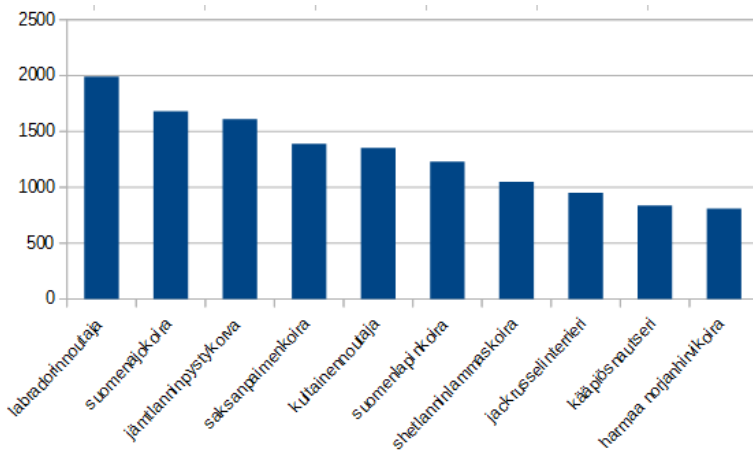
- c) Jakauma on kaikkien muuttujien arvojen frekvenssien tai suhteellisten frekvenssien kokoelma. Esimerkiksi opiskelijoiden massojen jakauma kilogrammoina voi olla seuraava:

Massa (kg)	f
54	2
60	1
62	1
71	3
89	1

- d) Luokka on väli, jolla osa tilastomuuttujan arvoista on. Esimerkiksi luokkaan 170–174 cm kuuluvat ne opiskelijat, joiden pituus on 170 cm, 171 cm, 172 cm, 173 cm tai 174 cm.
- e) Luokkavälin pituus on peräkkäisten luokkien alarajojen erotus. Esimerkiksi jos opiskelijoiden sähköpostilaatikossa olevien viestien lukumäärät on luokiteltu luokkiin 0–49, 50–99, 100–149 ja niin edelleen, niin luokkavälin pituus on $50 - 0 = 50$.
- f) Kertymä on niiden havaintoarvojen lukumäärä tai suhteellinen osuus, jotka ovat korkeintaan tietyn arvon suuruisia. Kertymällä tarkoitetaan siis joko summafrekvenssien tai suhteellisten summafrekvenssien kokoelmaa. Esimerkiksi, jos 17 opiskelijalta kysytään, kuinka monta kertaa he olivat ulkomailla viime kalenterivuoden aikana, kertymä voi seuraava.

Ulkomaanmatkojen määrä viime kalenterivuoden aikana	sf
0–1	8
2–3	14
4–5	16
6–7	17

412. a) Kopioidaan taulukon tiedot taulukkolaskentaohjelmaan ja piirretään pylväskaavio.



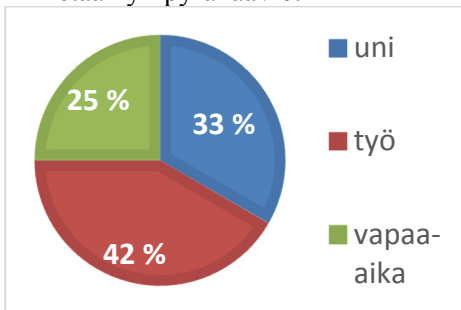
- b) Jos jakaumaa havainnollistettaisiin ympyräkaaviolla, voisi saada käsityksen, että kaaviossa on kaikki Suomen koirarodut, eli muita rotuja näiden 10 rodun lisäksi ei ole.

Vastaus: Voi syntyä kuva, että kaaviossa on kaikki Suomen koirarodut.

413. Taulukoidaan ajankäyttötavat ja lasketaan kutakin tapaa vastaava keskuskulma ja prosenttiosuus.

Ajankäyttötapa	Aika (h)	Keskuskulma	Suhteellinen osuus (%)
uni	8	$\frac{8}{24} \cdot 360^\circ = \frac{1}{3} \cdot 360^\circ = 120^\circ$	$\frac{8}{24} = \frac{1}{3} \approx 33\%$
työ	10	$\frac{10}{24} \cdot 360^\circ = \frac{5}{12} \cdot 360^\circ = 150^\circ$	$\frac{10}{24} \approx 42\%$
vapaa-aika	6	$\frac{6}{24} \cdot 360^\circ = \frac{1}{4} \cdot 360^\circ = 90^\circ$	$\frac{6}{24} = \frac{1}{4} = 25\%$

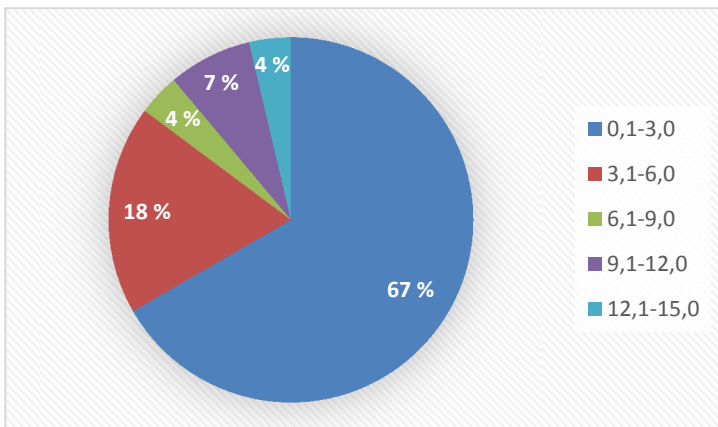
Piirretään ympyräkaavio.



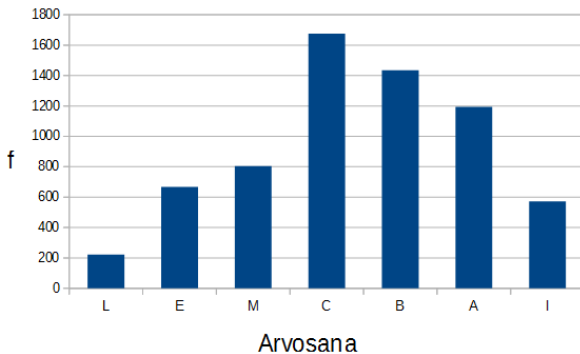
414. Luokitellaan koulumatkojen pituudet taulukkoon.

Koulumatkan pituus (km)	f
0,1–3,0	18
3,1–6,0	5
6,1–9,0	1
9,1–12,0	2
12,1–15,0	1

Havainnollistetaan jakaumaa ympyräkaaviolla.



- 415. a)** Kopioidaan taulukon luvut taulukkolaskentaohjelmaan ja piirretään pylväskaavio. Kaaviosta huomataan, että keskimmäistä arvosanaa on eniten ja korkeinta ja alinta arvosanaa vähiten.



- b)** Terveystiedon kirjoittajia oli yhteensä
 $223 + 668 + 805 + 1676 + 1435 + 1194 + 573 = 6574$.

Heistä arvosanan E tai L sai $223 + 668 = 891$. Tämä on
 $\frac{891}{6574} = 0,1355... \approx 13,6\%$ terveystiedon kirjoittajista.

Vastaus: 13,6 %

- c)** Hyväksytyistä suorituksista korkeintaan arvosanan C sai
 $1676 + 1435 + 1194 = 4305$ opiskelijaa. Tämä on
 $\frac{4305}{6001} = 0,7173... \approx 71,7\%$ terveystiedon hyväksytysti kirjoittaneista.

Vastaus: 71,7 %

- 416.** Työntekijöiden määrät ja sektorien keskuskulmat ovat suoraan verrannollisia. Taulukoidaan tunnetut tiedot ja merkitään tuntemattomia lukumääriä kirjaimilla.

Työntekijöiden paikka	Työntekijöiden määrä	Keskuskulma (°)
töissä	x	260
koulutuksessa	8	40
sairaana	y	60

Muodostetaan verranto ja ratkaistaan siitä töissä olevien työntekijöiden määrä.

$$\begin{aligned}\frac{x}{8} &= \frac{260}{40} \quad || \cdot 8 \\ x &= \frac{260}{40} \cdot 8 \\ x &= 52\end{aligned}$$

Töissä oli siis 52 työntekijää.

Muodostetaan verranto ja ratkaistaan siitä sairaana olevien työntekijöiden määrä.

$$\begin{aligned}\frac{y}{8} &= \frac{60}{40} \quad || \cdot 8 \\ y &= \frac{60}{40} \cdot 8 \\ y &= 12\end{aligned}$$

Sairaana oli siis 12 työntekijää.

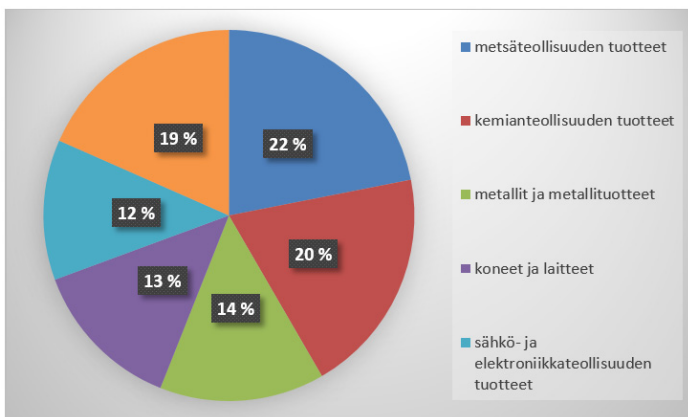
Vastaus: töissä 52, sairaana 12

417. Täydennetään taulukkoon alojen suhteelliset osuudet.

	A	B	C
1	metsäteollisuuden tuotteet	11344	=B1/\$B\$7
2	kemianteollisuuden tuotteet	10217	=B2/\$B\$7
3	metallit ja metallituotteet	7448	=B3/\$B\$7
4	koneet ja laitteet	6917	=B4/\$B\$7
5	sähkö- ja elektroniikkateollisuuden tuotteet	6332	=B5/\$B\$7
6	muut	9522	=B6/\$B\$7
7	yhteensä	=SUMMA(B1:B6)	

	A	B	C
1	metsäteollisuuden tuotteet	11344	22 %
2	kemianteollisuuden tuotteet	10217	20 %
3	metallit ja metallituotteet	7448	14 %
4	koneet ja laitteet	6917	13 %
5	sähkö- ja elektroniikkateollisuuden tuotteet	6332	12 %
6	muut	9522	18 %
7	yhteensä	51780	
8			

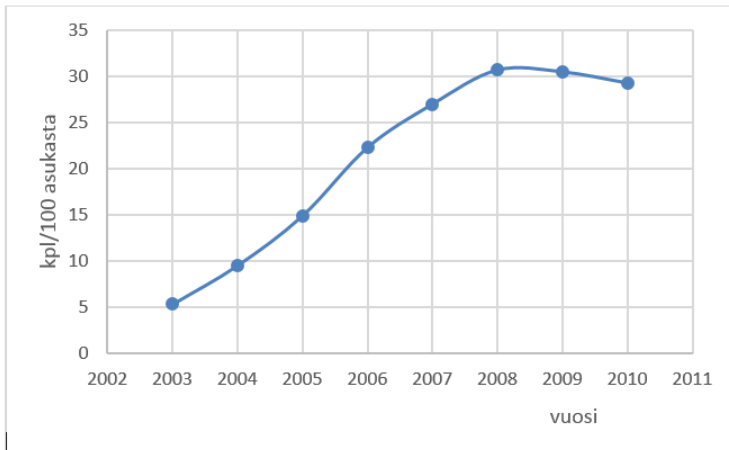
Piirretään
ympyräkaavio.



Vastaus:

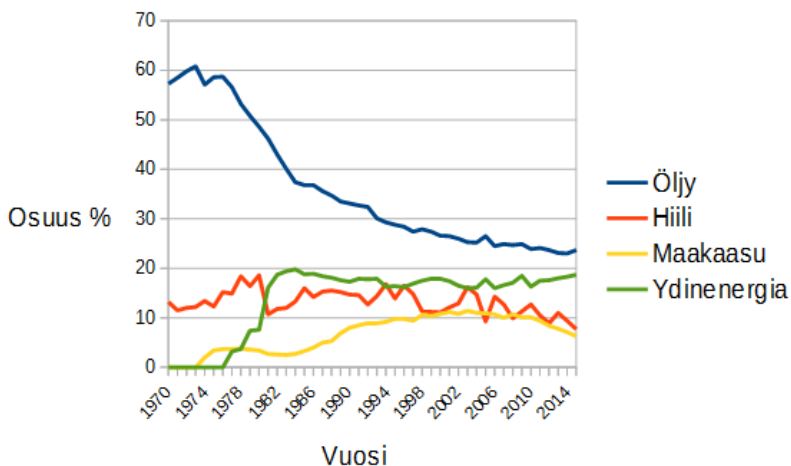
Vienti	Viennin arvo (miljoonaa euroa)	f%
metsäteollisuuden tuotteet	11 344	22 %
kemianteollisuuden tuotteet	10 217	20 %
metallit ja metallituotteet	7 448	14 %
koneet ja laitteet	6 917	13 %
sähkö- ja elektroniikkateollisuuden tuotteet	6 322	12 %
muut	9 522	18 %

418. a) Kopioidaan taulukon arvot taulukkolaskentaohjelmaan ja piirretään viivakaavio.



- b) Nopeiden internetyhteyksien määrää lieenee vuosina 2008-2010 vähentänyt älypuhelinien yleistyminen. Älypuhelinien internetyhteydet olivat hitaampia kuin esimerkiksi PC-koneiden yhteydet, joten niitä ei luokiteltu nopeiksi yhteyksiksi.

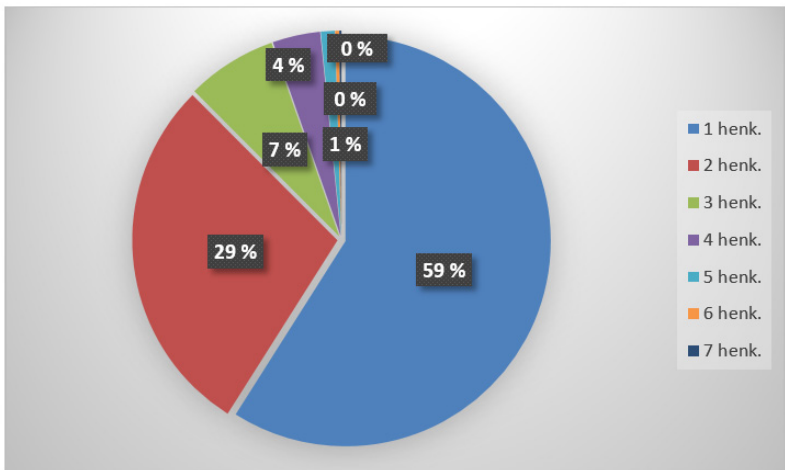
419. a) Kopioidaan taulukon tiedot sopivaan ohjelmaa ja piirretään viivakaavio.



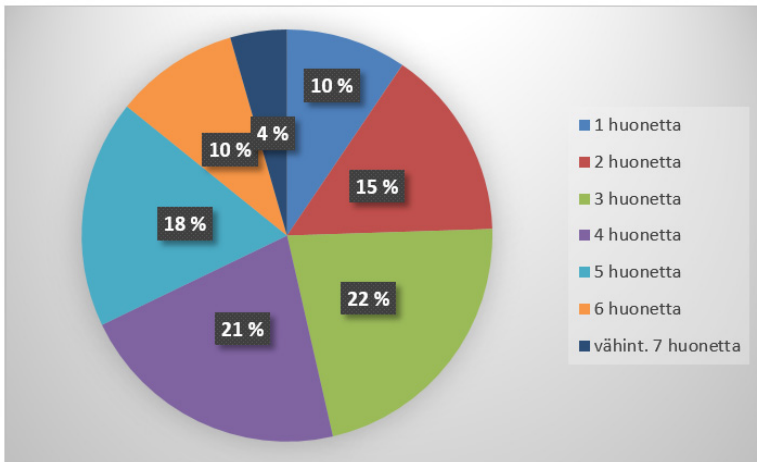
- b) Öljyllä tuotetun energian määrä on muuttunut eniten. Öljyn käytön suhteellinen osuus alkoi vähentyä 1970-luvun öljykriisin jälkeen ja öljyllä tuotettua energiaa alettiin korvata ydinvoimalla. Öljyn käytön suhteellinen osuus on vähentynyt myös siksi, että öljyvarat ovat vähentyneet ja on pyritty eroon fossiilisista polttoaineista.

Vastaus: Öljyllä tuotetun energian määrä on muuttunut eniten.

420. a) Kopioidaan taulukon luvut taulukkolaskentaohjelmaan ja piirretään ympyräkaavio.



- b) Kopioidaan taulukon luvut taulukkolaskentaohjelmaan ja piirretään ympyräkaavio.



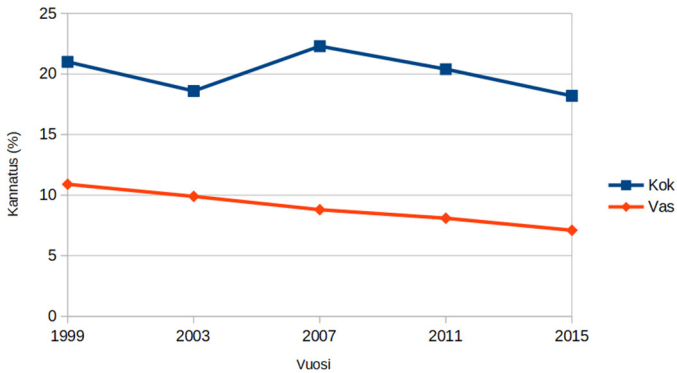
- c) Huomataan, että kerrostalojen asutokunnista suurin osa on 1–2 hengen asutokuntia, kun taas asuntojen huonelukumäärät ovat jakautuneet melko tasaisesti erikokoisiin asutuihin. Jakaumien ero johtunee siitä, että asutokuntien koot on tilastoitu vain kerrostaloista ja huonelukutilastossa on kaiken tyyppiset asunnot. Omakoti- ja rivitaloasuntojen huonelukumäärä on suurempi kuin kerrostaloasunnoissa, vaikka asutokunnan koko olisi 1–2 henkeä.

Vastaus: Kerrostalojen asutokunnista suurin osa on 1–2 hengen asutokuntia, kun taas asuntojen huonelukumäärät ovat jakautuneet melko tasaisesti erikokoisiin asutuihin.

421. a) Kaavion perusteella syntyy vaikutelma, että Vasemmistoliiton kannatus olisi vuonna 2015 ollut noin viidesosa Kansallisen Kokoomuksen kannatuksesta, vaikka kolmasosa on lähempänä totuutta. Tämä johtuu siitä, että pystyakseli ei ala nollasta.

Vastaus: Pystyakseli ei ala nollasta.

- b) Piirretään uusi kaavio katkaisematta pystyakselia arvioimalla kuvaajasta puolueiden kannatukset kunakin vuonna.



422. Täydennetään taulukkoon summafrekvenssit ja suhteelliset frekvenssit.

Lämpötila (°C)	f	sf	$f\%$
-21,0...-18,1	12	12	$\frac{12}{56} = 0,2142... \approx 21,4\%$
-18,0...-15,1	15	$12 + 15 = 27$	$\frac{15}{56} = 0,2678... \approx 26,8\%$
-15,0...-12,1	12	$27 + 12 = 39$	$\frac{12}{56} = 0,2142... \approx 21,4\%$
-12,0...-9,1	12	$39 + 12 = 51$	$\frac{12}{56} = 0,2142... \approx 21,4\%$
-9,0...-6,1	5	$51 + 5 = 56$	$\frac{5}{56} = 0,0892... \approx 8,9\%$

Täydennetään taulukkoon suhteelliset summafrekvenssit.

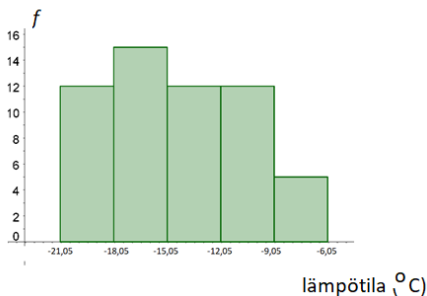
Lämpötila (°C)	f	sf	$f\%$	$sf\%$
-21,0...-18,1	12	12	21,4 %	$\frac{12}{56} = 0,2142... \approx 21,4\%$
-18,0...-15,1	15	27	26,8 %	$\frac{27}{56} = 0,4821... \approx 48,2\%$
-15,0...-12,1	12	39	21,4 %	$\frac{39}{56} = 0,6964... \approx 69,6\%$
-12,0...-9,1	12	51	21,4 %	$\frac{51}{56} = 0,9107... \approx 91,1\%$
-9,0...-6,1	5	56	8,9 %	100 %

Täydennetään taulukkoon todelliset rajat.

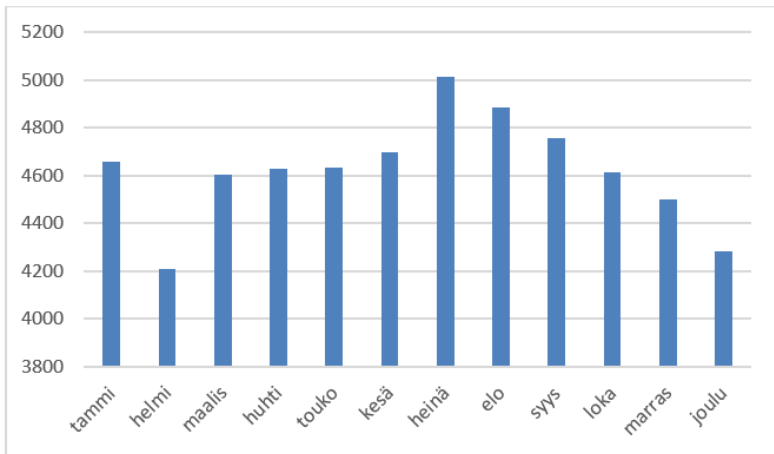
Lämpötilat on annettu yhden desimaalin tarkkuudella, joten ensimmäisen luokan alaraja on $-21,05$, koska se on suurin negatiivinen luku, joka pyöristyy luvuksi $-21,1$. Ensimmäisen luokan todellinen yläraja on $-18,05$, koska se on suurin negatiivinen luku, joka pyöristyy luvuksi $-18,1$.

Lämpötila ($^{\circ}\text{C}$)	Todellinen alaraja	Todellinen yläraja	f	sf	$f\%$	$sf\%$
$-21,0 \dots -18,1$	$-21,05$	$-18,05$	12	12	21,4 %	21,4 %
$-18,0 \dots -15,1$	$-18,05$	$-15,05$	15	27	26,8 %	48,2 %
$-15,0 \dots -12,1$	$-15,05$	$-12,05$	12	39	21,4 %	69,6 %
$-12,0 \dots -9,1$	$-12,05$	$-9,05$	12	51	21,4 %	91,1 %
$-9,0 \dots -6,1$	$-9,05$	$-6,05$	5	56	8,9 %	100 %

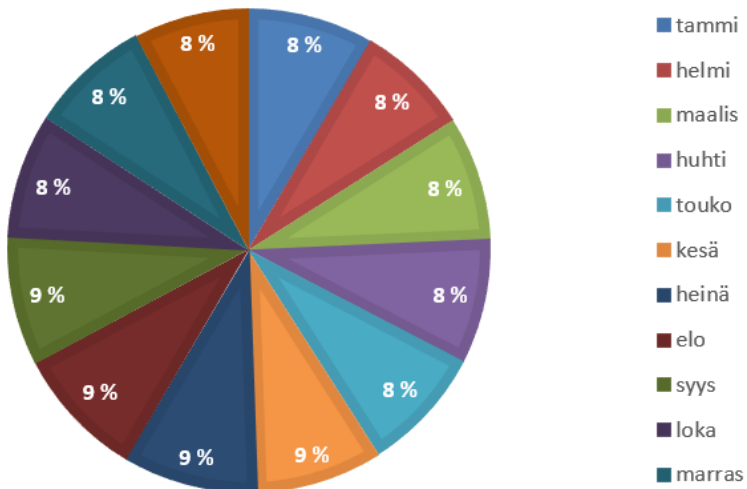
Havainnollistetaan aineistoa histogrammilla..



423. a) Kopioidaan taulukon tiedot taulukkolaskentaohjelmaan ja piirretään pylväskaavio.



Piirretään samasta aineistosta ympyräkaavio.



- b)** Pylväskaavion avulla on helpompi asettaa kuukaudet suuruusjärjestykseen kuin ympyräkaavion avulla, koska pylväiden korkeuksia on helpompi verrata kuin sektoreiden keskuskulmia.

Vastaus: pylväskaavio

- c)** Pylväskaavio antaa väärän kuvan suuruusluokista: helmikuussa syntyneiden määrä näyttää totuuden vastaisesti olevan alle puolet heinäkuussa syntyneiden määrästä. Tämä johtuu siitä, että pylväät eivät ala nollan korkeudelta. Ympyräkaaviosta nähdään, että kuukausittaiset määrät ovat suurin piirtein samansuuruisia.

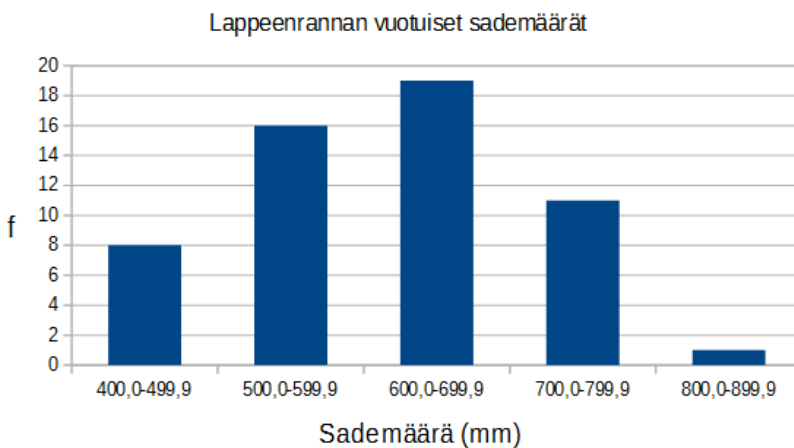
Vastaus: ympyräkaavio

- d)** Syntymien määrät vaihtelevat kuukausittain, joten kaikki kuukaudet eivät ole yhtä todennäköisiä.

Vastaus: eivät ole

424. a) Pienin vuosittainen sademäärä on 402,8 mm ja suurin 863,4 mm. Jos luokista tehdään tasavälisiä, on luokkavälin oltava millimetreinä vähintään $\frac{863,4 - 402,8}{5} = 92,12$. Yksinkertaisuuden vuoksi valitaan luokkaväliksi 100 mm. Kopioidaan luvut taulukkolaskentaohjelmaan, luokitellaan taulukon arvot ja piirretään pylväskaavio.

Sademäärä (mm)	f
400,0–499,9	8
500,0–599,9	16
600,0–699,9	19
700,0–799,9	11
800,0–899,9	1



- b) Keskimmaisessä luokassa 600,0–699,9 on eniten havaintoarvoja ja suurimman sademäärän luokassa 800,0–899,9 vähiten.

- 425. a)** Todellisia rajoja määritettäessä on otettava huomioon, että ikä pyöristyy aina alaspäin. Esimerkiksi ensimmäiseen ikäluokkaan 18–27 vuotta kuuluvat kaikki henkilöt, jotka ovat täyttäneet 18 vuotta, mutta eivät 28 vuotta.

Ikä (vuotta)	Todellinen alaraja (vuotta)	Todellinen yläraja (vuotta)
18–27	18	28
28–37	28	38
38–47	38	48
48–57	48	58
58–67	58	68

- b)** Kopioidaan luvut taulukkolaskentaohjelmaan ja lasketaan kaksi ensimmäistä summafrekvenssiä.

	A	B	C
1	Ikä (vuotta) f		sf
2	18-27	5	5
3	28-37	18	=C2+B3
4	38-47	14	
5	48-57	15	
6	58-67	9	

Kopioidaan summafrekvenssin lauseke alemmille riveille.

	A	B	C
1	Ikä (vuotta) f		sf
2	18-27	5	5
3	28-37	18	23
4	38-47	14	
5	48-57	15	
6	58-67	9	

	A	B	C
1	Ikä (vuotta) f		sf
2	18-27	5	5
3	28-37	18	23
4	38-47	14	37
5	48-57	15	52
6	58-67	9	61

Määritetään suhteelliset frekvenssit vastaavalla tavalla.

	A	B	C	D
1	Ikä (vuotta) f		sf	f %
2	18-27	5	5	=B2/61*100
3	28-37	18	23	
4	38-47	14	37	
5	48-57	15	52	
6	58-67	9	61	

	A	B	C	D
1	Ikä (vuotta) f		sf	f %
2	18-27	5	5	8.2
3	28-37	18	23	29.5
4	38-47	14	37	23
5	48-57	15	52	24.6
6	58-67	9	61	14.8

Sarakkeen D soluviittaukset viittaavat sarakkeeseen B. Suhteelliset summafrekvenssit saadaan suoraan kopioimalla sarakkeen D soluviittaukset sarakkeeseen E, koska tällöin uudet viittaukset viittaavat sarakkeeseen C.

	A	B	C	D	E
1	Ikä (vuotta)	f	sf	f %	sf %
2	18-27	5	5	8.2	8.2
3	28-37	18	23	29.5	37.7
4	38-47	14	37	23	60.7
5	48-57	15	52	24.6	85.2
6	58-67	9	61	14.8	100

Vastaus:

Ikä (vuotta)	f %	sf	sf %
18-27	8,2	5	8,2
28-37	29,5	23	37,7
38-47	23,0	37	60,7
48-57	24,6	52	85,2
58-67	14,8	61	100

- c) Vastaus saadaan laskemalla yhteen 28-37- ja 38-47- sekä 48-57- vuotiaiden määrät ja jakamalla kokonaismäärällä 61.

$$\frac{18+14+15}{61} = 0,7704... \approx 77,0 \%$$

Työntekijöistä 77,0 % on 28-57-vuotiaita.

Vastaus: 77,0 %

- d) Vertaamalla yrityksen ikäjakaumaa valtakunnalliseen työssä käyvien ikäjakaumaan huomataan, että yrityksessä on suhteellisesti enemmän nuoria kuin valtakunnallisesti, joten ei voida sanoa, että yrityksen ikäjakauma noudattaisi valtakunnallista ikäjakaumaa.

Vastaus: ei voida

SYVENNÄ YMMÄRRYSTÄ

426. Taulukon suurin summafrekvenssi on 50, joten havaintoaineiston koko on 50. Tällöin yksi henkilö on $\frac{1}{50} = 0,02$ eli 2 % koko aineistosta.

Täydennetään tällä perusteella frekvenssin ja suhteellisen frekvenssin sarakkeita.

Kengännumero	f	$f\%$	sf	$sf\%$
36	3	$2 \cdot 3 = 6$		
37	9	$2 \cdot 9 = 18$		
38	8	$2 \cdot 8 = 16$		
39				
40	$14 : 2 = 7$	14		
41	$6 : 2 = 3$	6		
42	$8 : 2 = 4$	8	45	
43			47	
44			50	

Täydennetään frekvenssisaraketta summafrekvenssien perusteella.

Kengännumero	f	$f\%$	sf	$sf\%$
36	3	6		
37	9	18		
38	8	16		
39				
40	7	14		
41	3	6		
42	4	8	45	
43	$47 - 45 = 2$		47	
44	$50 - 47 = 3$		50	

Kaikkien frekvenssien summan on oltava 50. Muita kokoja kuin kokoa 39 on yhteensä $3 + 9 + 8 + 7 + 3 + 4 + 2 + 3 = 39$ kpl, joten koon 39 frekvenssi on $50 - 39 = 11$. Täydennetään tämä frekvenssi ja loput suhteelliset frekvenssit taulukkoon.

Kengännumero	f	$f\%$	sf	$sf\%$
36	3	6		
37	9	18		
38	8	16		
39	11	$2 \cdot 11 = 22$		
40	7	14		
41	3	6		
42	4	8	45	
43	2	$2 \cdot 2 = 4$	47	
44	3	$2 \cdot 3 = 6$	50	

Täydennetään taulukkoon summafrekvenssit ja suhteelliset summafrekvenssit.

Kengännumero	f	$f\%$	sf	$sf\%$
36	3	6	3	6
37	9	18	$3 + 9 = 12$	$12 \cdot 2 = 24$
38	8	16	$12 + 8 = 20$	$20 \cdot 2 = 40$
39	11	22	$20 + 11 = 31$	$31 \cdot 2 = 62$
40	7	14	$31 + 7 = 38$	$38 \cdot 2 = 76$
41	3	6	$38 + 3 = 41$	$41 \cdot 2 = 82$
42	4	8	45	$45 \cdot 2 = 90$
43	2	4	47	$47 \cdot 2 = 94$
44	3	6	50	$50 \cdot 2 = 100$

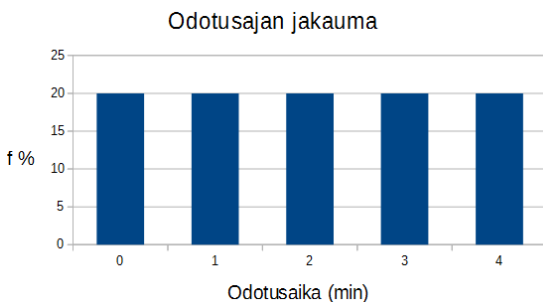
Vastaus:

Kengännumero	f	$f\%$	sf	$sf\%$
36	3	6	3	6
37	9	18	12	24
38	8	16	20	40
39	11	22	31	62
40	7	14	38	76
41	3	6	41	82
42	4	8	45	90
43	2	4	47	94
44	3	6	50	100

- 427.** Eri odotusaikoja on 5 kpl, joten jokaisen suhteellinen frekvenssi on 20 %. Jakauma on taulukon mukainen.

Odotusaika (min)	$f\%$
0	20 %
1	20 %
2	20 %
3	20 %
4	20 %

Piirretään jakaumaa kuvaava pylväskaavio.



428. a) Taulukoidaan kuvaajan avulla suhteelliset summafrekvenssit.

Massa (g)	$sf\%$
–485	0
486–490	2
491–495	12
496–500	35
501–505	59
506–510	88
511–515	98
516–520	100

Lisätään taulukkoon sarake suhteelliselle frekvenssille ja lasketaan suhteelliset frekvenssit yksi kerrallaan.

Massa (g)	$f\%$	$sf\%$
–485	0	0
486–490	$2 - 0 = 2$	2
491–495	$12 - 2 = 10$	12
496–500	$35 - 12 = 23$	35
501–505	$59 - 35 = 24$	59
506–510	$88 - 59 = 29$	88
511–515	$98 - 88 = 10$	98
516–520	$100 - 98 = 2$	100

Vastaus:

Massa (g)	$f\%$
–485	0
486–490	2
491–495	10
496–500	23
501–505	24
506–510	29
511–515	10
516–520	2

b) Korkeintaan 490 grammaa painavia juustoja on 2 % kaikista juustoista. 2000 kappaleen juustoerästä 2 % on $2000 \cdot 0,02 = 40$, joten B-luokan juustoja on odotettavasti 40.

Vastaus: 40 juustoa

429. Kopioidaan taulukon luvut taulukkolaskentaohjelmaan ja lasketaan yrityksen A vuoden 2011 kasvuprosentti soluviittausten avulla.

	A	B	C	D
1	vuosi	2010	2011	2012
2	liikevaihto A (milj. €)	52,1	54,1	55,3
3	kasvuprosentti A		=(C2-B2)/B2*100	

Kopioidaan kaava seuraaville vuosille.

	A	B	C	D	E	F	G	H
1	vuosi	2010	2011	2012	2013	2014	2015	2016
2	liikevaihto A (milj. €)	52,1	54,1	55,3	55,9	55,8	57,1	58,7
3	kasvuprosentti A		3,838771593					

	A	B	C	D	E	F	G	H
1	vuosi	2010	2011	2012	2013	2014	2015	2016
2	liikevaihto A (milj. €)	52,1	54,1	55,3	55,9	55,8	57,1	58,7
3	kasvuprosentti A		3,838771593	2,218114603	1,084990958	-0,178890877	2,329749104	2,802101576

Muotoillaan solut niin, että tulokset näkyvät 2 desimaalin tarkkuudella.

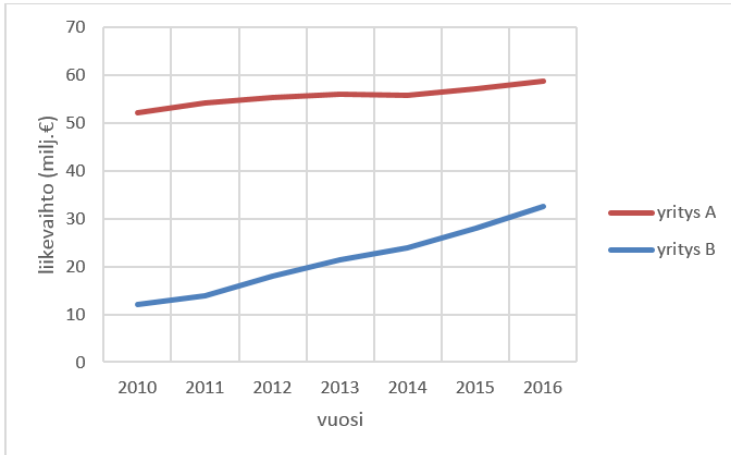
	A	B	C	D	E	F	G	H
1	vuosi	2010	2011	2012	2013	2014	2015	2016
2	liikevaihto A (milj. €)	52,1	54,1	55,3	55,9	55,8	57,1	58,7
3	kasvuprosentti A		3,84	2,22	1,08	-0,18	2,33	2,80

Tehdään sama yritykselle B. Nyt riittää kopioida vuotuiset liikevaihdot, sen jälkeen voidaan kopioida kasvuprosentin laskukaavat soluviittausten avulla yrityksen A kaavoista.

	A	B	C	D	E	F	G	H
1	vuosi	2010	2011	2012	2013	2014	2015	2016
2	liikevaihto A (milj. €)	52,1	54,1	55,3	55,9	55,8	57,1	58,7
3	kasvuprosentti A		3,84	2,22	1,08	-0,18	2,33	2,80
4	liikevaihto B (milj. €)	12	13,9	17,9	21,3	24	27,9	32,6
5	kasvuprosentti B		15,83	28,78	18,99	12,68	16,25	16,85

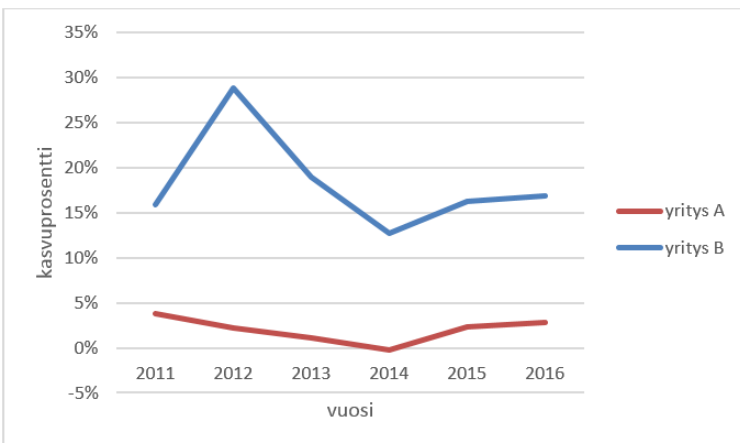
- a) Yrityksen A liikevaihto on koko ajan ollut suurempaa kuin yrityksen B, joten yrityksen A kannalta edullista on vertailla liikevaihtoja. Piirretään viivakaavio liikevaihdoista.

Vastaus:



- b) Yrityksen B liikevaihdon kasvuprosentti on koko ajan ollut suurempi kuin yrityksen A. Yrityksen B kannalta edullista on vertailla kasvuprosentteja. Piirretään niistä viivakaavio.

Vastaus:



- 430. a)** Koska yhdenmuotoisten kappaleiden tilavuuksien suhde on pituuksien suhteen kolmas potenssi, kuvan perusteella Ruotsin ja Suomen väkilukujen suhde on $\frac{9,9^3}{5,5^3} = 5,832$, joten Ruotsin väkiluku vaikuttaa kuutioiden perusteella noin kuusinkertaiselta Suomen väkilukuun verrattuna.

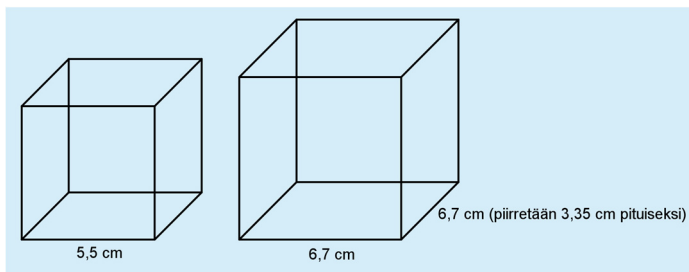
Vastaus: noin kuusinkertaiselta

- b)** Tilavuuksien suhde on mittakaavan kuutio. Muodostetaan yhtälö ja ratkaistaan sen avulla Ruotsin väkilukua esittävän kuution särmän pituus x .

$$\begin{aligned}\left(\frac{5,5}{x}\right)^3 &= \frac{5,5}{9,9} \\ \frac{5,5^3}{x^3} &= \frac{5,5}{9,9} \\ 5,5x^3 &= 9,9 \cdot 5,5^3 \quad \| : 5,5 \\ x^3 &= \frac{9,9 \cdot 5,5^3}{5,5} \\ x &= \sqrt[3]{\frac{9,9 \cdot 5,5^3}{5,5}} \\ x &= 6,690\dots\end{aligned}$$

Ruotsin väkilukua esittävän kuution särmän pituuden pitäisi olla noin 6,7 cm.

Piirretään kuutiot kavaljeeriperspektiivissä.



Vastaus: 6,7 cm

- 431. a)** Kertymä kohdassa $x = 20$ on $F(20) = 0,2 \cdot 0,375 - 7,3 = 0,2$, joten 20 % lapsista on nukahtanut klo 20:een mennessä.

Vastaus: 20 %

- b)** Tutkitaan yhtälöiden avulla, millä tilastomuuttujan x arvoilla kertymä on 0,9.
Etsitään ensin ratkaisua väliltä $[19, 20[$.

$$\begin{aligned} 0,2x - 3,8 &= 0,9 \\ 0,2x &= 0,9 + 3,8 \\ 0,2x &= 4,7 & \parallel : 0,2 \\ x &= 23,5 \end{aligned}$$

Tulos ei ole välillä $[19, 20[$, joten se ei kelpaa.

Etsitään sitten ratkaisua väliltä $[20, 22[$.

$$\begin{aligned} 0,375x - 7,3 &= 0,9 \\ 0,375x &= 0,9 + 7,3 \\ 0,375x &= 8,2 & \parallel : 0,375 \\ x &= 21,866... \end{aligned}$$

Tulos on välillä $[20, 22[$, joten se kelpaa.

Ratkaisu on $21,866... \text{ h} = 21 \text{ h } 52 \text{ min}$. 90 % lapsista on siis unessa klo 21.52.

Vastaus: klo 21.52

- e) Koska kertymäfunktion arvo kohdassa $x = 20$ on 0,2, niin 20 % lapsista nukahtaa välillä $[19, 20[$.

Kertymän arvo kohdassa $x = 22$ on

$F(22) = 0,05 \cdot 22 - 0,15 = 0,95$, joten 95 % lapsista on unessa klo 22.

Koska klo 20 lapsista oli unessa 20 %, niin 75 % lapsista nukahtaa välillä $[20, 22[$.

$F(23) = 0,05 \cdot 23 - 0,15 = 1$, joten 100 % lapsista on unessa klo 23.

Koska klo 22 lapsista oli unessa 95 %, niin 5 % lapsista nukahtaa välillä $[22, 23]$.

Täydennetään taulukko näiden tietojen avulla.

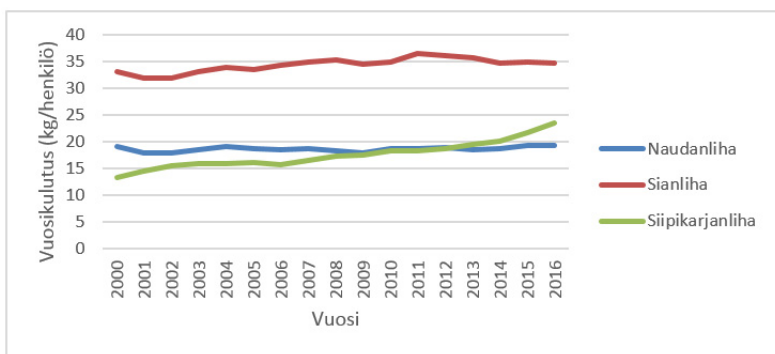
Nukahtamisaika	$f\%$
19–20	20 %
20–22	75 %
22–23	5 %

Vastaus:

Nukahtamisaika	$f\%$
19–20	20 %
20–22	75 %
22–23	5 %

432. Taulukossa on luonnonvarakeskuksen tilastoista vuonna 2017 löytynyt aineisto lihan vuosikulutuksesta (kg/henkilö).

Vuosi	Naudanliha	Sianliha	Siipikarjanliha
2000	19	33	13,3
2001	17,8	31,9	14,5
2002	17,9	31,9	15,4
2003	18,4	33	15,8
2004	19	33,8	15,9
2005	18,6	33,5	16,1
2006	18,5	34,3	15,7
2007	18,7	34,9	16,4
2008	18,2	35,3	17,2
2009	17,8	34,4	17,5
2010	18,6	34,9	18,2
2011	18,6	36,4	18,2
2012	18,9	36	18,7
2013	18,4	35,6	19,5
2014	18,7	34,6	20,1
2015	19,2	34,9	21,6
2016	19,2	34,7	23,5



Aineiston ja sen perusteella laaditun viivakaavion avulla voidaan sanoa, että siipikarjan lihan kulutus (kg/henkilö) on kasvanut 2000-luvun alusta lähtien, kun taas sian- ja naudanlihan kulutus on säilynyt jokseenkin ennallaan. Vuoden 2012 aikoihin siipikarjanlihan kulutus on ylittänyt naudanlihan kulutuksen.

4.2 Sijaintilukuja

ALOITA PERUSTEISTA

433. Moodi merkitään Mo , mediaani merkitään Md ja keskiarvo merkitään $\bar{x}$.

Vastaus: A: II, B: III ja C: I

434. Moodi Mo on yleisin arvo ja mediaani Md suuruusjärjestykseen asetetuista arvoista keskimmäinen tai kahden keskimmäisen keskiarvo.

a) Lukujen 1, 1, 2, 4 ja 7 moodi on $Mo = 1$ ja mediaani $Md = 2$.
Keskiarvo on

$$\bar{x} = \frac{1+1+2+4+7}{5} = \frac{15}{5} = 3.$$

Vastaus: $Mo = 1$, $Md = 2$, $\bar{x} = 3$

b) Kaikki neljä lukua ovat yhtä yleisiä, joten moodeja ovat 3, 6, 7 ja 8.

Keskimmäisiä ovat luvut 6 ja 7, joten mediaani on $Md = \frac{6+7}{2} = 6,5$.

Keskiarvo on $\bar{x} = \frac{3+6+7+8}{4} = \frac{24}{4} = 6$.

Vastaus: $Mo = 3$, $Mo = 6$, $Mo = 7$ ja $Mo = 8$, $Md = 6,5$, $\bar{x} = 6$

c) Järjestetään ensin luvut suuruusjärjestykseen: 10, 30, 40, 40.

Moodi on $Mo = 40$. Keskimmäisiä ovat luvut 30 ja 40, joten mediaani on $Md = \frac{30+40}{2} = 35$. Keskiarvo on $\bar{x} = \frac{10+30+40+40}{4} = \frac{120}{4} = 30$.

Vastaus: $Mo = 40$, $Md = 35$, $\bar{x} = 30$

435. Mediaani on se kurssien arvo, jonka kohdalla suhteellinen summafrekvenssi on 50 %. Kuvaajan perusteella mediaani on noin 39.

Vastaus: 39 kurssia

436. Kopioidaan luvut taulukkolaskentaohjelmaan ja määritetään keskiarvo, moodi ja mediaani ohjelman avulla.

	A	B	C	D
1	Askelia			
2	6790			
3	7712			
4	7543			
5	8924			
6	9972			
7	5214			
8	4505			
9	9972			
10	10076			
11	9521			
12	10374			
13	10642			
14	11021			
15	10989			
16	8528			
17	11153			
18				
19				
20	Keskiarvo	=KESKIARVO(A2:A17)		
21	Moodi			
22	Mediaani			
23				

	A	B	C
1	Askelia		
2	6790		
3	7712		
4	7543		
5	8924		
6	9972		
7	5214		
8	4505		
9	9972		
10	10076		
11	9521		
12	10374		
13	10642		
14	11021		
15	10989		
16	8528		
17	11153		
18			
19			
20	Keskiarvo	8933,5	
21	Moodi	=MOODI(A2:A17)	
22	Mediaani		
23			

	A	B	C
1	Askelia		
2	6790		
3	7712		
4	7543		
5	8924		
6	9972		
7	5214		
8	4505		
9	9972		
10	10076		
11	9521		
12	10374		
13	10642		
14	11021		
15	10989		
16	8528		
17	11153		
18			
19			
20	Keskiarvo	8933,5	
21	Moodi	9972	
22	Mediaani	=MEDIAANI(A2:A17)	
23			

Keskiarvo on 8933,5, moodi 9972 ja mediaani 9746,5.

Vastaus: $\bar{x} = 8933,5$, $Mo = 9972$, $Md = 9746,5$

437. a) Täydennetään taulukkoon summafrekvenssit ja suhteelliset summafrekvenssit.

Poissaoloja (oppituntia)	f	sf	$sf\%$
0–1	11	11	$\frac{11}{29} = 0,379... \approx 38\%$
2–3	10	$11 + 10 = 21$	$\frac{21}{29} = 0,724... \approx 72\%$
3–4	5	$21 + 5 = 26$	$\frac{26}{29} = 0,896... \approx 90\%$
5–6	3	$26 + 3 = 29$	100 %

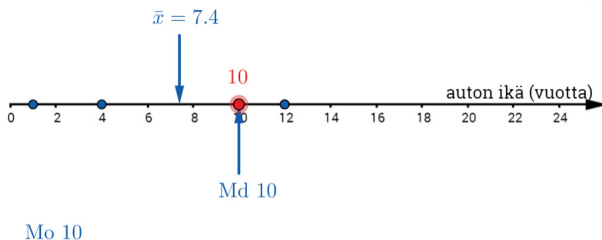
Vastaus:

Poissaoloja (oppituntia)	sf	$sf\%$
0–1	11	38 %
2–3	21	72 %
3–4	26	90 %
5–6	29	100 %

- b) Mediaaniluokka on luokka, jonka kohdalla summafrekvenssi ylittää 50 %. Mediaaniluokka on siis 2–3 oppituntia.

Vastaus: 2–3 oppituntia

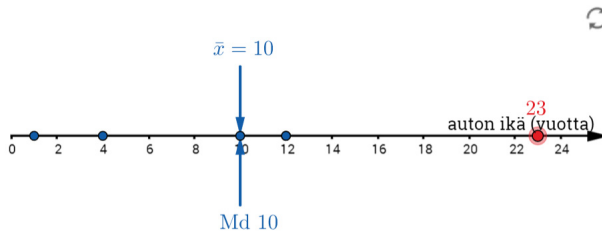
438. a) Tutkitaan appletin avulla, mikä viidennen auton iän pitäisi olla, jotta ikien moodi olisi 10 vuotta.



Viidennen auton iän on oltava 10 vuotta.

Vastaus: 10 vuotta

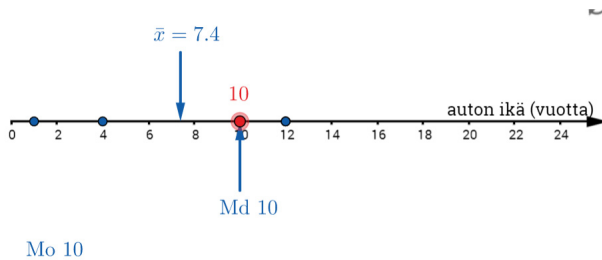
- b) Tutkitaan appletin avulla, mikä viidennen auton iän pitäisi olla, jotta ikien keskiarvo olisi 10 vuotta.



Viidennen auton iän on oltava 23 vuotta.

Vastaus: 23 vuotta

- c) Tutkitaan appletin avulla, mikä viidennen auton iän pitäisi olla, jotta ikien mediaani olisi 10 vuotta.



Ikien mediaani on 10 vuotta, kun viidennen auton ikä on vähintään 10 vuotta.

Vastaus: vähintään 10 vuotta

- 439. a)** Yöunen pituuden moodiluokka on luokka, jonka frekvenssi on suurin. Moodiluokka on siis 8–9 h.

Vastaus: 8–9 h

- b)** Täydennetään taulukkoon todelliset rajat ja luokkakeskukset.

Yöunen pituus (h)	Todellinen alaraja (h)	Todellinen yläraja (h)	Luokkakeskus (h)	f
4–5	3,5	5,5	$\frac{3,5 + 5,5}{2} = 4,5$	3
6–7	5,5	7,5	6,5	11
8–9	7,5	9,5	8,5	14
10–11	9,5	11,5	10,5	2

Vastaus:

Yöunen pituus (h)	Todellinen alaraja (h)	Todellinen yläraja (h)	Luokkakeskus (h)	f
4–5	3,5	5,5	4,5	3
6–7	5,5	7,5	6,5	11
8–9	7,5	9,5	8,5	14
10–11	9,5	11,5	10,5	2

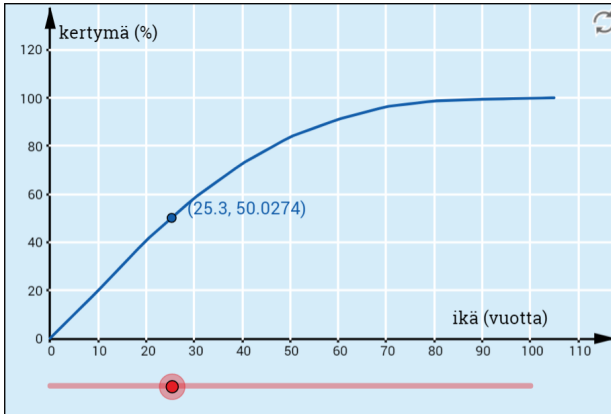
- c)** Lasketaan keskimääräinen yöunen pituus luokkakeskusten avulla.

$$\bar{x} = \frac{3 \cdot 4,5 + 11 \cdot 6,5 + 14 \cdot 8,5 + 2 \cdot 10,5}{3 + 11 + 14 + 2} = \frac{225}{30} = 7,5$$

Lukiolaisen keskimääräinen yöunen pituus on 7,5 h.

Vastaus: 7,5 h

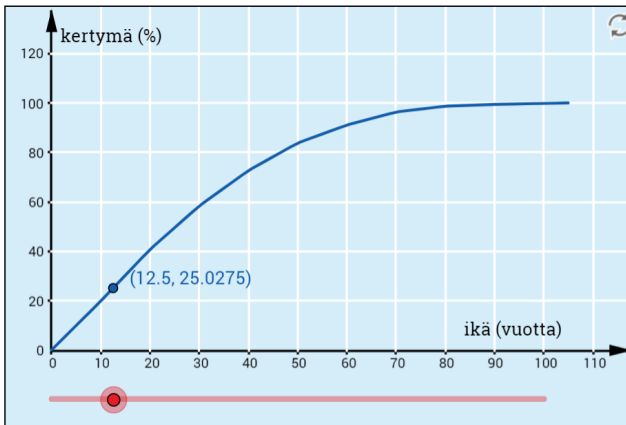
440. a) Määritetään appletin avulla mediaani-ikä.



Mediaani-ikä on appletin perusteella noin 25 vuotta.

Vastaus: 25 vuotta

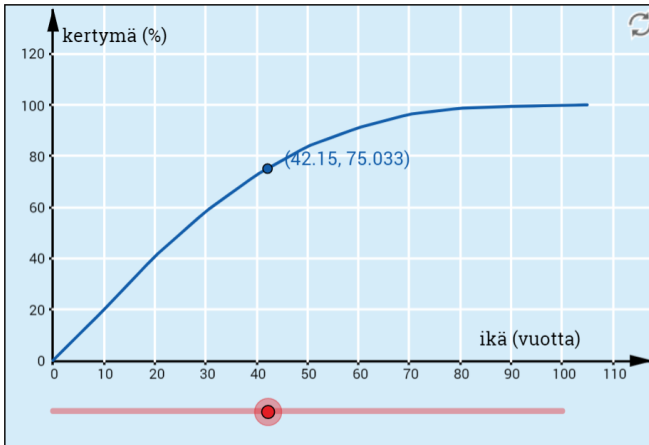
b) Määritetään appletin avulla ikä, jota nuorempia on 25 % intialaisista.



Appletin perusteella noin 25 % intialaisista on alle 13 vuotiaita.

Vastaus: 13 vuotta

- c) Määritetään appletin avulla ikä, jota vanhempia on 25 % intialaisista.



Appletin perusteella 25 % intialaisista on yli 42 vuotiaita.

Vastaus: 42 vuotta

- d) Intian syntyvyys ja toisaalta myös lapsikuolleisuus ovat suhteellisesti suurempia kuin Suomessa. Tästä johtuen lapsia on suhteellisesti paljon, mutta he eivät elä niin vanhoiksi kuin Suomessa.

441. Lasketaan erikokoisten lasien frekvenssit suhteellisten frekvenssien avulla, kun tiedetään, että laseja on yhteensä 300.

Tilavuus (dl)	$f\%$	f
2,0	48 %	$0,48 \cdot 300 = 144$
3,4	35 %	$0,35 \cdot 300 = 105$
5,6	17 %	$0,17 \cdot 300 = 51$

Lasketaan tilavuuksien keskiarvo frekvenssien avulla.

$$\bar{x} = \frac{144 \cdot 2,0 + 105 \cdot 3,4 + 51 \cdot 5,6}{300} = 3,102 \approx 3,1$$

Juomalasien tilavuuksien keskiarvo on noin 3,1 dl.

Vastaus: 3,1 dl

- 442.** Keskiarvoa laskettaessa voidaan ajatella, että kaikkien viiden ensimmäisen kurssin arvosana on 7. Kuudes arvosana on 10, saadaan keskiarvoksi $\bar{x} = \frac{5 \cdot 7 + 10}{6} = 7,5 \approx 8$.

Vastaus: Arvosanan 8

VAHVISTA OSAAMISTA

- 443. a)** Silmien värien moodi on sininen, koska se on yleisin havaintoarvo.

Vaatekokojen moodi on 42, koska se on yleisin havaintoarvo.

Kokoja on yhteensä $5 + 8 + 16 + 18 + 15 = 62$. Puolet siitä on $\frac{62}{2} = 31$, joten 31. ja 32. koko ovat keskimmäisiä. Kokoja 36, 38 ja 40 on yhteensä $5 + 8 + 16 = 29$ ja niitä seuraavat 18 ovat kokoa 42, joten keskimäinen koko on 42.

Lasketaan aineiston B keskiarvo

$$\bar{x} = \frac{5 \cdot 36 + 8 \cdot 38 + 16 \cdot 40 + 18 \cdot 42 + 15 \cdot 44}{62} = \frac{2540}{62} = 40,967... \approx 41.$$

Vastaus: A: Mo = sininen, B: Mo = 42, Md = 42, $\bar{x} \approx 41$

- b)** Mediaania ei ole mahdollista määrittää, koska havaintoarvoja ei voida asettaa suuruusjärjestykseen. Keskiarvoa ei voi määrittää, koska havaintoarvot eivät ole lukuarvoja.

- 444. a)** Moodiluokka on luokka, jonka frekvenssi on suurin. Moodiluokka on siis 160–164 cm.

Vastaus: 160–164 cm

b) Lasketaan summafrekvenssit mediaanin määrittämistä varten.

Pituus (cm)	f	sf
150–154	3	3
155–159	6	$3 + 6 = 9$
160–164	17	$9 + 17 = 26$
165–169	15	$26 + 15 = 41$
170–174	12	$41 + 12 = 53$
175–179	11	$53 + 11 = 64$
180–184	14	$64 + 14 = 78$
185–189	6	$78 + 6 = 84$
190–194	2	$84 + 2 = 86$

Mediaaniluokka on luokka, jossa keskimäinen pituus on. Pituuksia on yhteensä 86 ja $\frac{86}{2} = 43$. Pituudet 43. ja 44. ovat keskimäisiä. 43. ja 44. pituus ovat luokassa 170–174 cm, joten se on mediaaniluokka.

Vastaus: 170–174 cm

c) Täydennetään keskiarvon laskemista varten taulukkoon luokkakeskukset.

Pituus (cm)	f	Luokkakeskus (cm)
150–154	3	$\frac{149,5 + 154,5}{2} = 152$
155–159	6	157
160–164	17	162
165–169	15	167
170–174	12	172
175–179	11	177
180–184	14	182
185–189	6	187
190–194	2	192

Keskiarvo on

$$\bar{x} = \frac{3 \cdot 152 + 6 \cdot 157 + 17 \cdot 162 + \dots + 2 \cdot 192}{86} = 171,186\dots$$

Keskiarvo on noin 172 cm.

Vastaus: 172 cm

- 445.** Keskilukujen määrittämistä varten järjestetään vastaukset suuruusjärjestykseen: 1, 1, 2, 2, 2, 3, 3, 4, 4, 4, 4, 5, 5, 5, 5. Vastauksia on yhteensä 16.

a) Vastausten moodi on 4, koska se on yleisin vastaus.

Vastaus: $Mo = 4$

b) Vastausten keskiarvo on

$$\bar{x} = \frac{2 \cdot 1 + 3 \cdot 2 + 2 \cdot 3 + 5 \cdot 4 + 4 \cdot 5}{16} = 3,375 \approx 3,4.$$

Vastaus: $\bar{x} = 3,4$

c) Vastauksia on 16, joten 8. ja 9. vastaus ovat keskimmäiset vastaukset. Sekä 8. että 9. vastaus on 4, joten mediaani on 4.

Vastaus: $Md = 4$

- 446.** Kopioidaan arvot taulukkolaskentaohjelmaan. Määritetään keskiarvo, moodi ja mediaani ohjelman avulla.

	A	B	C
1	Voitto		
2	25	Keskiarvo	=KESKJARVO(A2:A21)
3	-3	Moodi	=MOODI(A2:A21)
4	-2	Mediaani	=MEDIAANI(A2:A21)
5	-9		
6	17		
7	-6		
8	-8		
9	-2		
10	0		
11	-4		
12	-11		
13	-5		
14	-3		
15	-2		
16	-9		
17	-12		
18	10		
19	-9		
20	-2		
21	-1		

	A	B	C
1	Voitto		
2	25	Keskiarvo	-1,8
3	-3	Moodi	-2
4	-2	Mediaani	-3
5	-9		
6	17		
7	-6		
8	-8		
9	-2		
10	0		
11	-4		
12	-11		
13	-5		
14	-3		
15	-2		
16	-9		
17	-12		
18	10		
19	-9		
20	-2		
21	-1		

Voiton keskiarvo on -1,8 euroa, moodi -2 euroa ja mediaani -3 euroa. Negatiiviset voitot tarkoittavat sitä, että pelaajat ovat jääneet tappiolle.

Vastaus: $\bar{x} = -1,8$ €, $Mo = -2$ €, $Md = -3$ €. Pelaajat jäivät tappiolle.

- 447.** Keskilukujen määrittämistä varten järjestetään arvot suuruusjärjestykseen: 1, 5, 6, 6, 7, 9, 9, 9, 12, 12. Arvoja on yhteensä 10 kpl.

Aineiston moodi on $M_o = 9$, koska se on yleisin arvo.

Aineiston mediaani on viidennen ja kuudennen arvon keskiarvo, eli

$$M_d = \frac{7+9}{2} = 8.$$

$$\text{Aineiston keskiarvo on } \bar{x} = \frac{1+5+2\cdot 6+7+3\cdot 9+2\cdot 12}{10} = \frac{76}{10} = 7,6.$$

- a)** Moodi on 9 ja mediaani 8, joten moodi on suurempi kuin mediaani.
Väite on siis epätosi.

Vastaus: epätosi, suurempi

- b)** Mediaani on 8, joten väite on tosi.

Vastaus: tosi

- c)** Mediaani on 8 ja keskiarvo 7,6, joten väite on tosi.

Vastaus: tosi

- d)** Keskiarvo on 7,6, joten väite on epätosi.

Vastaus: epätosi; 7,6

- 448.** Mediaani on 7 kaaviossa II, koska havaintoarvot ovat 4, 5, 6, 6, 7, 7, 7, 7, 8 ja 9. Havaintoarvoja on yhteensä 10, joten viides ja kuudes havaintoarvo ovat keskimmäisiä. Viides ja kuudes havaintoarvo ovat molemmat 7, joten mediaani on 7. Tunnusluku A kuuluu kaavioon II.

Moodi on 6 kaaviossa III, koska lukua 6 kuvaava sektori on suurin. Tunnusluku B kuuluu kaavioon III.

Keskiarvo on 7 kaaviossa I, koska luvut ovat jakautuneet symmetrisesti luvun 7 ympärille. Tunnusluku C kuuluu kaavioon I.

Vastaus: A: II, B: III ja C: I

- 449.** Appletin avulla saadaan lukujen 1, 2, 3, 4, 7, 7, 8 ja 9 moodiksi 7, mediaaniksi 5,5 ja keskiarvoksi 5,13.
- a) Kun pienin luvuista kasvaa seitsemällä, ovat moodina luvut 7 ja 8, mediaani on 7 ja keskiarvo 6. Moodiksi tulevat 7 ja 8, mediaani kasvaa 1,5:llä ja keskiarvo kasvaa 0,87:llä.

Vastaus: Moodiksi tulevat 7 ja 8, mediaani kasvaa 1,5:llä ja keskiarvo kasvaa 0,87:llä.

- b) Kun suurin luvuista kaksinkertaistuu, on moodi 7, mediaani 5,5 keskiarvo 6,25. Moodi ja mediaani eivät muutu, keskiarvo kasvaa 1,12:lla.

Vastaus: Moodi ja mediaani eivät muutu, keskiarvo kasvaa 1,12:lla.

- c) Kun jokainen luku kasvaa seitsemällä, moodi 14, mediaani 12,5 ja keskiarvo 12,13. Kaikki kolme tunnuslukua kasvavat 7:llä.

Vastaus: Kaikki kolme tunnuslukua kasvavat 7:llä.

- d) Kun jokainen luku kaksinkertaistuu, on moodi 14 ja mediaani 11 keskiarvo 10,26. Jokainen tunnusluku siis kaksinkertaistuu.

Vastaus: Kaikki kolme tunnuslukua kaksinkertaistuvat.

- 450.** Valitaan luvut esimerkiksi niin, että luvuista kaksi on samoja ja kaikki muut keskenään eri lukuja, ja yksi luku poikkeaa muista suuresti. Tällaiset luvut ovat esimerkiksi 0, 0, 1, 2 ja 10. Lukujen moodi on 0 ja mediaani 1. Keskiarvo on $\frac{0+0+1+2+10}{5} = 2,6$.

Vastaus: esim. 0, 0, 1, 2 ja 10

- 451.** Merkitään viidennen kurssin vähimmäisarvosanaa kirjaimella x ja ratkaistaan se yhtälön avulla.

$$\begin{aligned}\frac{8+7+6+8+x}{5} &= 7,5 \\ \frac{29+x}{5} &= 7,5 \quad \| \cdot 5 \\ 29+x &= 37,5 \\ x &= 37,5 - 29 \\ x &= 8,5\end{aligned}$$

Karoliinan pitäisi saada viidennestä kurssista vähintään arvosana yli 8,5 ja koska kurssiarvosanat ovat kokonaislukuja, on hänen saatava arvosana 9.

Vastaus: arvosana 9

452. a) Kokojen moodiluokka on 40, koska se on yleisin koko.

Määritetään kaksi ensimmäistä suhteellista summafrekvenssiä.

	A	B	C
1	Kokonro	f %	sf %
2	32	0,2	0,2
3	34	3,2	=C2+B3
4	36	8,4	
5	38	17,8	
6	40	20	
7	42	13,7	
8	44	11,1	
9	46	10,6	
10	48	7,6	
11	50	4,1	
12	52	1,5	
13	54	1,1	
14	56	0,7	

Kopioidaan suhteellisen summafrekvenssin laskukaava seuraaville riveille.

	A	B	C
1	Kokonro	f %	sf %
2	32	0,2	0,2
3	34	3,2	3,4
4	36	8,4	11,8
5	38	17,8	29,6
6	40	20	49,6
7	42	13,7	63,3
8	44	11,1	74,4
9	46	10,6	85
10	48	7,6	92,6
11	50	4,1	96,7
12	52	1,5	98,2
13	54	1,1	99,3
14	56	0,7	100

Suhteellinen summafrekvenssi ylittää 25 % koon 38 kohdalla, joten alakvartiili on 38.

Suhteellinen summafrekvenssi ylittää 50 % koon 42 kohdalla, joten mediaani on 42.

Suhteellinen summafrekvenssi ylittää 75 % koon 46 kohdalla, joten yläkvartiili on 46.

Koska vaatteiden kokonaismäärää ei ole kerrottu, merkitään sitä kirjaimella n . Tällöin kokoa 32 olevien vaatteiden määrä on $0,002n$ jne.

Lasketaan vaatekoon keskiarvo.

$$\begin{aligned}\bar{x} &= \frac{0,002n \cdot 32 + 0,032n \cdot 34 + \dots + 0,007n \cdot 56}{n} \\ &= \frac{41,918n}{n} \\ &= 41,918n^{(n)} \\ &= 41,918 \\ &\approx 42\end{aligned}$$

Kokojen keskiarvo on noin 42.

Vastaus: $Mo = 40$, $Q_1 = 38$, $Q_2 = 42$, $Q_3 = 46$, $\bar{x} \approx 42$

- b) Koska kokonaistilausmäärä on noin 100, on kutakin kokoa tilattava koon suhteellisen frekvenssin verran.

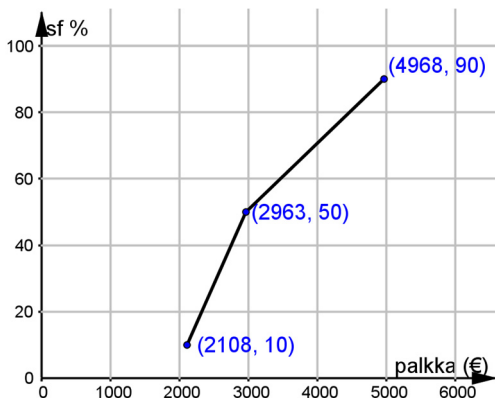
Pyöristetään saadut desimaaliluvut ylöspäin seuraavaan kokonaislukuun, jotta jokaista kokoa on jakauman mukaan riittävästi.

	A	B	C	D	E	F	G
1	Kokonro	f %	tilattava määrä	pyöristys seuraavaan kokonaislukuun			
2	32	0,2	0,2	1			
3	34	3,2	3,2	4			
4	36	8,4	8,4	9			
5	38	17,8	17,8	18			
6	40	20	20	20			
7	42	13,7	13,7	14			
8	44	11,1	11,1	12			
9	46	10,6	10,6	11			
10	48	7,6	7,6	8			
11	50	4,1	4,1	5			
12	52	1,5	1,5	2			
13	54	1,1	1,1	2			
14	56	0,7	0,7	1			
15			yhteensä	107			

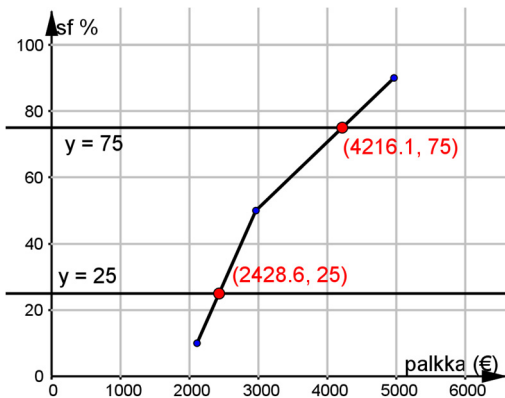
Vastaus: esim.

Koko	32	34	36	38	40	42	44	46	48	50
Tilattava määrä	1	4	9	18	20	14	12	11	8	5

453. Suhteellisen summafrekvenssin kuvaaja kulkee pisteiden (2108, 10), (2963, 50) ja (4968, 90) kautta. Piirretään niiden kautta kertymän kuvaaja sopivalla ohjelmalla.



Määritetään ala- ja yläkvartiili piirtämällä suorat $y = 25$ ja $y = 75$ ja määrittämällä niiden ja kertymän kuvaajan leikkauspisteet.



Leikkauspisteet ovat (2428,6; 25) ja (4216,1; 75), joten alakvartiili on $2428,6 \approx 2429$ euroa ja yläkvartiili $4216,1 \approx 4216$ euroa.

Vastaus: $Q_1 \approx 2429$ €, $Q_3 \approx 4216$ €

- 454.** a) Jos aineiston jokaisesta havaintoarvosta vähennetään yksi, keskiarvokin pienenee yhdellä. Väite on siis tosi.

Vastaus: tosi

- b) Jos vain suurin havaintoarvo kasvaa muiden pysyessä samoina, havaintoarvojen suuruusjärjestys ei muutu. Tällöin mediaanikaan ei muutu. Väite on epätosi.

Vastaus: epätosi, pysyy samana

- c) Havaintoarvoista 75 % on suurempia tai yhtä suuria kuin alakvartiili Q_1 , joten väite on tosi.

Vastaus: tosi

- 455.** Kaikkien viiden luvun summa on $2 \cdot 42 + 3 \cdot 54 = 246$. Kaikkien viiden luvun keskiarvo on $\frac{246}{5} = 49,2$.

Vastaus: 49,2

- 456.** Keskiarvo on satunnaisista kovista pakkasista huolimatta noin $-5\text{ }^{\circ}\text{C}$. 25 asteen pakkaset poikkeavat keskiarvosta noin 20 asteen verran alaspäin, mutta yhtä suuria poikkeamia keskiarvosta ylöspäin ei ole. Koska suuria poikkeamia keskiarvosta ylöspäin ei ole, täytyy sellaisia päiviä, joina lämpötilan keskiarvo ylittyy, olla enemmän kuin päiviä joina keskiarvo alittuu. Lämpötilan mediaanin on siten oltava korkeampi kuin keskiarvo.

Vastaus: mediaani

- 457. a)** Kaikkien yhdentoista henkilön pituuksien summa on
 $10 \cdot 172 \text{ cm} + 182 \text{ cm} = 1902 \text{ cm}$. Keskiarvo on
$$\frac{1902 \text{ cm}}{11} = 172,909... \text{ cm} \approx 173 \text{ cm}.$$

Vastaus: 173 cm

- b)** Kaikkien 21 henkilön pituuksien summa on
 $20 \cdot 172 \text{ cm} + 182 \text{ cm} = 3622 \text{ cm}$. Keskiarvo on
$$\frac{3622 \text{ cm}}{21} = 172,476... \text{ cm} \approx 172 \text{ cm}.$$

Vastaus: 172 cm

- c)** Merkitään henkilön pituutta senttimetreinä kirjaimella x . Muodostetaan yhtälö ja ratkaistaan se.

$$\begin{aligned}\frac{20 \cdot 172 + x}{21} &= 173 & \parallel \cdot 21 \\ 3440 + x &= 3633 \\ x &= 3633 - 3440 \\ x &= 193\end{aligned}$$

Henkilön pitäisi olla 193 cm pitkä.

Vastaus: 193 cm

- 458.** a) Luokkakeskus on luokan todellisen ala- ja todellisen ylärajan keskiarvo. Täydennetään taulukkoon todelliset rajat ja luokkakeskukset.

Odotusaika (min)	Todellinen alaraja (min)	Todellinen yläraja (min)	Luokkakeskus (min)	f
0–4	0	4,5	$\frac{0+4,5}{2} = 2,25$	3
5–9	4,5	9,5	$\frac{4,5+9,5}{2} = 7,0$	12
10–14	9,5	14,5	$\frac{9,5+14,5}{2} = 12$	40
15–19	14,5	19,5	$\frac{14,5+19,5}{2} = 17,0$	31
20–24	19,5	24,5	$\frac{19,5+24,5}{2} = 22$	10
25–29	24,5	29,5	$\frac{24,5+29,5}{2} = 27$	2

Keskiarvo on

$$\begin{aligned}\bar{x} &= \frac{3 \cdot 2,25 + 12 \cdot 7,0 + 40 \cdot 12,0 + 31 \cdot 17,0 + 10 \cdot 22,0 + 2 \cdot 27,0}{3 + 12 + 40 + 31 + 10 + 2} \\ &= \frac{1371,75}{98} \\ &= 13,997... \\ &\approx 14,0\end{aligned}$$

Vastaus:

Odotusaika (min)	Luokkakeskus (min)
0–4	2,25
5–9	7,0
10–14	12
15–19	17,0
20–24	22
25–29	27

$$\bar{x} \approx 14,0 \text{ min}$$

b) Luokitellaan odotusajat 10 minuutin pituisiin luokkiin.

Aika (min)	Todellinen alaraja (min)	Todellinen yläraja (min)	Luokkakeskus (min)	f
0–9	0	9,5	$\frac{0+9,5}{2} = 4,75$	$3 + 12 = 15$
10–19	9,5	19,5	$\frac{9,5+19,5}{2} = 14,5$	$40 + 31 = 71$
20–29	19,5	29,5	$\frac{19,5+29,5}{2} = 24,5$	$10 + 2 = 12$

Keskiarvo on

$$\bar{x} = \frac{15 \cdot 4,75 + 71 \cdot 14,5 + 12 \cdot 24,5}{98} = 14,232... \approx 14,2$$

a- ja b-kohdan tulokset ovat likimain yhtä suuria.

Vastaus:

Odotusaika (min)	f
0–9	15
10–19	71
20–29	12

$\bar{x} \approx 14,2$ min. Tulokset ovat likimain yhtä suuria.

459. Merkitään ryhmän B opiskelijoiden keskiarvoa kirjaimella x . Opiskelijoita oli yhteensä $27 + 34 = 61$. Muodostetaan yhtälö ja ratkaistaan siitä x .

$$\begin{aligned} \frac{27 \cdot 7,25 + 34x}{61} &= 7,14 \\ \frac{195,75 + 34x}{61} &= 7,14 && \parallel \cdot 61 \\ 195,75 + 34x &= 435,54 \\ 34x &= 435,54 - 195,75 \\ 34x &= 239,79 && \parallel : 34 \\ x &= 7,052... \end{aligned}$$

Ryhmän B opiskelijoiden keskiarvo oli noin 7,05.

Vastaus: 7,05

SYVENNÄ YMMÄRRYSTÄ

- 460.** Kopioidaan taulukon tiedot taulukkolaskentaohjelmaan ja määritetään luokkakeskukset sekä tulojen moodi ja keskiarvo. Ensimmäisen luokan todellinen alaraja on 0 ja todellinen yläraja 4999,5. Keskiarvon laskemista varten kerrotaan luokan luokkakeskus luokan frekvenssillä.

	A	B	C	D
1	tulot (€)	f	luokkakeskus x	frekvenssi*luokkakeskus
2	– 4 999	430404	2499,75	1075902399,00
3	5 000 – 9 999	469221	7499,5	3518922889,50
4	10 000 – 14 999	597134	12499,5	7463876433,00
5	15 000 – 19 999	516298	17499,5	9034956851,00
6	20 000 – 24 999	463222	22499,5	10422263389,00
7	25 000 – 29 999	464417	27499,5	12771235291,50
8	30 000 – 39 999	745083	34999,5	26077532458,50
9	40 000 – 49 999	397752	44999,5	17898641124,00
10	50 000 – 59 999	206779	54999,5	11372741610,50
11	60 000 – 79 999	174565	69999,5	12219462717,50
12	80 000 – 99 999	62249	89999,5	5602378875,50
13	Ylimmän tuloluokan tulojen keskiarvo 893 523	67301	893523	60134991423,00
14	yhteensä	4594425		177592905462,00
15				

Huomataan, että moodiluokka on 30 000 – 39 999 €.

Tulojen keskiarvo saadaan jakamalla viimeisen sarakkeen summa 177 592 905 462 € tulonsaajien kokonaismäärällä 4 594 925. Keskiarvo

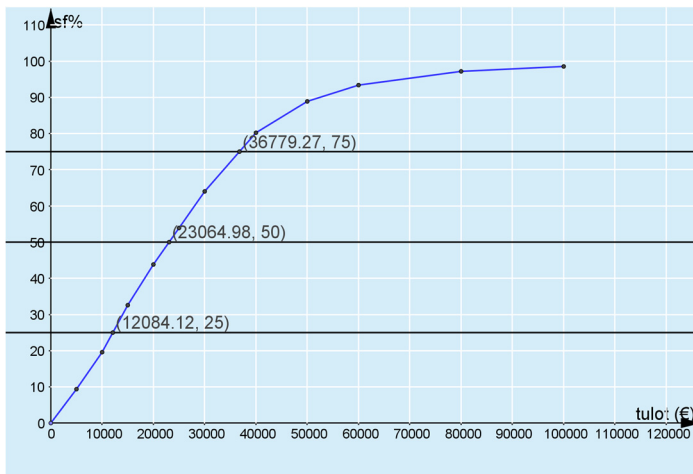
$$\text{on } \bar{x} = \frac{177\,592\,905\,462\,€}{4\,594\,925} = 38\,649,794...€ \approx 38\,600\,€.$$

Kvartiilien määrittämiseksi lasketaan jakauman todelliset ylärajat ja suhteelliset summafrekvenssit.

tulot (€)	todellinen yläraja (€)	f	$f\%$	$sf\%$
– 4 999	4999,5	430404	9,37	9,37
5 000 – 9 999	9999,5	469221	10,21	19,58
10 000 – 14 999	14999,5	597134	13,00	32,58
15 000 – 19 999	19999,5	516298	11,24	43,82
20 000 – 24 999	24999,5	463222	10,08	53,90
25 000 – 29 999	29999,5	464417	10,11	64,01
30 000 – 39 999	39999,5	745083	16,22	80,22
40 000 – 49 999	49999,5	397752	8,66	88,88
50 000 – 59 999	59999,5	206779	4,50	93,38
60 000 – 79 999	79999,5	174565	3,80	97,18
80 000 – 99 999	99999,5	62249	1,35	98,54
Ylimmän tuloluokan tulojen keskiarvo 893 523		67301	1,46	100,00
yhteensä		4594425		

Kvartiilien määrittämiseksi piirretään kertymää kuvaava viivakaavio. Viivakaavion kuvaajan pisteiden x -koordinaatti on luokan todellinen yläraja ja y -koordinaatti luokan suhteellinen frekvenssi. Ylimmällä tuloluokalla ei ole ylärajaa, joten kyseistä riviä ei voida täyttää. Tällä ei kuitenkaan ole merkitystä kvartiilien kannalta.

Selvitetään kvartiilit piirtämällä suorat $y = 25$, $y = 50$ ja $y = 75$ ja määrittämällä suorien ja viivakaavion kuvaajan leikkauspisteet.



Alakvartiili on noin 12 100 euroa, mediaani noin 23 100 euroa ja yläkvartiili noin 36 800 euroa.

Vastaus: $Mo \approx 30\,000 - 39\,999$ €, $\bar{x} \approx 38\,600$ €, $Q_1 \approx 12\,100$ €, $Q_2 \approx 23\,100$ € ja $Q_3 \approx 36\,800$ €

- b)** Keskiarvo on noin $38\,600$ € – $23\,100$ € = $15\,500$ € suurempi kuin mediaani. Keskiarvo- ja mediaanitulon välinen ero johtuu väestön tuloeroista. Suuret poikkeamat vaikuttavat keskiarvoon enemmän kuin mediaaniin. Jos pienellä joukolla ihmisiä on todella suuret tulot, heidän tulonsa nostavat keskiarvoa selvästi, mutta eivät juuri vaikuta mediaaniin, koska hyvin suurituloisia ihmisiä on vain vähän.

Vastaus: Keskiarvo on n. $15\,500$ € suurempi kuin mediaani.

- 461. a)** Esimerkiksi lukujen 0, 1 ja 11 keskiarvo on $\frac{0+1+11}{3} = 4 \neq 1$, joten väite on epätosi.

Vastaus: epätosi

- b)** Merkitään pienintä lukua kirjaimella x . Seuraavat kaksi kokonaislukua ovat $x + 1$ ja $x + 2$. Lukujen keskiarvo on
- $$\frac{x + (x + 1) + (x + 2)}{3} = \frac{3x + 3}{3} = x + 1, \text{ eli sama kuin keskimääräinen luku.}$$

Väite on siis tosi.

Vastaus: tosi

- c)** Esimerkiksi lukujen 0, 1, 2, 3, 3 mediaani on 2 ja keskiarvo 1,8. Vain 2 lukua on keskiarvoa pienempiä

Vastaus: epätosi

- 462.** Muodostetaan kuvan perusteella frekvenssitaulukko 13–17-vuotiaiden internetissä viettämästä ajasta.

Aika (h)	$f\%$
0–1	1
1–2	5
3–5	13
6–9	14
10–14	17
15–20	16
21–30	15
31–40	9
41–50	5
51–70	6

Keskiarvon laskemista varten täydennetään taulukkoon todelliset rajat ja luokkakeskukset.

Aika (h)	Todellinen alaraja (h)	Todellinen yläraja (h)	Luokkakeskus (h)	$f\%$
0–1	0	1,5	$\frac{0+1,5}{2} = 0,75$	1
1–2	1,5	2,5	1,5	5
3–5	2,5	5,5	4	13
6–9	5,5	9,5	7,5	14
10–14	9,5	14,5	12	17
15–20	14,5	29,5	17,5	16
21–30	20,5	30,5	25,5	15
31–40	30,5	40,5	35,5	9
41–50	40,5	50,5	45,5	5
51–70	50,5	70,5	60,5	6

Koska nuorten kokonaismäärää ei ole kerrottu, merkitään sitä kirjaimella n . Tällöin luokan 0–1 h frekvenssi on $0,01n$ jne.

Lasketaan käytetyn ajan keskiarvo.

$$\begin{aligned}
 \bar{x} &= \frac{0,01n \cdot 0,75 + 0,05n \cdot 1,5 + \dots + 0,06n \cdot 60,5}{n} \\
 &= \frac{19,4175n}{n} \\
 &= 19,4175
 \end{aligned}$$

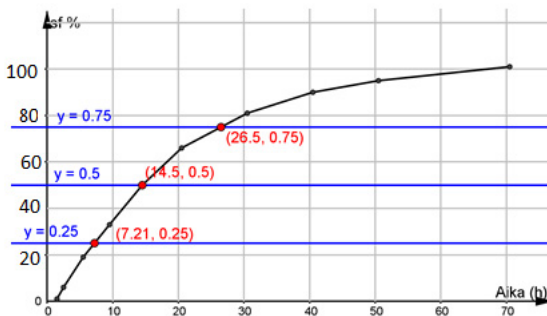
Keskiarvoksi saadaan $19,4175 \text{ h} = 19 \text{ h} + 0,4175 \cdot 60 \text{ min} \approx 19 \text{ h } 25 \text{ min}$.

Kvartiilien määrittämiseksi lasketaan jakauman suhteelliset summafrekvenssit.

Aika (h)	Todellinen yläraja (h)	$sf\%$
0–1	1,5	1
1–2	2,5	6
3–5	5,5	19
6–9	9,5	33
10–14	14,5	50
15–20	29,5	66
21–30	30,5	81
31–40	40,5	90
41–50	50,5	95
51–70	70,5	101

Kvartiilien määrittämiseksi piirretään kertymää kuvaava viivakaavio. Viivakaavion kuvaajan pisteiden x -koordinaatti on luokan todellinen yläraja ja y -koordinaatti luokan suhteellinen frekvenssi.

Määritetään kvartiilit piirtämällä suorat $y = 0,25$, $y = 0,5$ ja $y = 0,75$ ja määrittämällä niiden ja viivakaavion kuvaajan leikkauspisteet.



Alakvartiili on 7,21 h, keskikvartiili eli mediaani 14,5 h ja yläkvartiili 26,5 h. Tunneiksi ja minuuteiksi muutettuina nämä ovat noin 7 h 13 min, 14 h 30 min ja 26 h 30 min.

Vastaus: $\bar{x} \approx 19$ h 25 min, $Q_1 \approx 7$ h 13 min, $Q_2 \approx 14$ h 30 min, $Q_3 \approx 26$ h 30 min

463. Keskilukujen ja kvartiilien määrittämiseksi lasketaan jakauman todelliset rajat ja suhteelliset summafrekvenssit sekä luokkakeskukset.

Oletetaan viimeisen luokan ylärajaksi 90 vuotta, jotta voidaan laskea jakauman keskiarvo. Ylärajojen määrittämisessä on otettava huomioon, että ikä pyöristetään aina alaspäin.

	A	B	C	D	E	F
1	Ikä (vuotta)	f	f %	todellinen alaraja	todellinen yläraja	Luokkakeskus
2	15-19	506	2.07	15	20	17.5
3	20-24	3453	14.12	20	25	22.5
4	25-29	6878	28.12	25	30	27.5
5	30-34	5184	21.19	30	35	32.5
6	35-39	2737	11.19	35	40	37.5
7	40-44	1630	6.66	40	45	42.5
8	45-49	1939	7.93	45	50	47.5
9	50-54	937	3.83	50	55	52.5
10	55-59	597	2.44	55	60	57.5
11	60-64	318	1.3	60	65	62.5
12	65-69	185	0.76	65	70	67.5
13	70-90	98	0.4	70	90	80
14		24462				

Tehdään sarake frekvenssin ja luokkakeskuksen tulolle.

Keskiarvo saadaan jakamalla tulosten summa 831 910 kokonaisfrekvenssillä 24 462.

$$\bar{x} = \frac{831\,910}{24\,462} = 34,008... \approx 34.$$

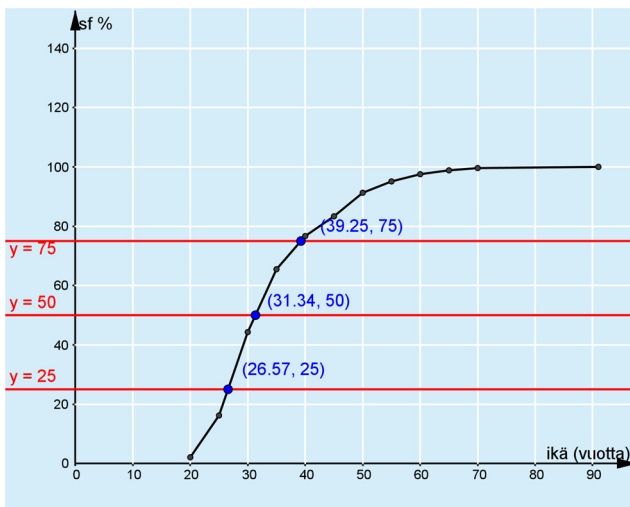
Avioitumisiän keskiarvo on noin 34 vuotta.

Luokkakeskus	f	f * luokkakeskus
17.5	506	8855
22.5	3453	77692.5
27.5	6878	189145
32.5	5184	168480
37.5	2737	102637.5
42.5	1630	69275
47.5	1939	92102.5
52.5	937	49192.5
57.5	597	34327.5
62.5	318	19875
67.5	185	12487.5
80	98	7840
	24462	831910

Kvartiilit määritetään kertymän kuvaajasta. Kuvaajan pisteiden x -koordinaatit ovat luokkien ylärajoja ja y -koordinaatit suhteellisia summafrekvenssejä.

todellinen yläraja	sf %
20	2.07
25	16.18
30	44.3
35	65.49
40	76.68
45	83.35
50	91.27
55	95.1
60	97.54
65	98.84
70	99.6
91	100

Piirretään kertymän kuvaaja.
Määritetään kvartiilit piirtämällä suorat $y = 25$, $y = 50$ ja $y = 75$ ja määrittämällä niiden ja kertymän kuvaajan leikkauspisteet.



Alakvartiili on noin 27 vuotta, mediaani noin 31 vuotta ja yläkvartiili 39 vuotta.

Vastaus: $Mo = 25\text{--}29$ vuotta, $\bar{x} \approx 34$ vuotta, $Q_1 \approx 27$ vuotta, $Q_2 \approx 31$ vuotta, $Q_3 \approx 39$ vuotta

464. a) Täydennetään taulukkoon suhteelliset frekvenssit.

	Miehet	Naiset	f % miehet	f % naiset
0-9	132	103	0.53	0.42
10-19	129	63	0.52	0.26
20-29	376	121	1.52	0.5
30-39	519	161	2.09	0.66
40-49	1232	527	4.97	2.17
50-59	3228	1366	13.01	5.63
60-69	4565	2195	18.4	9.04
70-79	6598	4522	26.6	18.63
80-89	6399	9845	25.79	40.57
90-99	1600	5184	6.45	21.36
100-	31	181	0.12	0.75
Yht.	24809	24268		

Määritetään luokkien todelliset rajat ja luokkakeskukset. Otetaan huomioon, että henkilö on 9-vuotias vielä päivää ennen kuin hän täyttää 10. Keskiarvon laskemiseksi oletetaan viimeisen luokan ylärajaksi 110 vuotta.

	Miehet	Naiset	f % mie...	f % naiset	Tod. alaraja	Tod. yläraja	Luokkakeskus
0-9	132	103	0.53	0.42	0	10	5
10-19	129	63	0.52	0.26	10	20	15
20-29	376	121	1.52	0.5	20	30	25
30-39	519	161	2.09	0.66	30	40	35
40-49	1232	527	4.97	2.17	40	50	45
50-59	3228	1366	13.01	5.63	50	60	55
60-69	4565	2195	18.4	9.04	60	70	65
70-79	6598	4522	26.6	18.63	70	80	75
80-89	6399	9845	25.79	40.57	80	90	85
90-99	1600	5184	6.45	21.36	90	100	95
100-	31	181	0.12	0.75	100	110	105
Yht.	24809	24268					

Tehdään sarake suhteellisen frekvenssin ja luokkakeskusten tulolle.
 Kuoliniän keskiarvo on tämän tulosarakkeen summa.

Ikä (vuotta)	Miehet	Naiset	f%miehet	f%naiset	Luokkakeskus LK	f%miehet*LK	f%naiset*LK
0-9	132	103	0,0053	0,0042	5	0,03	0,02
10-19	129	63	0,0052	0,0026	15	0,08	0,04
20-29	376	121	0,0152	0,0050	25	0,38	0,12
30-39	519	161	0,0209	0,0066	35	0,73	0,23
40-49	1232	527	0,0497	0,0217	45	2,23	0,98
50-59	3228	1366	0,1301	0,0563	55	7,16	3,10
60-69	4565	2195	0,1840	0,0904	65	11,96	5,88
70-79	6598	4522	0,2660	0,1863	75	19,95	13,98
80-89	6399	9845	0,2579	0,4057	85	21,92	34,48
90-99	1600	5184	0,0645	0,2136	95	6,13	20,29
100-	31	181	0,0012	0,0075	105	0,13	0,78
	24809	24268				70,70	79,90

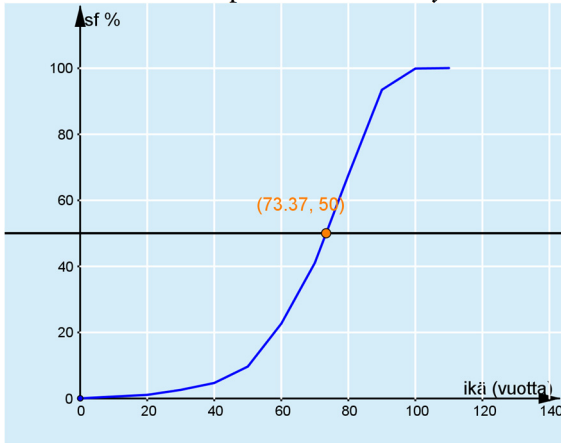
Miesten kuoliniän keskiarvo on noin 70,7 vuotta ja naisten
 noin 79,9 vuotta.

Mediaanin määrittämiseksi tehdään sarakkeet suhteellisille
 summafrekvensseille (sf %m ja sf %n).

Ikä (vuotta)	Miehet	Naiset	f%miehet	f%naiset	Luokkakeskus LK	f%miehet*LK	f%naiset*LK	sf%miehet	sf%naiset
0-9	132	103	0,0053	0,0042	5	0,03	0,02	0,53 %	0,42 %
10-19	129	63	0,0052	0,0026	15	0,08	0,04	1,05 %	0,68 %
20-29	376	121	0,0152	0,0050	25	0,38	0,12	2,57 %	1,18 %
30-39	519	161	0,0209	0,0066	35	0,73	0,23	4,66 %	1,85 %
40-49	1232	527	0,0497	0,0217	45	2,23	0,98	9,63 %	4,02 %
50-59	3228	1366	0,1301	0,0563	55	7,16	3,10	22,64 %	9,65 %
60-69	4565	2195	0,1840	0,0904	65	11,96	5,88	41,04 %	18,69 %
70-79	6598	4522	0,2660	0,1863	75	19,95	13,98	67,63 %	37,32 %
80-89	6399	9845	0,2579	0,4057	85	21,92	34,48	93,43 %	77,89 %
90-99	1600	5184	0,0645	0,2136	95	6,13	20,29	99,88 %	99,25 %
100-	31	181	0,0012	0,0075	105	0,13	0,78	100,00 %	100,00 %
	24809	24268				70,70	79,90		

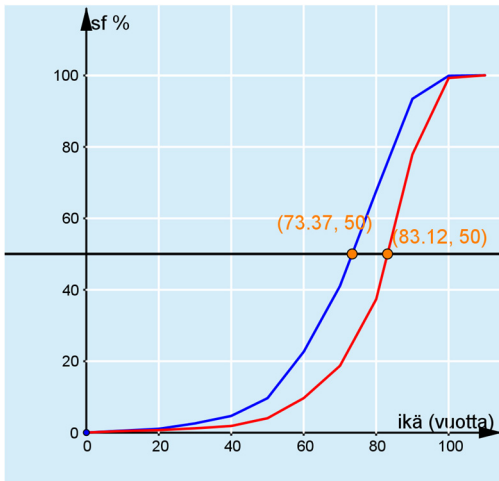
Piirretään miesten suhteellisista summafrekvensseistä (sf % m) kertymän kuvaaja. Mediaani määritetään kertymän kuvaajasta. Kuvaajan pisteiden x -koordinaatit ovat luokkien ylärajoja ja y -koordinaatit suhteellisia summafrekvenssejä.

Mediaani määritetään piirtämällä suora $y = 50$.



Miesten kuolinikien mediaani on noin 73,37 vuotta.

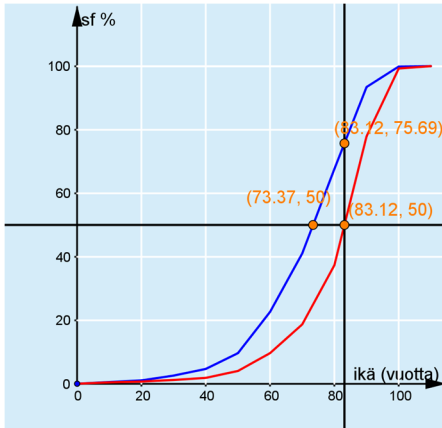
Määritetään naisten kuolinikien mediaani vastaavalla tavalla.



Naisten kuolinikien mediaani on noin 83,12 vuotta.

Vastaus: miehet: $\bar{x} \approx 70,7$ vuotta ja $M_d \approx 73,4$ vuotta, naiset: $\bar{x} \approx 79,9$ vuotta ja $M_d \approx 83,1$ vuotta

- b) Naisten kuoliniän mediaani oli 83,12 vuotta. On selvítettävä, kuinka monta prosenttia miehistä kuoli yli 83,12-vuotiaina. Piirretään x -akselille normaali pisteen $(83,12; 0,5)$ kautta ja määritetään normaalin ja miesten kuoliniän kertymän kuvaajan leikkauspiste.



Miehistä 76 % kuoli naisten kuoliniän mediaania nuorempina, joten 24 % kuoli kyseistä ikää vanhempina.

Vastaus: 24 %

465. Merkitään tuntematonta frekvenssiä x :llä ja muodostetaan keskiarvon avulla yhtälö. Ratkaistaan yhtälöstä x .

$$\frac{4 \cdot 0 + 9 \cdot 1 + 12 \cdot 2 + 11 \cdot 3 + x \cdot 4 + 5 \cdot 5 + 3 \cdot 6 + 1 \cdot 7}{4 + 9 + 12 + 11 + x + 5 + 3 + 1} = 2,72$$

$$\frac{4x + 116}{45 + x} = 2,72 \quad \| \cdot (45 + x)$$

$$4x + 116 = 2,72 \cdot 45 + 2,72x$$

$$4x + 116 = 122,4 + 2,72x$$

$$4x - 2,72x = 122,4 - 116$$

$$1,28x = 6,4 \quad \| : 1,28$$

$$x = 5$$

Tuntematon frekvenssi on 5.

Vastaus: 5

466. Merkitään oikeaa keskiarvoa kirjaimella x ja oppilaiden määrää kirjaimella n . Arvosanojen oikea summa on tällöin nx .

16 % oppilaista on $0,16n$. Koska $0,16n$ oppilasta sai yhden numeron liian pienen arvosanan, jäi summa $0,16n$ liian pieneksi.

6 % oppilaista on $0,06n$. Koska $0,06n$ oppilasta sai kaksi numeroa liian pienen arvosanan, jäi summa $0,06n \cdot 2 = 0,12n$ liian pieneksi.

Merkittyjen arvosanojen summa on $nx - 0,16n - 0,12n = nx - 0,28n$.

Keskiarvoksi tuli $\frac{nx - 0,28n}{n} = x - 0,28$. Keskiarvo jäi siis $0,28$ numeroa liian pieneksi.

Vastaus: $0,28$ numeroa alaspäin

467. a) Määritetään luokkakeskukset. Ensimmäisen luokan todellinen alaraja on 0 ja todellinen yläraja 2,5.

Jonotusaika (min)	Luokkakeskus (min)	Asiakkaiden lukumäärä yrityksessä A	Asiakkaiden lukumäärä yrityksessä B
0–2	$\frac{0 + 2,5}{2} = 1,25$	45	7
3–5	4	52	42
6–8	7	38	59
9–1	10	20	61
12–4	13	18	17
15–17	16	16	0
18–20	19	13	0

Puheluita tuli yrityksen A puhelinpalveluun yhteensä
 $45 + 52 + 38 + 20 + 18 + 16 + 13 = 202$

ja yrityksen B puhelinpalveluun
 $7 + 42 + 59 + 61 + 17 = 186$.

Yrityksen A jonotusaikojen keskiarvo on

$$\bar{x}_A = \frac{45 \cdot 1,25 + 52 \cdot 4 + 38 \cdot 7 + 20 \cdot 10 + 18 \cdot 13 + 16 \cdot 16 + 13 \cdot 19}{202} = 7,263...$$

Yrityksen B jonotusaikojen keskiarvo on

$$\bar{x}_B = \frac{7 \cdot 1,25 + 42 \cdot 4 + 59 \cdot 7 + 61 \cdot 10 + 17 \cdot 13}{186} = 7,638...$$

Minuuteiksi ja sekunneiksi muutettuna keskiarvot ovat ja 7 min 16 s ja 7 min 38 s.

Vastaus: A: 7 min 16 s; B: 7 min 38 s

- b)** Todennäköisyys, jolla yrityksen A puhelinpalveluun soittanut asiakas joutui odottamaan yli 15 minuuttia, oli

$$P(\text{odottaa yli 15 min}) = \frac{16+13}{202} = 0,143\dots$$

Todennäköisyys, jolla yrityksen B puhelinpalveluun soittanut asiakas joutui odottamaan yli 15 minuuttia, oli 0, koska näin ei käynyt kenellekään.

Vastaus: A: 0,14; B: 0

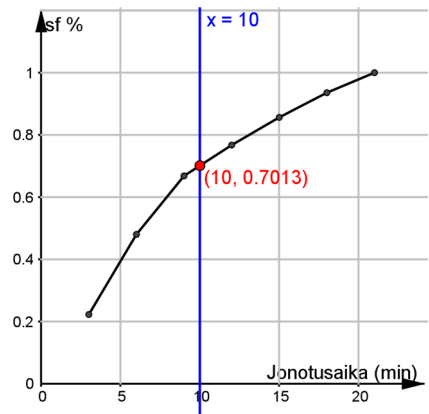
- c) Todennäköisyys, jolla asiakas joutui odottamaan yli 10 minuuttia, määritetään kertymän kuvaajasta. Kuvaajan pisteiden x -koordinaatit ovat luokkien todellisia ylärajoja ja y -koordinaatit niitä vastaavia suhteellisen summafrekvenssin arvoja. On otettava huomioon, että jonotusajat on pyöristetty alaspäin minuutin tarkkuuteen.

Tehdään sarake suhteelliselle summafrekvenssille. Määritetään kaksi ensimmäistä summafrekvenssiä soluviittausten avulla ja kopioidaan kaava seuraaville riveille.

Todellinen yläraja (min)	sf % A	Todellinen yläraja (min)	sf % B
3	0.2228	3	0.0376
6	0.4802	6	0.2634
9	0.6683	9	0.5806
12	0.7673	12	0.9086
15	0.8564	15	1
18	0.9356	18	1
21	1	21	1

Piirretään yrityksen A kertymän kuvaaja.

Todennäköisyys, jolla asiakas joutui odottamaan yli 10 minuuttia, saadaan piirtämällä suora $x = 10$ ja määrittämällä suoran viivakaaviokuvaajan leikkauspiste.

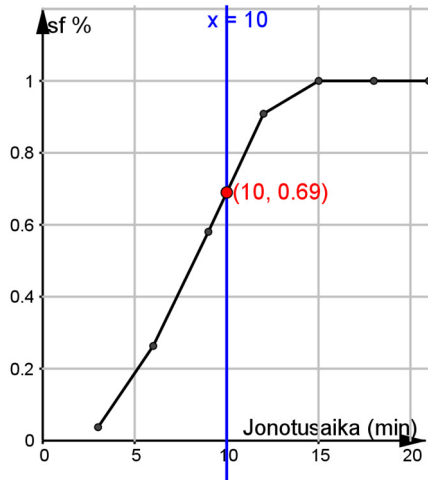


70,13 % asiakkaista jonotti alle 10 minuuttia, joten yli 10 minuuttia odotti $100 \% - 70,13 \% = 29,87 \% \approx 30 \%$ asiakkaista.

Määritetään vastaava todennäköisyys yrityksen B puhelinpalvelulle vastaavalla tavalla.

69 % asiakkaista jonotti alle 10 minuuttia, joten yli 10 minuuttia odotti
 $100 \% - 69 \% = 31 \%$
 asiakkaista.

Vastaus: A: 0,30; B: 0,31



- d) Ne 9 minuuttia jonottaneet asiakkaat, jotka joutuivat odottamaan vielä vähintään 6 minuuttia, odottivat yhteensä vähintään 15 minuuttia. Heidän osuutensa 9 minuuttia jonottaneista asiakkaista oli yrityksen A puhelinpalvelussa

$$\frac{16 + 13}{20 + 18 + 16 + 13} = 0,432...$$

ja yrityksen B puhelinpalvelussa 0.

Vastaus: A: 0,43; B: 0

468. a) Taulukoidaan arvosanat painokertoimineen. MAB2-kurssin painokerroin on 2, MAB3:n 3 jne.

Arvosana	Painokerroin
9	2
8	3
7	4
5	5
6	6

$$\text{Arvosanaksi tulee } \frac{9 \cdot 2 + 8 \cdot 3 + 7 \cdot 4 + 5 \cdot 5 + 6 \cdot 6}{2 + 3 + 4 + 5 + 6} = \frac{131}{20} = 6,55 \approx 7.$$

Vastaus: arvosanan 7

- b) Jos Olli haluaa nostaa päättöarvosanaansa, hänen on saatava keskiarvoksi vähintään 7,5.

Ratkaistaan yhtälön avulla, mikä arvosana x Ollin on vähintään saatava kurssista MAB7, jos hän käy vain kurssin MAB7.

$$\frac{131 + 7x}{20 + 7} = 7,5$$

$$\frac{131 + 7x}{27} = 7,5 \quad \parallel \cdot 27$$

$$131 + 7x = 202,5$$

$$7x = 202,5 - 131$$

$$7x = 71,5 \quad \parallel : 7$$

$$x = 10,214...$$

Olli ei voi saada yli arvosanaa 10, joten tämä vaihtoehto on mahdoton.

Tutkitaan vastaavasti, mikä kurssin MAB8 arvosanan pitäisi vähintään olla, jos Olli käy vain kurssin MAB8.

$$\frac{131 + 8x}{20 + 8} = 7,5$$

$$\frac{131 + 8x}{28} = 7,5 \quad \parallel \cdot 28$$

$$131 + 8x = 210$$

$$8x = 210 - 131$$

$$8x = 79 \quad \parallel : 8$$

$$x = 9,875$$

Jos Olli käy vain kurssin MAB8, niin hänen on saatava siitä arvosana 10.

Tutkitaan taulukkolaskentaohjelmalla, mikä arvosanojen painotetuksi keskiarvoksi tulee kaikilla kurssien MAB7 ja MAB8 arvosanojen yhdistelmillä.

Täytetään taulukkoon jo olemassa olevat arvosanat ja asetetaan kurssien MAB7 ja MAB8 arvosanoiksi 4. Lasketaan arvosanojen ja painokertoimien tulot sekä tulojen summa.

	A	B	C	D	E	F	G	H	I
1									
2	Kurssin nro	2	3	4	5	6	7	8	
3	Arvosana	9	8	7	5	6	4	4	
4	Tulo	18	24	28	25	36	28	32	=SUMMA(B4:H4)

Painokertoimien summa on
 $1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 = 36$.

Painotettu keskiarvo saadaan jakamalla tulojen summa painokertoimien summalla.

	A	B	C	D	E	F	G	H	I	J
1										
2	Kurssin nro	2	3	4	5	6	7	8		
3	Arvosana	9	8	7	5	6	4	4		
4	Tulo	18	24	28	25	36	28	32	191	=I4/36

Painotetuksi keskiarvoksi tulee nyt 5,31.

	A	B	C	D	E	F	G	H	I	J
1										
2	Kurssin nro	2	3	4	5	6	7	8		
3	Arvosana	9	8	7	5	6	4	4		
4	Tulo	18	24	28	25	36	28	32	191	5.31

Muutteleamalla kurssien 7 ja 8 arvosanoja löydetään kaikki mahdollisuudet, joissa painotettu keskiarvo on vähintään 7,5.

	A	B	C	D	E	F	G	H	I	J
1										
2	Kurssin nro	2	3	4	5	6	7	8		
3	Arvosana	9	8	7	5	6	10	9		
4	Tulo	18	24	28	25	36	70	72	273	7.58

Muodostetaan taulukko kaikista arvosanayhdistelmistä, joilla Ollin päättöarvosana nouse yhdellä numerolla.

MAB7 arvosana	8	10	9	10	10
MAB8 arvosana	10	8	10	9	10

Vastaus:

MAB7 arvosana	8	10	9	10	10
MAB8 arvosana	10	8	10	9	10

ALOITUSAUKEAMAAN LIITTYVIÄ TEHTÄVIÄ

1. Keskiarvoksi saadaan taulukkolaskentaohjelmalla 47 163,25 dollaria asukasta kohti.

Vastaus: 47 163,25 \$

2. Keskiarvoksi Luokitellaan aineisto LibreOfficen TAAJUUS-komennon avulla.

LibreOffice Calc screenshot showing the TAAJUUS function dialog box. The dialog box is titled "Oletettu funktion luonti" and shows the function "TAAJUUS" with arguments "data" (B2:B51) and "classes" (F3:F8). The function result is shown as an array: {0;37;8;2;2;1;0}.

G3:G9								
	A	B	C	D	E	F	G	H
1	Country Nam	GDP per capita (current international \$) YR2015						
2	Qatar	141542,6628				Yläraja	f	
3	Macao SAR,	111496,6294				25000	0	
4	Luxembourg	101926,4236				50000	37	
5	Singapore	85382,30227				75000	8	
6	Brunei Daru	78369,29102				100000	2	
7	Kuwait	74645,48246				125000	2	
8	United Arab	69970,81534				150000	1	
9	Ireland	65144,1137					0	
10	Norway	61196,68819						
11	Switzerland	61086,32681						

Luokiteltu aineisto

Asukaskohtainen bruttokansantuote (\$)	f
0 – 25 000	0
26 000 – 50 000	37
51 000 – 75 000	8
76 000 – 100 000	2
101 000 – 125 000	2
126 000 – 150 000	1

Vastaus:

Asukaskohtainen bruttokansantuote (\$)	f
0 – 25 000	0
26 000 – 50 000	37
51 000 – 75 000	8
76 000 – 100 000	2
101 000 – 125 000	2
126 000 – 150 000	1