1 Tilastot

1.1 Tilastotieteen peruskäsitteitä

1.

a) Koska tutkitaan eniten myydyn kahvimerkin hintaa eri puolilla Suomea, perusjoukkona (eli kaikkien havaintoyksiköiden joukkona) on Suomen eniten myytyjen kahvipakettien joukko.

b) Tutkimuksen kohteena, eli havaintoyksikkönä ovat ne kahvipaketit, jotka on valittu tutkimukseen. Havaintoyksiköt ovat kahvipaketteja.

c) Tutkittava ominaisuus, eli muuttuja on eniten myydyn kahvimerkin hinta euroina per kilo.

d) Koska tutkimusta varten ei ole kysytty kaikkien Suomessa valintamyymälöissä myynnissä olevien kahvipaketin hintoja, havaintoaineisto on otos.

2.

a) Koska tutkitaan yleistä arvonlisäverokantaa EU-maissa, perusjoukkona ovat kaikki EU-maat.

b) Tutkimuksen kohteena, eli havaintoyksikkönä on EU-maa.

c) Tutkittava ominaisuus, eli muuttuja on yleinen arvonlisäverokanta prosentteina tuotteen tai palvelun verottomasta hinnasta.

d) Koska tutkimusta varten tiedot saatiin jokaisen EU-maan verohallinnolta, havaintoaineisto on kokonaisaineisto.

3.

a) Koska tutkitaan talitiaispoikueiden kokoa luonnonpuistossa, perusjoukko on kaikkien kyseisen luonnonpuiston talitiaispoikueiden joukko.

b) Tutkimuksen kohteena, eli havaintoyksikkönä ovat ne talitiaispoikueet, jotka on valittu tutkimukseen. Havaintoyksiköt ovat talitiaispoikueita.

c) Tutkittava ominaisuus, eli muuttuja on talitiaispoikueen koko munien lukumääränä ilmoitettuna.

d) Koska muuttujan arvo, eli munien lukumäärä voi saada vain tiettyjä kokonaislukuarvoja, on muuttuja diskreetti.

e) Koska tutkimusta varten tietoa ei kerätty kaikista luonnonpuiston talitiaispoikueista, havaintoaineisto on otos.

4.

a) Koska valtion pinta-ala voi saada mitä tahansa positiivisia arvoja, se on jatkuva muuttuja.

b) Koska asuinhuoneiston huoneiden lukumäärä voi saada vain kokonaislukuarvoja, se on diskreetti muuttuja.

c) Koska auton huippunopeus voi saada mitä tahansa positiivisia arvoja, se on jatkuva muuttuja.

d) Koska henkilön yleiskunto voi tällä mittarilla mitattuna saada ainoastaan tietyt arvot 1, 2, 3 tai 4, se on diskreetti muuttuja.

e) Koska yli miljoonan asukkaan kaupunkien lukumäärä voi eri valtioissa saada vain kokonaislukuarvoja, se on diskreetti muuttuja.

5.

Summafrekvenssi lasketaan frekvenssien avulla summaamalla yhteen siihenastiset frekvenssit. Näin saadaan kaksi ensimmäistä summafrekvenssiä.

Kolmannen muuttujan arvon summafrekvenssi on 43. Tämän avulla saadaan laskettua muuttujan arvoa 3 vastaava frekvenssi:

$$43 - (7 + 12) = 43 - 19 = 24.$$

Tämän jälkeen saadaan neljännen muuttujan arvon 4 summafrekvenssi:

$$7 + 12 + 24 + 5 = 43 + 5 = 48.$$

Viimeisen muuttujan arvon summafrekvenssi on 50. Tämän avulla saadaan laskettua muuttujan arvoa 5 vastaava frekvenssi:

$$50 - (7 + 12 + 24 + 5) = 50 - 48 = 2.$$

x	***f***	***f %***	***sf***	***sf %***
1	7		7	
2	12		7 + 12 = 19	
3	24		19 + 24 = 43	
4	5		43 + 5 = 48	
5	2		48 + 2 = 50	

Kun frekvenssisarake on saatu valmiiksi, voidaan täydentää taulukon kolmas sarake, eli laskea suhteelliset frekvenssit. Verrataan muuttujan frekvenssiä havaintoyksiköiden kokonaismäärään, joka on 50, ja ilmoitetaan osamäärä prosentteina.

$\boldsymbol{x}$	$\boldsymbol{f}$	$\boldsymbol{f\,\%}$
1	7	$\frac{7}{50}\cdot 100 = \mathbf{14}$
2	12	$\frac{12}{50}\cdot 100 = \mathbf{24}$
3	24	$\frac{24}{50}\cdot 100 = \mathbf{48}$
4	5	$\frac{5}{50}\cdot 100 = \mathbf{10}$
5	2	$\frac{2}{50}\cdot 100 = \mathbf{4}$

Viimeinen sarake, eli suhteellinen summafrekvenssi lasketaan vertaamalla summafrekvenssiä kokonaismäärään, ja ilmoittamalla osamäärä prosentteina.

x	***f***	***f* %**	***sf***	***sf* %**
1	7	14	7	$\frac{7}{50} \cdot 100 = 14$
2	12	24	19	$\frac{19}{50} \cdot 100 = 38$
3	24	48	43	$\frac{43}{50} \cdot 100 = 86$
4	5	10	48	$\frac{48}{50} \cdot 100 = 96$
5	2	4	50	$\frac{50}{50} \cdot 100 = 100$

6.

Summafrekvenssi, eli taulukon neljäs sarake, lasketaan frekvenssien avulla summaamalla yhteen siihenastiset frekvenssit.

Havaintoyksiköiden kokonaismäärää on kaikkien frekvenssien summa, eli 100.

Suhteelliset frekvenssit, eli taulukon kolmas sarake, saadaan vertaamalla muuttujan frekvenssiä havaintoyksiköiden kokonaismäärään ja ilmoittamalla osamäärä prosentteina. Koska kokonaismäärä on 100, ovat frekvenssi ja suhteellinen frekvenssi sama luku.

Suhteelliset summafrekvenssit, eli taulukon viimeinen sarake, saadaan vertaamalla summafrekvenssiä kokonaismäärään ja ilmoittamalla osamäärä prosentteina.

x	***f***	***f* %**	***sf***	***sf* %**
2	6	$\frac{6}{100} \cdot 100 = 6$	6	$\frac{6}{100} \cdot 100 = 6$
3	10	$\frac{10}{100} \cdot 100 = 10$	$6 + 10 = 16$	$\frac{16}{100} \cdot 100 = 16$
4	22	$\frac{22}{100} \cdot 100 = 22$	$16 + 22 = 38$	$\frac{38}{100} \cdot 100 = 38$
5	31	$\frac{31}{100} \cdot 100 = 31$	$38 + 31 = 69$	$\frac{69}{100} \cdot 100 = 69$
6	26	$\frac{26}{100} \cdot 100 = 26$	$69 + 26 = 95$	$\frac{95}{100} \cdot 100 = 95$
7	5	$\frac{5}{100} \cdot 100 = 5$	$95 + 5 = 100$	$\frac{100}{100} \cdot 100 = 100$

7.

Ensimmäinen frekvenssi on 2 ja sen suhteellinen frekvenssi on 10. Lukumäärä 2 on siis 10 % havaintoyksiköiden kokonaismäärästä. Kokonaismäärä on näin ollen 20.

Summafrekvenssi lasketaan summaamalla siihenastiset frekvenssit. Havaintoyksiköiden kokonaismäärä on 20, eli viimeinen summafrekvenssi on **20**.

Toinen frekvenssi saadaan vastaavasta summafrekvenssistä: $6-2=\mathbf{4}$.

Suhteellinen frekvenssi saadaan vertaamalla frekvenssiä kokonaismäärään:

$$\frac{4}{20}\cdot 100=\mathbf{20}.$$

Suhteellinen summafrekvenssi lasketaan summaamalla siihenastiset suhteelliset frekvenssit. Kolmas suhteellinen frekvenssi saadaan vastaavasta summafrekvenssistä: $80-30=\mathbf{50}$. Kolmas frekvenssi on siis 50 % havaintoyksiköiden kokonaismäärästä. Kokonaismäärä on 20, joten kolmas frekvenssi on $0{,}5\cdot 20=\mathbf{10}$.

Lopulta täydennetään taulukot vielä puuttuvat frekvenssit:

f	***f %***	***sf***	***sf %***
2	**10**	2	10
4	**20**	**6**	$10+20=30$
10	**50**	$6+10=16$	**80**
$20-16=4$	$\frac{4}{20}\cdot 100=20$	**20**	$\frac{20}{20}\cdot 100=100$

8.

a) Opiskelijat, jotka eivät myöhästyneet kertaakaan, kuuluvat ”nolla kertaa myöhästyneisiin”. Näiden prosenttiosuus nähdään muuttujan arvon 0 suhteellisesta frekvenssistä, joka on 20 %.

b) Kaksi kertaa myöhästyneiden lukumäärä nähdään muuttujan arvon 2 frekvenssistä, joka on 6 (oppilasta).

c) Korkeintaan kahtena aamuna myöhästyneiden prosenttiosuus nähdään muuttujan arvon 2 suhteellisesta summafrekvenssistä, joka on 76 %.

d) Enintään kolmena aamuna myöhästyneiden määrä nähdään muuttujan arvon 3 summafrekvenssistä, joka on 22 (oppilasta).

e) Vähintään neljänä aamuna myöhästyneisiin kuuluvat 4 tai 5 kertaa myöhästyneet. Näitä oli prosentteina yhteensä $8\ \% + 4\ \% = 12\ \%$.

Tämä voidaan laskea myös kertymien avulla. Kaikkien oppilaiden määrä prosentteina on 100 %. Enintään kolmena aamuna myöhästyneiden prosenttiosuus nähdään muuttujan arvon 3 suhteellisesta summafrekvenssistä, joka on 88 %. Vähintään neljänä aamuna myöhästyneitä on silloin $100\ \% - 88\ \% = 12\ \%$.

9.

a) Kolme kertaa viikossa hiuksensa pesevien lukumäärä nähdään muuttujan arvon 3 frekvenssistä, joka on 78 (henkilöä).

b) Korkeintaan neljä kertaa viikossa hiuksensa pesevien lukumäärä nähdään muuttujan arvon 4 summafrekvenssistä, joka on 186 (henkilöä).

c) Korkeintaan viisi kertaa viikossa hiuksensa pesevien prosenttiosuus nähdään muuttujan arvon 5 suhteellisesta summafrekvenssistä, joka on 88 %.

d) Ainakin kuusi kertaa viikossa hiuksensa peseviin kuuluvat 6 kertaa pesevät sekä 7 kertaa pesevät. Näitä oli prosentteina yhteensä
$8\ \% + 4\ \% = 12\ \%$.

Tämä voidaan laskea myös kertymien avulla. Kaikkien tutkimukseen osallistuneiden määrä prosentteina on 100 %. Korkeintaan viisi kertaa viikossa pesevien prosenttiosuus laskettiin kohdassa c: näitä on 88 %. Ainakin kuusi kertaa viikossa hiuksensa peseviä on silloin
$100\ \% - 88\ \% = 12\ \%$.

10.

a) Opiskelijoiden lukumäärä, eli havaintoyksiköiden kokonaismäärä, saadaan laskemalla frekvenssit yhteen:

$$1 + 3 + 4 + 8 + 5 + 3 + 1 = 25.$$

Suhteelliset frekvenssit saadaan vertaamalla frekvenssiä kokonaismäärään, ja ilmaisemalla osamäärä prosentteina.

Summafrekvenssi lasketaan summaamalla yhteen siihenastiset frekvenssit.

Suhteelliset summafrekvenssit saadaan vertaamalla summafrekvenssiä kokonaismäärään, ja ilmaisemalla osamäärä prosentteina. Suhteellinen summafrekvenssi saadaan myös summaamalla siihenastiset suhteelliset frekvenssit, kuten alla taulukossa on tehty.

Arvosana	***f***	***f* %**	***sf***	***sf* %**
4	1	$\frac{1}{25} \cdot 100 = 4$	1	4
5	3	$\frac{3}{25} \cdot 100 = 12$	$1 + 3 = 4$	$4 + 12 = 16$
6	4	$\frac{4}{25} \cdot 100 = 16$	$4 + 4 = 8$	$16 + 16 = 32$
7	8	$\frac{8}{25} \cdot 100 = 32$	$8 + 8 = 16$	$32 + 32 = 64$
8	5	$\frac{5}{25} \cdot 100 = 20$	$16 + 5 = 21$	$64 + 20 = 84$
9	3	$\frac{3}{25} \cdot 100 = 12$	$21 + 3 = 24$	$84 + 12 = 96$
10	1	$\frac{1}{25} \cdot 100 = 4$	$24 + 1 = 25$	$94 + 4 = 100$

b) Arvosanaksi korkeintaan kahdeksan saaneiden lukumäärä nähdään muuttujan arvon 8 summafrekvenssistä, joka on 21 (opiskelijaa).

c) Arvosanaksi vähintään seitsemän saaneisiin kuuluvat arvosanan 7, 8, 9 ja 10 saaneet opiskelijat. Näiden prosenttiosuudet on laskettu suhteellisten frekvenssien sarakkeeseen.
Prosenttiosuudet ovat yhteensä

$$32\ \% + 20\ \% + 12\ \% + 4\ \% = 68\ \%.$$

Tämä voidaan laskea myös kertymien avulla. Kaikkien opiskelijoiden määrä on 100 %. Korkeintaan arvosanan 6 saaneiden prosenttiosuus nähdään muuttujan arvon 6 suhteellisesta summafrekvenssistä, joka on 32 %. Arvosanaksi vähintään seitsemän saaneita oli silloin

$$100\ \% - 32\ \% = 68\ \%.$$

d) Arvosanan 7 tai 8 saaneiden prosenttiosuudet nähdään muuttujan arvojen 7 ja 8 suhteellisista frekvensseistä, jotka ovat yhteensä

$$32\ \% + 20\ \% = 52\ \%.$$

11.

a) Määritetään eri heittotulosten frekvenssit:

Heittotulos	**4**	**7**	**8**	**9**	**10**
f	2	4	1	2	1

Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Heittotuloksia 7 on eniten (neljä kappaletta) eli heittotulos 7 on muuttujan arvoista yleisin. Moodi Mo = 7.

b) Mediaani on suuruusjärjestykseen laitetun aineiston keskimmäinen luku. Järjestetään heittotulokset suuruusjärjestykseen:

4 4 7 7 **7** **7** 8 9 9 10

Heittotuloksia on parillinen määrä (10 kappaletta), joten aineistossa ei ole keskimmäistä lukua. Mediaani on kahden keskimmäisen luvun keskiarvo:

Mediaani Md $= \frac{7+7}{2} = \frac{14}{2} = 7$.

c) Lasketaan heittotulosten keskiarvo jakamalla heittotulosten summa lukumäärällä eli luvulla 10. Summan laskemisessa voidaan hyödyntää frekvenssiä, joka kertoo muuttujan arvojen lukumäärän. Keskiarvo on

$$\bar{x} = \frac{2 \cdot\ 4 + 4 \cdot 7 + 8 + 2 \cdot 9 + 10}{10} = \frac{72}{10} = 7{,}2.$$

12.

a) Järjestetään aineisto frekvenssitaulukoksi:

Tölkkien kulutus	***f***	***f* %**	***sf* %**
0	2	$\frac{2}{20} = 0{,}1 = 10\%$	10
1	6	$\frac{6}{20} = 0{,}3 = 30\%$	$10 + 30 = 40$
2	6	$\frac{6}{20} = 0{,}3 = 30\%$	$40 + 30 = 70$
3	2	$\frac{2}{20} = 0{,}1 = 10\%$	$70 + 10 = 80$
4	2	$\frac{2}{20} = 0{,}1 = 10\%$	$80 + 10 = 90$
5	2	$\frac{2}{20} = 0{,}1 = 10\%$	$90 + 10 = 100$

Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Suurin frekvenssi on 6 ja se on arvoilla 1 ja 2. Moodeja on tässä aineistossa kaksi: moodi Mo = 1 (tölkki) ja Mo =2 (tölkkiä).

Mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää 50 % eli Md = 2 (tölkkiä).

Koska aineisto on kohtuullisen pieni, mediaani voidaan määrittää myös järjestämällä muuttujan arvot suuruusjärjestykseen:

0 0 1 1 1 1 1 1 2 **2** **2** 2 2 2 3 3 4 4 5 5

Vastauksia on parillinen määrä (20 kappaletta), joten järjestetyssä aineistossa ei ole keskimmäistä lukua. Mediaani on kahden keskimmäisen luvun keskiarvo:

Mediaani $\mathrm{Md} = \frac{2+2}{2} = \frac{4}{2} = 2$ (tölkkiä).

b) Lasketaan tölkkien kulutuksen keskiarvo. Keskiarvo lasketaan jakamalla summa lukumäärällä eli luvulla 20. Summan laskemisessa voidaan hyödyntää frekvenssejä, jotka kertovat muuttujan arvojen lukumäärän.

Keskiarvo on

$$\bar{x} = \frac{2 \cdot 0 + 6 \cdot 1 + 6 \cdot 2 + 2 \cdot 3 + 2 \cdot 4 + 2 \cdot 5}{20} = \frac{42}{20} = 2{,}1.$$

Kyselyn perusteella juoman kulutus on keskimäärin 2,1 tölkkiä opiskelijaa kohden. Tämän perusteella automaatti hankitaan.

13.

Lasketaan liikevoiton keskiarvo:

$$\bar{x} = \frac{8 + 8 + 10 + 12 + 11 + 11 + 10 + 10}{8} = \frac{80}{8} = 10.$$

Keskimääräinen vuotuinen liikevoitto vuosina 2008–2015 on siis 10 miljoonaa euroa.

14.

Luvun -1 vastaluku on $-(-1) = 1$ ja luvun 5 käänteisluku on $\frac{1}{5}$.

Lukujen 1 ja $\frac{1}{5}$ summa on

$$1 + \frac{1}{5} = \frac{5}{5} + \frac{1}{5} = \frac{5+1}{5} = \frac{6}{5}.$$

Keskiarvo lasketaan jakamalla summa lukumäärällä:

$$\frac{\frac{6}{5}}{2} = \frac{3}{5}.$$

Kysytty keskiarvo on siis $\frac{3}{5}$.

15.

a) Moodi on se muuttujan arvo, jonka frekvenssi on suurin eli tässä aineistossa Mo = 1 (elokuva).

Mediaanin määrittämistä varten lasketaan frekvenssitaulukkoon suhteelliset summafrekvenssit:

Katsottujen elokuvien määrä	***f***	***f* %**	***sf* %**
0	6	12	12
1	15	30	$12 + 30 = 42$
2	14	28	$42 + 28 = \mathbf{70}$
3	8	16	$70 + 16 = 86$
4	5	10	$86 + 10 = 96$
5	2	4	$96 + 4 = 100$

Mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää 50 % eli tässä aineistossa Md = 2 (elokuvaa).

b) Lasketaan keskiarvo jakamalla elokuvien lukumäärän summa vastaajien lukumäärällä eli luvulla 50:

$$\frac{6 \cdot 0 + 15 \cdot 1 + 14 \cdot 2 + 8 \cdot 3 + 5 \cdot 4 + 2 \cdot 5}{50} = \frac{97}{50} = 1{,}94 \approx 1{,}9.$$

Keskiarvo on yhden desimaalin tarkkuudella $\bar{x} \approx 1{,}9$ (elokuvaa).

16.

Täydennetään frekvenssitaulukkoon suhteelliset frekvenssit ja suhteelliset summafrekvenssit:

Tulos	***f***	***f*%**	***sf*%**
1	10	10	10
2	10	10	20
3	10	10	30
4	25	25	**55**
5	15	15	70
6	**30**	30	100

Moodi on se muuttujan arvo, jonka frekvenssi on suurin, eli Mo = 6.

Mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää 50 %, eli Md = 4.

Mediaani voidaan päätellä myös järjestämällä heittotulokset suuruusjärjestykseen ja etsimällä keskimmäisin arvo eli se arvo, joka jakaa havaintoyksiköt kahteen yhtä suureen osaan.

Keskiarvo lasketaan jakamalla muuttujan arvojen summa arvojen lukumäärällä:

$$\frac{10 \cdot 1 + 10 \cdot 2 + 10 \cdot 3 + 25 \cdot 4 + 15 \cdot 5 + 30 \cdot 6}{100} = \frac{415}{100} = 4{,}15.$$

Keskiarvo on siis $\bar{x} = 4{,}15$.

17.

Havaintoyksiköiden lukumäärä on frekvenssien summa eli

$$12 + 18 + 11 + 6 + 3 = 50.$$

Täydennetään frekvenssitaulukkoon suhteelliset frekvenssit ja suhteelliset summafrekvenssit:

Terälehtiä	***f***	***f* %**	***sf* %**
4	12	$\frac{12}{50} = 0{,}24 = 24\ \%$	24
5	**18**	$\frac{18}{50} = 0{,}36 = 36\ \%$	$24 + 36 = \mathbf{60}$
6	11	$\frac{11}{50} = 0{,}22 = 22\ \%$	$60 + 22 = 82$
7	6	$\frac{6}{50} = 0{,}12 = 12\ \%$	$82 + 12 = 94$
8	3	$\frac{3}{50} = 0{,}06 = 6\ \%$	$94 + 6 = 100$

Moodi on se muuttujan arvo, jonka frekvenssi on suurin, eli Mo = 5 (terälehteä).

Mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää 50 %, eli Md = 5 (terälehteä).

Mediaani voidaan päätellä myös järjestämällä aineisto suuruusjärjestykseen ja etsimällä se muuttujan arvo, joka jakaa havainnot kahteen yhtä suureen osaan.

b) Lasketaan terälehtien keskiarvo otoksessa. Keskiarvo lasketaan jakamalla muuttujan arvojen summa arvojen lukumäärällä:

$$\frac{12 \cdot 4 + 18 \cdot 5 + 11 \cdot 6 + 6 \cdot 7 + 3 \cdot 8}{50} = \frac{270}{50} = 5{,}4.$$

Keskiarvo on siis $\bar{x} = 5{,}4$ (terälehteä).

18.

a) Merkitään Hannan viimeisen pakollisen kurssin arvosanaa kirjaimella *x.*

Kaikkien kurssien arvosanojen summa on tällöin

$$8 + 6 + 7 + 8 + 6 + x = x + 35$$

ja arvosanojen keskiarvo on $\frac{x+35}{6}$.

Tiedetään, että keskiarvo on 7,0 joten saadaan yhtälö:

$$\frac{x + 35}{6} = 7{,}0 \quad | \cdot 6$$

$$x + 35 = 42$$

$$x = 42 - 35 = 7$$

Viimeisen pakollisen kurssin arvosana oli siis 7.

b) Merkitään Iiriksen kahden viimeisen pakollisen kurssin arvosanoja kirjaimella x.

Kaikkien kurssien arvosanojen summa on tällöin

$$8 + 6 + 7 + 8 + x + x = 2x + 29$$

ja arvosanojen keskiarvo on $\frac{2x+29}{6}$.

Tiedetään, että keskiarvo on 6,5 joten saadaan yhtälö:

$$\frac{2x + 29}{6} = 6{,}5 \quad | \cdot 6$$

$$2x + 29 = 39$$

$$2x = 10$$

$$x = \frac{10}{2} = 5.$$

Kahden viimeisen pakollisen kurssin arvosana oli siis 5.

19.

a) Merkitään Laurin neljännellä ja viidennellä heittokerralla saamaa silmälukua kirjaimella *x.*

Kaikkien heittojen silmälukujen summa on tällöin

$$4 + 6 + 5 + x + x = 2x + 15$$

ja silmälukujen keskiarvo on $\frac{2x+15}{5}$.

Tiedetään, että keskiarvo on 3,8 joten saadaan yhtälö:

$$\frac{2x + 15}{5} = 3{,}8 \quad | \cdot 5$$

$$2x + 15 = 19$$

$$2x = 4$$

$$x = \frac{4}{2} = 2.$$

Neljännen ja viidennen heiton silmäluku on siis 2.

b) Merkitään Laurin neljännellä heittokerralla saamaa silmälukua kirjaimella *x.*

Lauri saa viidennellä heittokerralla yhtä suuremman silmäluvun, joten viidennen heittokerran silmäluku on $x + 1$.

Kaikkien heittojen silmälukujen summa on tällöin

$$4 + 6 + 5 + x + (x + 1) = 2x + 16$$

ja silmälukujen keskiarvo on $\frac{2x+16}{5}$.

Tiedetään, että keskiarvo on 4,4 joten saadaan yhtälö:

$$\frac{2x + 16}{5} = 4{,}4 \quad | \cdot 5$$

$$2x + 16 = 22$$

$$2x = 6$$

$$x = \frac{6}{2} = 3.$$

Neljännen heiton silmäluku on siis 3 ja viidennen heiton silmäluku on

$3 + 1 = 4.$

20.

Seitsemän omenan keskimääräinen massa on 150 g, eli omenien massan keskiarvo on 150 g. Keskiarvo lasketaan jakamalla omien massan summa niiden lukumäärällä eli luvulla 7. Omenien yhteenlaskettu massa on siis $150 \text{ g} \cdot 7 = 1050 \text{ g}$.

Kaikkien hedelmien yhteenlaskettu massa on mandariinien massan ja omenien massan summa eli 250 g +1050 g = 1300 g.

Lasketaan kaikkien hedelmien massan keskiarvo. Keskiarvo lasketaan jakamalla kaikkien hedelmien yhteenlaskettu massa niiden lukumäärällä. Keskiarvo on

$$\frac{1300 \text{ g}}{10} = 130 \text{ g}.$$

Kaikkien hedelmien keskimääräinen massa on siis 130 grammaa.

21.

Merkitään kyseistä lukua kirjaimella *x.* Luvun neliö on silloin x^2.
Näiden keskiarvo on $\frac{x+x^2}{2}$.

Tiedetään, että keskiarvo on 10, joten saadaan yhtälö:

$$\frac{x+x^2}{2} = 10 \quad | \cdot 2$$

$$x + x^2 = 20$$

$$x^2 + x - 20 = 0$$

Yhtälö ratkaistaan toisen asteen yhtälön ratkaisukaavalla:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

Tässä tehtävässä $a = 1, b = 1, c = -20$.

$$x = \frac{-1 \pm \sqrt{1^2 - 4 \cdot 1 \cdot (-20)}}{2 \cdot 1}$$

$$= \frac{-1 \pm \sqrt{1 + 80}}{2}$$

$$= \frac{-1 \pm \sqrt{81}}{2}$$

$$= \frac{-1 \pm 9}{2}.$$

Tehtävällä on kaksi ratkaisua:

$$x = \frac{-1 + 9}{2} = \frac{8}{2} = 4$$

tai

$$x = \frac{-1 - 9}{2} = \frac{-10}{2} = -5.$$

Kyseinen luku on siis joko 4 tai -5.

22.

a) Koska tutkitaan kaikkien ensimmäisen vuoden opiskelijoiden kemian arvosanojen jakautumista, perusjoukkona on kaikkien kemian pakollisella kurssilla olevien ensimmäisen vuoden opiskelijoiden joukko.

b) Tutkimuksen kohteena, eli havaintoyksikkönä on kemian pakollisen kurssin ensimmäisen vuoden opiskelija.

c) Tutkittava ominaisuus, eli muuttuja on kemian arvosana.

d) Koska muuttujan arvo, eli kemian arvosana, voi saada vain tiettyjä kokonaislukuarvoja (arvot 4–10), on muuttuja diskreetti.

e) Koska tutkimusta varten tieto kerättiin kaikilta opiskelijoilta, havaintoaineisto on kokonaisaineisto.

23.

Viimeisen muuttujan arvon 14 summafrekvenssi on 20, joten havaintoyksiköitä on yhteensä 20.

Muuttujan arvon 11 frekvenssi saadaan vähentämällä kokonaismäärästä muiden muuttujan arvojen frekvenssit:

$$20 - (2 + 5 + 8 + 4) = 20 - 19 = \mathbf{1}.$$

Tämän jälkeen suhteelliset frekvenssit saadaan vertaamalla muuttujan frekvenssiä kokonaismäärään, eli lukuun 20, ja ilmaisemalla osamäärä prosentteina.

Summafrekvenssi lasketaan summaamalla siihenastiset frekvenssit.

Suhteellinen summafrekvenssi lasketaan summaamalla siihenastiset suhteelliset frekvenssit.

x	f	$f\,\%$	sf	$sf\,\%$
10	**2**	$\frac{2}{20} \cdot 100 = 10$	2	10
11	**1**	$\frac{1}{20} \cdot 100 = 5$	$2 + 1 = 3$	$10 + 5 = 15$
12	**5**	$\frac{5}{20} \cdot 100 = 25$	$3 + 5 = 8$	$15 + 25 = 40$
13	**8**	$\frac{8}{20} \cdot 100 = 40$	$8 + 8 = 16$	$40 + 40 = 80$
14	**4**	$\frac{4}{20} \cdot 100 = 20$	**20**	$80 + 20 = 100$

24.

Havaintoyksiköiden lukumäärä on 100.

a) Neljä matkaa tekevien lukumäärä nähdään muuttujan arvon 4 frekvenssistä, joka on 12. Prosenttiosuus saadaan vertaamalla frekvenssiä kokonaismäärään:

$$\frac{12}{100} = 0{,}12 = 12\ \%.$$

Siis, 12 % henkilöistä tekee neljä matkaa.

b) Korkeintaan kolme matkaa tekevien lukumäärä nähdään muuttujan arvon 3 summafrekvenssistä, joka on 85 (henkilöä).

c) Korkeintaan kaksi matkaa tekevien lukumäärä nähdään muuttujan arvon 2 summafrekvenssistä, joka on 72. Prosenttiosuus saadaan vertaamalla frekvenssiä kokonaismäärään:

$$\frac{72}{100} = 0{,}72 = 72\ \%.$$

Siis, 72 % henkilöistä tekee korkeintaan kaksi matkaa.

d) Ainakin kaksi matkaa tekevien lukumäärä saadaan summaamalla 2, 3, 4 ja 5 matkaa tekevien frekvenssit:

$$25 + 13 + 12 + 3 = 53.$$

Tämä voitaisi laskea myös vähentämällä havaintoyksiköiden kokonaismäärästä korkeintaan yhden matkan tekevien lukumäärä:

$$100 - (11 + 36) = 100 - 47 = 53.$$

Ainakin kaksi matkaa tekevien määrä on siis 53 (henkilöä).

25.

Kootaan luettujen kirjojen lukumäärät frekvenssitaulukkoon:

Luettujen kirjojen lukumäärä	***f***	***sf***	***sf %***
0	7	7	28
1	5	12	48
2	6	18	**72**
3	3	21	84
4	2	22	92
5	1	23	96
6	1	25	100

a) Moodi on se muuttujan arvo, jonka frekvenssi on suurin eli Mo = 0 (kirjaa).

b) Mediaani on se muuttujan arvo, jonka summafrekvenssi ensimmäisen kerran ylittää 50 % eli Md = 2 (kirjaa).

c) Luettujen kirjojen keskiarvo on

$$\bar{x} = \frac{7 \cdot 0 + 5 \cdot 1 + 6 \cdot 2 + 3 \cdot 3 + 2 \cdot 4 + 5 + 6}{25} = \frac{45}{25} = 1{,}8 \text{ (kirjaa).}$$

d) Uusi opiskelija kasvattaa muuttujan arvon 1 frekvenssiä yhdellä ja havaintoyksiköiden kokonaismäärää yhdellä. Myös summafrekvenssit muuttuvat.

Laaditaan uusi frekvenssitaulukko:

Luettujen kirjojen lukumäärä	***f***	***sf***	***sf* %**
0	7	7	26,9
1	6	13	50,0
2	6	19	73,1
3	3	22	84,5
4	2	24	92,3
5	1	25	96,2
6	1	26	100,0

Taulukossa suhteelliset summafrekvenssit on muuttujan arvosta 2 alkaen pyöristetty prosentin kymmenesosan tarkkuuteen.

Muuttujan arvolla 0 on edelleen suurin frekvenssi, joten Mo = 0 (kirjaa). Moodi ei muutu.

Havaintoyksiköiden kokonaismäärä on 26, joka on parillinen luku. Suuruusjärjestykseen asetetussa aineistossa ei siis ole keskimmäistä lukua. Muuttujan arvon 1 suhteellinen summafrekvenssi on tasan 50 %, joten keskimmäiset kaksi lukua ovat 1 ja 2. Mediaani on kahden keskimmäisen luvun keskiarvo:

$$\mathrm{Md} = \frac{1+2}{2} = \frac{3}{2} = 1{,}5.$$

Mediaani siis muuttuu, ja uusi mediaani on Md = 1,5 (kirjaa).

Luettujen kirjojen uusi keskiarvo on

$$\bar{x} = \frac{7 \cdot 0 + 6 \cdot 1 + 6 \cdot 2 + 3 \cdot 3 + 2 \cdot 4 + 5 + 6}{26}$$

$$= \frac{46}{26}$$

$$= 1{,}769\ldots \approx 1{,}77 \text{ (kirjaa).}$$

Uusi keskiarvo on siis hivenen pienempi kuin vanha keskiarvo 1,8 kirjaa.

26.

a) Merkitään viidennen kortin pistearvoa kirjaimella *x.*

Kaikkien korttien pistearvojen summa on tällöin

$$3 + 5 + 3 + 7 + x = x + 18$$

ja pistearvojen keskiarvo on $\frac{x+18}{5}$.

Tiedetään, että keskiarvo on tasan 5, joten saadaan yhtälö:

$$\frac{x + 18}{5} = 5 \quad | \cdot 5$$

$$x + 18 = 25$$

$$x = 25 - 18 = 7.$$

Viides kortti on siis ruutu7.

b) Merkitään neljännen kortin pistearvoa kirjaimella x. Neljäs ja viides kortti ovat peräkkäisiä ruutuja, joten niiden pistearvot ovat peräkkäisiä kokonaislukuja. Viidennen kortin pistearvo on siis $x + 1$.

Kaikkien korttien pistearvojen summa on tällöin

$$3 + 5 + 3 + x + (x + 1) = 2x + 12$$

ja pistearvojen keskiarvo on $\frac{2x+12}{5}$.

Tiedetään, että keskiarvo on 6,4 joten saadaan yhtälö:

$$\frac{2x + 12}{5} = 6{,}4 \quad | \cdot 5$$

$$2x + 12 = 32$$

$$2x = 20$$

$$x = \frac{20}{2} = 10.$$

Neljännen kortin pistearvo on siis 10 ja viidennen kortin pistearvo on $10 + 1 = 11$.

Kaksi viimeistä korttia ovat siis ruutu10 ja ruutu11 (eli ruutusotilas).

27.

a) Tutkittava ominaisuus eli muuttuja on A-englannin yo-kokeen arvosana pistemäärän avulla ilmoitettuna.

b) Muuttuja voi saada vain tiettyjä kokonaislukuarvoja (arvot 0, 2–7), joten muuttuja on diskreetti.

c) Kyseessä on kokonaisaineisto, joten tutkimusta varten on kerätty tieto kaikilta perusjoukkoon kuuluvilta. Koska tutkitaan erään koulun A-englannin yo-kokeen arvosanoja, perusjoukkona on kaikkien kyseisen koulun A-englannin yo-kokeen kirjoittaneiden joukko.

d) Moodi on se muuttujan arvo, jonka frekvenssi on suurin eli Mo = 4.

e) Mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää 50 % eli Md = 4.

1 Tilastot

1.2 Tilastokuvaajien tulkitseminen

28.

a) Kukin pylvään korkeus kertoo muuttujan arvon frekvenssin. Muuttujan arvolla 5 pylvään korkeus on 45. Tutkimukseen osallistuneista kotitalouksista viisi matkapuhelinta on siis 45 kotitaloudessa.

b) Vähintään kolme matkapuhelinta sisältää matkapuhelinten määrät 3–7. Lasketaan näiden frekvenssit yhteen:

$$70 + 95 + 45 + 20 + 10 = 240.$$

Niitä kotitalouksia, joissa on vähintään kolme matkapuhelinta, on 240.

c) Kaikkien tutkimukseen osallistuneiden kotitalouksien lukumäärä saadaan, kun lasketaan kaikki frekvenssit yhteen:

$$5 + 15 + 40 + 70 + 95 + 45 + 20 + 10 = 300.$$

Tutkimukseen osallistui 300 kotitaloutta.

d) Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Frekvenssijakauman pylväsdiagrammissa moodi on siis se muuttujan arvo, jonka pylväs on kaikkein korkein. Kotitalouksissa olevien matkapuhelimien lukumäärän moodi on Mo = 4 (matkapuhelinta).

29.

a) Sektoridiagrammin mukaan asunnon arvo on 67 % varallisuudesta. Lasketaan, kuinka paljon on 67 % perheen varallisuudesta, joka on 200 000 euroa.

Muutetaan prosenttiosuus desimaaliluvuksi ja kerrotaan: $67\ \% = 0{,}67$ ja

$$0{,}67 \cdot 200\ 000\ \text{eur} = 134\ 000\ \text{eur}.$$

Perheen asunnon arvo on 134 000 euroa.

b) Sektoridiagrammin mukaan talletusten määrä on 9 % varallisuudesta. Lasketaan, kuinka paljon on 9 % perheen varallisuudesta, joka on 200 000 euroa.

Muutetaan prosenttiosuus desimaaliluvuksi ja kerrotan: $9\ \% = 0{,}09$ ja

$$0{,}09 \cdot 200\ 000\ \text{eur} = 18\ 000\ \text{eur}.$$

Perheen talletusten määrä on 18 000 euroa.

c) Sektoridiagrammin mukaan arvopaperisijoitusten määrä on 6 % varallisuudesta. Lasketaan, kuinka paljon on 6 % perheen varallisuudesta, joka on 200 000 euroa.

Muutetaan prosenttiosuus desimaaliluvuksi ja kerrotaan: $6\ \% = 0{,}06$ ja

$$0{,}06 \cdot 200\ 000\ \text{eur} = 12\ 000\ \text{eur}.$$

Perheen arvopaperisijoitusten määrä on 12 000 euroa.

30.

a) ”Täydet pisteet” vastaa muuttujan arvoa 6. Pylvään korkeus kertoo muuttujan arvon frekvenssin. Muuttujan arvolla 6 pylvään korkeus on 7, joten seitsemän opiskelijaa sai tehtävästä täydet pisteet.

b) Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Pistejakaumassa muuttujan arvoilla 3 ja 5 on kaikista korkein pylväs eli kaikkein suurin frekvenssi (frekvenssi on 8). Pistemäärällä on siis kaksi moodia: Mo = 3 pistettä ja Mo = 5 pistettä.

c) Kaikkien ensimmäiseen tehtävään vastanneiden opiskelijoiden lukumäärä saadaan, kun lasketaan frekvenssit yhteen:

$$2 + 5 + 3 + 8 + 1 + 8 + 7 = 34.$$

Opiskelijoista 34 vastasi ensimmäiseen tehtävään.

d) ”Yli puolet pisteistä” vastaa muuttujan arvoja 4–6. Lasketaan näiden frekvenssit yhteen:

$$1 + 8 + 7 = 16.$$

Opiskelijoista 16 sai ensimmäisestä tehtävästä yli puolet pisteistä.

31.

a) Indeksin arvo luetaan kuvaajalta pystyasteikolta. Heinäkuussa 2015 (eli 7.2015) kuvaajan pystyasteikon arvo, eli käyrän korkeus, on noin 5500. Indeksin arvo oli heinäkuussa 2015 noin 5500.

b) Luetaan kuvaajaa vasemmalta oikealle ja katsotaan missä vaaka-asteikon kohdassa kuvaajan korkeus ylittää arvon 5500. Kuvaajan korkeus saavuttaa arvon 5500 ensimmäisen kerran kohdassa 1.2007. Indeksin arvo ylittää ensimmäisen kerran arvon 5500 siis tammikuussa 2007.

Vastaus voidaan päätellä myös seuraavasti: piirretään kuvaan vaakaviiva korkeudelle 5500 ja etsitään kohta, jossa kuvaaja leikkaa sen ensimmäisen kerran. Vastaus luetaan vaaka-asteikolta.

c) Etsitään kuvaajalta kohta, missä kuvaajan korkeus on suurimmillaan. Vastaus luetaan vaaka-asteikolta: 7.2007. Indeksin arvo oli kaikkein korkeimmillaan heinäkuussa 2007.

d) Etsitään kuvaajalta kohta, missä kuvaajan korkeus on pienimmillään. Vastaus luetaan vaaka-asteikolta: 1.2009. Indeksin arvo oli matalimmillaan tammikuussa 2009.

32.

Frekvenssijakauman histogrammissa kunkin pylvään korkeus kertoo muuttujan arvon frekvenssin. Summafrekvenssi lasketaan frekvenssien avulla summaamalla siihenastiset frekvenssit.

Frekvenssijakaumassa a frekvenssit ovat: 5, 10, 15 ja 5.

Lasketaan summafrekvenssit näiden avulla. Summafrekvenssit ovat:

$\mathbf{5,} \qquad 5 + 10 = \mathbf{15,} \qquad 15 + 15 = \mathbf{30}, \qquad 30 + 5 = \mathbf{35.}$

Summafrekvenssikuvaajassa f pisteiden korkeudet vastaavat frekvenssijakauman a summafrekvenssin arvoja. Siis, frekvenssijakauman a summafrekvenssijakauma on f.

Frekvenssijakaumassa c frekvenssit ovat: 15, 10, 6 ja 4.

Lasketaan summafrekvenssit näiden avulla. Summafrekvenssit ovat:

$\mathbf{15,} \qquad 15 + 10 = \mathbf{25,} \qquad 25 + 6 = \mathbf{31}, \qquad 31 + 4 = \mathbf{35.}$

Summafrekvenssikuvaajassa b pisteiden korkeudet vastaavat frekvenssijakauman c summafrekvenssin arvoja. Siis, frekvenssijakauman c summafrekvenssijakauma on b.

Frekvenssijakaumassa e frekvenssit ovat: 4, 6, 10 ja 15.

Lasketaan summafrekvenssit näiden avulla. Summafrekvenssit ovat:

$\mathbf{4,} \qquad 4 + 6 = \mathbf{10,} \qquad 10 + 10 = \mathbf{20}, \qquad 20 + 15 = \mathbf{35.}$

Summafrekvenssikuvaajassa d pisteiden korkeudet vastaavat frekvenssijakauman e summafrekvenssin arvoja. Siis, frekvenssijakauman e summafrekvenssijakauma on d.

Vastaus: c – b, a – f ja e – d

33.

a) Suhteellinen summafrekvenssijakauma kuvaa tiettyyn muuttujan arvoon mennessä kertyneiden havaintoyksiköiden määrää prosentteina. Päiväkodin lapsista korkeintaan kolmevuotiaita on 40 %.

b) Päiväkodin lapsista korkeintaan kaksivuotiaita on 20 %. Koska kaikkien päiväkodin lapsien määrä prosentteina on 100 %, niin yli kaksivuotiaita on $100\ \% - 20\ \% = 80\ \%$.

c) Mediaani on se muuttujan arvo, joka sijaitsee jakauman puolivälissä. Kuvaajassa mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi on 50 %. Luetaan 50 prosentin summafrekvenssiä vastaava muuttujan arvo kuvaajan vaaka-akselilta: muuttujan arvo on 3,5 (vuotta). Päiväkodin lasten mediaani-ikä on siis 3,5 vuotta.

d) Päiväkodin lapsista korkeintaan neljävuotiaita on 60 %. Lasketaan, kuinka paljon on 60 % 80 lapsesta. Muutetaan prosenttiosuus desimaaliluvuksi ja kerrotaan: $60\ \% = 0{,}6$ ja

$$0{,}6 \cdot 80 = 48.$$

Päiväkodin lapsista 48 lasta on iältään korkeintaan neljä vuotta.

34.

a) Mediaani on se muuttujan arvo, joka sijaitsee jakauman puolivälissä eli mediaania pienempiä on 50 % aineistosta. Luetaan 50 %:n summafrekvenssiä vastaava muuttujan arvo kuvaajan vaaka-akselilta: muuttujan arvo sijaitsee arvojen 165 ja 170 puolivälissä, joten se on 167,5 (cm). Harrastekerholaisten mediaanipituus on 167,5 cm.

b) Alakvartiili on muuttujan arvo, jota pienempiä on 25 % aineistosta. Luetaan 25 %:n summafrekvenssiä vastaava muuttujan arvo kuvaajan vaaka-akselilta: muuttujan arvo sijaitsee arvojen 160 ja 165 puolivälissä, joten se on 162,5 (cm). Harrastekerholaisten pituuden alakvartiili on 162,5 cm.

c) Yläkvartiili on muuttujan arvo, jota suurempia on 25 % aineistosta. Yläkvartiilia pienempiä on näin ollen $100\ \% - 25\ \% = 75\ \%$ aineistosta. Luetaan 75 %:n summafrekvenssiä vastaava muuttujan arvo kuvaajan vaaka-akselilta: muuttujan arvo sijaitsee arvojen 170 ja 175 puolivälissä, joten se on 172,5 (cm). Harrastekerholaisten pituuden yläkvartiili on 172,5 cm.

35.

a) Moodi on se muuttujan arvo, jonka pylväs on kaikkein korkein, eli Mo = 1.

b) Mediaani on suuruusjärjestykseen laitetun aineiston keskimmäinen luku. Järjestetään muuttujan arvot suuruusjärjestykseen. Kuvaajassa pylvään korkeus kertoo muuttujan arvon frekvenssin, eli esiintymiskertojen lukumäärän:

1 1 1 1 **2** **2** 3 4 5 5

Havaintoyksiköitä on parillinen määrä (10 kappaletta), joten aineistossa ei ole keskimmäistä lukua. Mediaani on kahden keskimmäisen luvun keskiarvo:

Mediaani Md $= \frac{2+2}{2} = \frac{4}{2} = 2.$

c) Lasketaan muuttujan A keskiarvo jakamalla muuttujan arvojen summa lukumäärällä eli luvulla 10. Summan laskemisessa voidaan hyödyntää frekvenssiä, joka kertoo muuttujan arvojen lukumäärän:

Keskiarvo $\bar{x} = \frac{4\cdot 1+2\cdot 2+3+4+2\cdot 5}{10} = \frac{25}{10} = 2{,}5.$

36.

Määritetään frekvenssijakaumien kuvaajista jakaumien keskikohdat eli mediaanit.

Mediaani on se muuttujan arvo, joka jakaa aineiston kahteen yhtä suureen osaan, eli jonka ala- ja yläpuolelle jää yhtä paljon havaintoja. Kaikkien opiskelijoiden lukumäärä saadaan, kun lasketaan kaikki frekvenssi yhteen: opiskelijoita on ryhmässä yhteensä 30. Koska havaintoaineiston kokonaismäärä on parillinen luku, keskimmäistä lukua ei ole vaan mediaani on kahden keskimmäisen luvun keskiarvo.

Kunkin arvosanajakauman mediaanin määrittämiseksi etsitään sitä muuttujan arvoa, jonka alapuolelle jää 14 havaintoa ja sitä muuttujan arvoa, jonka yläpuolelle jää 14 havaintoa:

Fysiikan kurssilla mediaani on $\frac{6+6}{2} = 6$.

Kemian kurssilla mediaani on $\frac{9+9}{2} = 9$.

Biologian kurssilla mediaani on $\frac{6+7}{2} = 6{,}5$.

Maantieteen kurssilla mediaani on $\frac{7+7}{2} = 7$.

a) Kemian kurssilla arvosanojen mediaani on suurin (Md = 9), joten kemian kurssilla arvosanojen keskiarvokin on suurin. Kemian kurssilla menestyttiin parhaiten.

b) Fysiikan kurssilla arvosanojen mediaani on pienin (Md = 6), joten fysiikan kurssilla arvosanojen keskiarvokin on suurin. Fysiikan kurssilla menestyttiin heikoiten.

c) Arvosanajakauma on tasaisin silloin, kun keskihajonta on pienin eli silloin, kun suuri osa muuttujan arvoista on lähellä keskiarvoa. Maantiedon kurssilla on vähiten jakauman keskikohdasta poikkeavia arvosanoja, joten maantiedon kurssin keskihajonta on pienin. Maantiedon kurssin arvosanat jakaantuvat tasaisimmin.

d) Arvosanajakauma on epätasaisin silloin, kun keskihajonta on suurin eli silloin, kun suuri osa muuttujan arvoista on kaukana keskiarvosta. Biologian kurssilla on eniten jakauman keskikohdasta poikkeavia arvosanoja, joten biologian kurssin keskihajonta on suurin. Biologian kurssin arvosanat jakaantuvat epätasaisimmin.

37.

a) Vaihteluvälin pituus lasketaan vähentämällä aineiston suurimmasta muuttujan arvosta aineiston pienin muuttujan arvo.

Muuttujan A suurin arvo on 8 ja pienin arvo on 1 joten vaihteluvälin pituus on 8 – 1 = 7.

Muuttujan B suurin arvo on 10 ja pienin arvo on 1, joten vaihteluvälin pituus on 10 – 1 = 9.

b) Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Muuttujan A moodi on $Mo_A = 5$ ja muuttujan B moodi on $Mo_B = 4$. Muuttujan A moodi on suurempi.

c) Keskihajonta on suuri silloin, kun muuttujan arvot poikkeavat paljon keskiarvosta. Muuttujan B jakaumassa kaukana keskiarvosta olevilla muuttujan arvoilla 1 ja 10 on suuret frekvenssit, joten suuri osa arvoista on kaukana keskiarvosta 5.

Muuttujan B keskihajonta on siis suurempi.

38.

a) Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Muuttujan A moodi on $Mo_A = 2$ ja muuttujan B moodi on $Mo_B = 4$. Muuttujan B moodi on suurempi.

b) Muuttuja A saa eniten pieniä arvoja 1 ja 2 (frekvenssit ovat 5 ja 6) ja vähiten suuria arvoja 4 ja 5 (frekvenssit ovat 2 ja 1) kun taas muuttuja B saa eniten suuria arvoja 4 ja 5 (frekvenssi ovat 6 ja 5) ja vähiten pieniä arvoja 1 ja 2 (frekvenssi ovat 1 ja 2). Tällä perusteella muuttujan B keskiarvo on suurempi.

Jakaumien perusteella voidaan myös laskea keskiarvojen tarkat arvot:

Muuttujan A keskiarvo on

$$\bar{x}_A = \frac{5 \cdot 1 + 6 \cdot 2 + 3 \cdot 3 + 2 \cdot 4 + 5}{17} = \frac{39}{17} = 2{,}294 \ldots \approx 2{,}3$$

Muuttujan B keskiarvo on

$$\bar{x}_B = \frac{1 + 2 \cdot 2 + 3 \cdot 3 + 6 \cdot 4 + 5 \cdot 5}{17} = \frac{63}{17} = 3{,}705 \ldots \approx 3{,}7$$

Tämän laskun perusteella muuttujan B keskiarvo on suurempi.

39.

a) Muuttuja A saa arvot 1–5. Suurin osa muuttujan A arvoista on välillä 1–3. Muuttuja B saa arvot 4–8. Suurin osa muuttujan B arvoista on välillä 6–8. Tällä perusteella muuttujan B keskiarvo on suurempi kuin muuttujan A keskiarvo.

b) Keskihajonta kuvaa muuttujan arvojen hajaantumista keskiarvon ympärille. Muuttujien A ja B saamien arvojen frekvenssit ovat kummassakin jakaumassa 2, 6, 5, 4 ja 1, kun frekvenssit on lueteltu järjestyksessä pienimmästä muuttujan arvosta suurimpaan muuttujan arvoon. Muuttujien arvojen frekvenssit ovat siis jakaumissa samat. Tästä johtuen molempien muuttujien arvot hajaantuvat samalla tavoin keskiarvon ympärille. Muuttujien A ja B keskihajonnat ovat yhtä suuret.

40.

a) Jos nopeus poikkeaa korkeintaan kahden keskihajonnan verran keskiarvosta, niin poikkeama ei ole merkitsevä.

Lasketaan ne nopeudet jotka poikkeavat tasan kahden keskihajonnan verran keskinopeudesta $\bar{x} = 40 \text{ cm/s}$.

Keskinopeudesta kaksi keskihajontaa alaspäin oleva nopeus on:

$$\bar{x} - 2s = 40 \text{ cm/s} - 2 \cdot 1{,}4 \text{ cm/s} = 40 \text{ cm/s} - 2{,}8 \text{ cm} = 37{,}2 \text{ cm/s}.$$

Keskinopeudesta kaksi keskihajontaa ylöspäin oleva nopeus on:

$$\bar{x} + 2s = 40 \text{ cm/s} + 2 \cdot 1{,}4 \text{ cm/s} = 40 \text{ cm/s} + 2{,}8 \text{ cm} = 42{,}8 \text{ cm/s}.$$

Jotta nopeus ei poikkea merkitsevästi keskiarvosta, niin nopeuden on oltava välillä 37,2 cm/s – 42,8 cm/s.

b) Merkitään etsittyä ilmatyynyradalla kulkevan vaunun nopeutta kirjaimella *x.* Lasketaan nopeuden normitettu arvo. Ilmatyynyradalla kulkevien vaunujen keskinopeus on $\bar{x} = 40$ (cm/s) ja keskihajonta on $s = 1{,}4$ (cm/s), joten nopeuden x normitettu arvo on

$$z_{\text{vaunu}} = \frac{x - \bar{x}}{s} = \frac{x - 40}{1{,}4}.$$

Lasketaan poliisin mittaaman kaahariauton nopeuden 150 km/h normitettu arvo. Moottoritiellä kulkevien autojen keskinopeus on $\bar{x} = 105$ (km/h) ja keskihajonta on $s = 9{,}0$ (km/h), joten nopeuden $x = 150$ (km/h) normitettu arvo on

$$z_{\text{auto}} = \frac{x - \bar{x}}{s} = \frac{150 - 105}{9{,}0} = \frac{45}{9} = 5.$$

Jos ilmatyynyradalla kulkeva vaunu ja moottoritiellä kaahannut auto kulkevat suhteessa yhtä nopeasti, on nopeuksien normitettujen arvojen oltava yhtä suuret.

Muodostetaan yhtälö, jossa nopeuksien normitetut arvot on merkitty yhtä suuriksi:

$$\frac{x - 40}{1{,}4} = 5$$

Ratkaistaan yhtälöstä tuntematon nopeus *x:*

$$\frac{x - 40}{1{,}4} = 5 \quad | \cdot 1{,}4$$

$$x - 40 = 7$$

$$x = 7 + 40 = 47 \text{ (cm/s)}.$$

Vaunun nopeus olisi siis 47 cm/s.

41.

a) Omenien keskimääräinen massa on $\bar{x} = 68{,}0$ (g) ja massan keskihajonta on $s = 4{,}0$ (g). Massan $x = 74$ (g) normitettu arvo on

$$z = \frac{x - \bar{x}}{s} = \frac{74 - 68{,}0}{4{,}0} = \frac{6}{4} = 1{,}5.$$

b) Omenien keskimääräinen massa on $\bar{x} = 68{,}0$ (g) ja massan keskihajonta on $s = 4{,}0$ (g). Massan $x = 58$ (g) normitettu arvo on

$$z = \frac{x - \bar{x}}{s} = \frac{58 - 68{,}0}{4{,}0} = \frac{-10}{4} = -2{,}5.$$

Negatiivinen tulos merkitsee, että massa $x = 58$ (g) on alle omenien keskimääräisen massan.

c) Muuttujan arvo poikkeaa merkitsevästi keskiarvosta, jos se on yli kahden keskihajonnan päässä keskiarvosta. Tällöin normitetun arvon itseisarvo on suurempi kuin 2.

Massan 74 g normitetun arvon itseisarvo on $|z| = 1{,}5 < 2$, joten 74 grammaa painavan omenan massa ei poikkea merkitsevästi keskiarvosta.

Massan 58 g normitetun arvon itseisarvo on $|z| = 2{,}5 > 2$, joten 58 grammaa painavan omenan massa poikkeaa merkitsevästi keskiarvosta.

Tämä voidaan myös ratkaista myös laskemalla ne massat, jotka poikkeavat tasan kahden keskihajonnan verran keskimääräisestä massasta $\bar{x} = 68{,}0$ g.

Keskimassasta kaksi keskihajontaa alaspäin oleva massa on:

$$\bar{x} - 2s = 68{,}0 \text{ g} - 2 \cdot 4{,}0 \text{ g} = 68 \text{ g } - 8 \text{ g} = 60 \text{ g}.$$

Keskimassasta kaksi keskihajontaa ylöspäin oleva massa on:

$$\bar{x} + 2s = 68{,}0 \text{ g} + 2 \cdot 4{,}0 \text{ g} = 68 \text{ g} + 8 \text{ g} = 76 \text{ g}.$$

Jotta omenan massa poikkeaisi merkitsevästi keskiarvosta, niin massan on oltava pienempi kuin 60 grammaa tai suurempi kuin 76 grammaa.

74 grammaa painavan omenan massa ei poikkea merkitsevästi keskiarvosta.

58 grammaa painavan omenan massa poikkeaa merkitsevästi keskiarvosta, koska se on pienempi kuin 60 grammaa.

42.

Aloitetaan normittamalla iät:

Lasketaan 140 vuotta vanhan papukaijan normitettu arvo.
Papukaijojen keski-ikä on $\bar{x} = 92$ (vuotta) ja keskihajonta on $s = 8{,}0$ (vuotta), joten iän $x = 140$ (vuotta) normitettu arvo on

$$z_{\text{kaija}} = \frac{x - \bar{x}}{s} = \frac{140 - 92}{8{,}0} = \frac{48}{8} = 6.$$

Merkitään etsittyä suomalaisen naisen ikää kirjaimella *x.* Lasketaan iän normitettu arvo. Naisten keski-ikä on $\bar{x} = 82$ (vuotta) ja keskihajonta on $s = 4{,}5$ (vuotta), joten iän *x* normitettu arvo on

$$z_{\text{nainen}} = \frac{x - \bar{x}}{s} = \frac{x - 82}{4{,}5}.$$

Jos suomalainen nainen ja supervanha papukaijan ovat suhteessa yhtä vanhat, on ikien normitettujen arvojen oltava yhtä suuret.
Muodostetaan yhtälö, jossa ikien normitetut arvot on merkitty yhtä suuriksi:

$$\frac{x - 82}{4{,}5} = 6$$

Ratkaistaan yhtälöstä tuntematon ikä *x:*

$$\frac{x - 82}{4{,}5} = 6 \quad | \cdot 4{,}5$$

$$x - 82 = 27$$

$$x = 27 + 82 = 109 \text{ (vuotta).}$$

Suomalaisen naisen ikä olisi 109 vuotta.

43.

a) Jos arvosana poikkeaa korkeintaan kahden keskihajonnan verran keskiarvosta, niin poikkeama ei ole merkitsevä. Tällöin arvosanan normitetun arvon itseisarvo on suurempi kuin 2.

Rasmuksen arvosanan $x = 8{,}6$ normitettu arvo on

$$z = \frac{x - \bar{x}}{s} = \frac{8{,}6 - 7{,}2}{0{,}8} = \frac{1{,}4}{0{,}8} = 1{,}75.$$

Normitetun arvon itseisarvo on $|z| = 1{,}75 < 2$, joten Rasmuksen arvosana ei poikkea merkitsevästi keskiarvosta.

Millan arvosanan $x = 5{,}5$ normitettu arvo on

$$z = \frac{x - \bar{x}}{s} = \frac{5{,}5 - 7{,}2}{0{,}8} = \frac{-1{,}7}{0{,}8} = -2{,}125.$$

Normitetun arvon itseisarvo on $|z| = 2{,}125 > 2$, joten Millan arvosana poikkeaa merkitsevästi keskiarvosta.

Tämä voidaan ratkaista myös laskemalla ne historian kokeen arvosanat, jotka poikkeavat tasan kahden keskihajonnan verran arvosanojen keskiarvosta $\bar{x} = 7{,}8$.

Keskiarvosta kaksi keskihajontaa alaspäin oleva arvosana on:

$$\bar{x} - 2s = 7{,}8 - 2 \cdot 0{,}8 = 7{,}8 - 1{,}6 = 6{,}2.$$

Keskiarvosta kaksi keskihajontaa ylöspäin oleva arvosana on:

$$\bar{x} + 2s = 7{,}8 + 2 \cdot 0{,}8 = 7{,}8 + 1{,}6 = 9{,}4.$$

Jotta historian kokeen arvosana poikkeaisi merkitsevästi keskiarvosta, niin arvosanan on oltava pienempi kuin 6,2 tai suurempi kuin 9,4.

Rasmuksen arvosana 8,6 ei poikkea merkitsevästi keskiarvosta. Millan arvosana 5,5 poikkeaa merkitsevästi keskiarvosta, koska se on pienempi kuin 6,2.

b) Normitetaan Rasmuksen arvosanat:

Lasketaan Rasmuksen historian kokeen arvosanan normitettu arvo. Historian kokeen arvosanojen keskiarvo oli $\bar{x} = 7{,}2$ ja keskihajonta oli $s = 0{,}8$ joten Rasmuksen historian kokeen arvosanan $x = 8{,}6$ normitettu arvo on

$$z_{\text{historia}} = \frac{x - \bar{x}}{s} = \frac{8{,}6 - 7{,}2}{0{,}8} = \frac{1{,}4}{0{,}8} = 1{,}75.$$

Lasketaan Rasmuksen ruotsin kokeen arvosanan normitettu arvo. Merkitään ruotsin kokeen keskiarvoa kirjaimella $\bar{x}$. Arvosanojen keskihajonta oli ruotsin kokeessa sama kuin historian kokeessa, eli $s = 0{,}8$. Rasmuksen ruotsin kokeen arvosanan $x = 7{,}5$ normitettu arvo on

$$z_{\text{ruotsi}} = \frac{x - \bar{x}}{s} = \frac{7{,}5 - \bar{x}}{0{,}8}.$$

Rasmus menestyi kummassakin kokeessa suhteessa yhtä hyvin, joten arvosanojen normitetut arvot ovat yhtä suuret. Muodostetaan yhtälö, jossa arvosanojen normitetut arvot on merkitty yhtä suuriksi:

$$\frac{7{,}5 - \bar{x}}{0{,}8} = 1{,}75.$$

Ratkaistaan yhtälöstä tuntematon ruotsin kokeen keskiarvo $\bar{x}$:

$$\frac{7{,}5 - \bar{x}}{0{,}8} = 1{,}75 \quad | \cdot 0{,}8$$

$$7{,}5 - \bar{x} = 1{,}4$$

$$-\bar{x} = 1{,}4 - 7{,}5 = -6{,}1$$

$$\bar{x} = 6{,}1$$

Ruotsin kokeen keskiarvo oli 6,1.

44.

a) ”Ei lainkaan” vastaa muuttujan arvoa 0. Pylvään korkeus kertoo muuttujan arvon frekvenssin. Muuttujan arvolla 0 pylvään korkeus on 2. Siis kaksi työntekijää ei harrasta liikuntaa lainkaan.

b) ”Ainakin kolmena päivänä” vastaa muuttujan arvoja 3–6. Lasketaan näiden frekvenssit yhteen:

$$6 + 2 + 3 + 1 = 12.$$

Työntekijöistä 12 harrastaa liikuntaa ainakin kolmena päivänä.

c) Koska tutkimusta varten tieto urheilupäivistä kerättiin kaikilta yrityksen työntekijöiltä, niin tutkimusaineisto on kokonaisaineisto.

d) Kaikkien tutkimukseen osallistuneiden työntekijöiden lukumäärä saadaan, kun lasketaan kaikki frekvenssit yhteen:

$$2 + 3 + 6 + 6 + 2 + 3 + 1 = 23.$$

Yrityksessä on siis 23 työntekijää.

e) Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Frekvenssijakauman kuvaajassa muuttujan arvoilla 2 ja 3 on kaikista korkein pylväs eli kaikkein suurin frekvenssi (frekvenssi on 6). Urheilupäivien lukumäärällä on siis kaksi moodia: Mo = 2 ja Mo = 3 (urheilupäivää viikossa).

45.

a) Moodi on se muuttujan arvo, jonka frekvenssi, ja siis myös suhteellinen frekvenssi, on suurin. Diagrammin mukaan suurin suhteellinen frekvenssi on arvosanalla C ($f\,\%$ = 24). Ylioppilaskokeen lyhyen matematiikan arvosanan moodi on siis Mo = C.

b) Diagrammin mukaan arvosanan I sai 5 % osallistujista ja arvosanan A sai 11 % osallistujista. Jommankumman heikoimmista arvosanoista sai siis 5 % + 11 % = 16 % kokeeseen osallistujista.

c) Diagrammin mukaan arvosanan E sai 15 % osallistujista. Lasketaan, kuinka paljon on 15 % osallistujien kokonaismäärästä 80.

Muutetaan prosenttiosuus desimaaliluvuksi ja kerrotaan: $15\ \% = 0{,}15$ ja

$$0{,}15 \cdot 80 = 12.$$

Koulun lyhyen matematiikan kirjoittajista 12 opiskelijaa sai arvosanan E.

d) Mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää 50 %. Muodostetaan arvosanoista frekvenssitaulukko, jossa arvosanat on asetettu suuruusjärjestykseen (I on pienin, L on suurin):

Arvosana	***f* %**	***sf* %**
I	5	5
A	11	$5 + 11 = 16$
B	20	$16 + 20 = 36$
C	24	$36 + 24 = \mathbf{60}$
M	20	$60 + 20 = 80$
E	15	$80 + 15 = 95$
L	5	$95 + 5 = 100$

Suhteellisesta summafrekvenssisarakkeesta nähdään, että lyhyen matematiikan yo-kokeen arvosanan mediaani on Md = C.

46.

a) Vaaka-asteikolla muuttujan arvoa 34 (minuuttia) vastaava summafrekvenssin arvo pystyasteikolla on noin 75 %, joten korkeintaan 34 minuuttia 5,0 kilometrin juoksuun käyttää noin 75 % opiskelijoista.

b) Vaaka-asteikolla muuttujan arvoa 40 (minuuttia) vastaava summafrekvenssin arvo pystyasteikolla on 90 %, joten korkeintaan 40 minuuttia 5,0 kilometrin juoksuun käyttää noin 90 % opiskelijoista. Yli 40 minuuttia juoksuun käyttää silloin 100 % – 90 % = 10 % opiskelijoista.

c) Mediaani on se muuttujan arvo, jota vastaava suhteellinen summafrekvenssi on 50 %. Pystyasteikolla summafrekvenssiä 50 % vastaava muuttujan arvo vaaka-akselilla on 30 (minuuttia), joten juoksuajan mediaani on Md = 30 (minuuttia).

d) Vaaka-asteikolla muuttujan arvoa 24 (minuuttia) vastaava summafrekvenssin arvo pystyasteikolla on 10 %, joten opiskelijoista 10 % käyttää 5,0 kilometrin juoksuun korkeintaan 24 minuuttia. Lasketaan, kuinka paljon 10 % on 30 opiskelijasta:

$$0{,}1 \cdot 30 = 3.$$

Alle 24 minuuttia juoksumatkaan käyttää kolme opiskelijaa.

47.

a) Jos aluskasvillisuuden korkeus poikkeaa korkeintaan kahden keskihajonnan verran keskiarvosta, niin poikkeama ei ole merkitsevä.

Lasketaan ne aluskasvillisuuden korkeudet jotka poikkeavat tasan kahden keskihajonnan verran keskikorkeudesta $\bar{x} = 15{,}0 \text{ cm}$.

Keskikorkeudesta kaksi keskihajontaa alaspäin oleva arvo on:

$$\bar{x} - 2s = 15{,}0 \text{ cm} - 2 \cdot 3{,}0 \text{ cm} = 15 \text{ cm} - 6 \text{ cm} = 9 \text{ cm}.$$

Keskikorkeudesta kaksi keskihajontaa ylöspäin oleva arvo on:

$$\bar{x} + 2s = 15{,}0 \text{ cm} + 2 \cdot 3{,}0 \text{ cm} = 15 \text{ cm} + 6 \text{ cm} = 21 \text{ cm}.$$

Jotta aluskasvillisuuden korkeus poikkeaisi merkitsevästi keskiarvosta, niin korkeuden on oltava joko alle 9,0 cm tai yli 21 cm.

b) Aloitetaan normittamalla korkeudet:

Lasketaan tiheässä metsässä kasvavan varvun korkeuden normitettu arvo. Tiheässä metsässä aluskasvillisuuden keskikorkeus on $\bar{x} = 15{,}0$ (cm) ja keskihajonta on $s = 3{,}0$ (cm), joten varvun korkeuden $x = 12{,}0$ (cm) normitettu arvo on

$$z_{\text{varpu}} = \frac{x - \bar{x}}{s} = \frac{12{,}0 - 15{,}0}{3{,}0} = \frac{-3}{3} = -1.$$

Lasketaan metsäaukealla kasvavan heinän korkeuden normitettu arvo. Metsäaukealla aluskasvillisuuden keskikorkeus on $\bar{x} = 40{,}0$ (cm). Merkitään keskihajontaa kirjaimella s (cm). Heinän korkeuden $x = 30{,}0$ (cm) normitettu arvo on

$$z_{\text{heinä}} \frac{x - \bar{x}}{s} = \frac{30{,}0 - 40{,}0}{s} = \frac{-10}{s}.$$

Jos tiheässä metsässä kasvava varpu ja metsäaukealla kasvava heinä ovat suhteessa yhtä korkeat, on korkeuksien normitettujen arvojen oltava yhtä suuret. Muodostetaan yhtälö, jossa korkeuksien normitetut arvot on merkitty yhtä suuriksi:

$$\frac{-10}{s} = -1.$$

Ratkaistaan yhtälöstä tuntematon keskihajonta s:

$$s = 10.$$

Metsäaukiolla kasvavan aluskasvillisuuden korkeuden keskihajonta on siis s = 10,0 (cm).

48.

Tässä aineistossa muuttujana on matematiikan ja luonnontieteiden kurssien lukumäärä, joka saa arvot 1–5 (kurssia).

a) Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Moodit ovat:

Jakso 1 Mo = 2 (kurssia)

Jakso 2 Mo = 3 (kurssia)

Jakso 3 Mo = 1 (kurssia)

Jakso 4 Mo = 5 (kurssia)

Jaksossa 5 kaikilla muuttujan arvoilla on sama frekvenssi 20, joten kaikki muuttujan arvot ovat moodeja: Mo = 1 ja Mo = 2 ja Mo = 3 ja Mo = 4 ja Mo = 5.

Matematiikan ja luonnontieteiden kurssien määrän moodi oli suurin jaksossa 4.

b) Arvioidaan frekvenssijakaumien kuvaajista jakaumien keskikohdat, eli mediaanit: mediaani on se muuttujan arvo, joka jakaa havaintoarvot kahteen yhtä suuren ryhmään.

Jakso 1 Md = 2 (kurssia)

Jakso 2 Md = 3 (kurssia)

Jakso 3 Md = 2,5 (kurssia)

Jakso 4 Md = 5 (kurssia)

Jakso 5 Md = 3 (kurssia)

Jaksossa 4 arvosanojen mediaani on suurin, joten jaksossa 4 myös arvosanojen keskiarvo on suurin. Jaksossa 4 luettiin keskimäärin eniten matematiikkaa ja luonnontieteitä.

c) Jaksossa 2 arvosanojen mediaani on pienin, joten jaksossa 2 myös arvosanojen keskiarvo on pienin. Jaksossa 2 luettiin keskimäärin vähiten matematiikkaa ja luonnontieteitä.

d) Jakson 2 jakaumassa on vähiten keskikohdasta 3 poikkeavia arvoja, joten keskihajonta on pienin jaksossa 2.

e) Kaikkien opiskelijoiden lukumäärä saadaan, kun lasketaan kaikki frekvenssit yhteen. Jaksossa 5 kaikilla muuttujan arvoilla on sama frekvenssi 20, joten lasketaan lukumäärä jakson 5 frekvenssien avulla:

$$20 + 20 + 20 + 20 + 20 = 5 \cdot 20 = 100.$$

1 Tilastot

1.3 Frekvenssijakauman laatiminen

49.

Hyödynnetään taulukkolaskentaohjelmaa frekvenssien määrittämisessä.

Taulukon ensimmäiselle riville kirjoitetaan sarakkeiden otsikot. Sarakkeeseen A kirjoitetaan muuttujan arvot 1–5 ja sarakkeeseen B frekvenssien arvot.

Aloitetaan frekvenssitaulukon täyttäminen laskemalla summafrekvenssit sarakkeeseen C. Summafrekvenssi lasketaan summaamalla siihenastiset frekvenssit:

- soluun C2 kirjoitetaan ensimmäisen muuttujan arvon frekvenssi eli luku 4
- soluun C3 kirjoitetaan laskukaava: $= \mathrm{C2} + \mathrm{B3}$
- kaava kopioidaan sarakkeen C seuraaville riveille

Viimeisen muuttujan arvon 5 summafrekvenssi kertoo havaintoyksiköiden lukumäärän. Solusta C6 nähdään, että havaintoyksiköiden lukumäärä on 61.

Lasketaan suhteelliset frekvenssit sarakkeeseen D. Suhteellinen frekvenssi lasketaan jakamalla muuttujan frekvenssi havaintoyksiköiden lukumäärällä 61 ja ilmaisemalla osamäärä prosentteina:

- soluun D2 kirjoitetaan laskukaava: $= \mathrm{B2}/61 \cdot 100$
- kaava kopioidaan sarakkeen D seuraaville riveille

Lasketaan suhteelliset summafrekvenssit sarakkeeseen E. Suhteellinen summafrekvenssi lasketaan summaamalla siihenastiset suhteelliset frekvenssit:

- soluun E2 kirjoitetaan ensimmäinen suhteellinen frekvenssi eli laskukaava: $= \mathrm{D2}$
- soluun E3 kirjoitetaan laskukaava: $= \mathrm{E2} + \mathrm{D3}$
- kaava kopioidaan sarakkeen E seuraaville riveille

	A	B	C	D	E
1	**x**	**f**	**sf**	**f%**	**sf%**
2	1	4	4	7	7
3	2	10	14	16	23
4	3	23	37	38	61
5	4	15	52	25	85
6	5	9	61	15	100

Huomaa, että vastauksessa prosenttiluvut pyöristetään prosenttien tarkkuuteen.

50.

Laaditaan frekvenssijakauma taulukkolaskentaohjelmalla.

Taulukon ensimmäiselle riville kirjoitetaan sarakkeiden otsikot. Sarake A nimetään muuttujan mukaan: bongausten lukumäärä. Muuttujan arvo, eli bongausten lukumäärä, vaihtelee välillä 1–6. Kirjoitetaan sarakkeeseen A muuttujan arvot 1–6.

Lasketaan kunkin muuttujan arvon frekvenssi, eli esiintymiskertojen lukumäärä. Kirjoitetaan frekvenssi sarakkeeseen B.

	A	B
1	**Bongausten lukumäärä**	**f**
2	1	4
3	2	3
4	3	5
5	4	3
6	5	4
7	6	3

Aloitetaan frekvenssitaulukon täyttäminen laskemalla summafrekvenssit sarakkeeseen C. Summafrekvenssi lasketaan summaamalla siihenastiset frekvenssit:

- soluun C2 kirjoitetaan ensimmäisen muuttujan arvon frekvenssi eli luku 4
- soluun C3 kirjoitetaan laskukaava: $= \mathrm{C2} + \mathrm{B3}$
- kaava kopioidaan sarakkeen C seuraaville riveille

Viimeisen muuttujan arvon 6 summafrekvenssi kertoo havaintoyksiköiden lukumäärän. Solusta C7 nähdään, että havaintoyksiköiden lukumäärä on 22.

Lasketaan suhteelliset frekvenssit sarakkeeseen D. Suhteellinen frekvenssi lasketaan jakamalla muuttujan frekvenssi havaintoyksiköiden lukumäärällä 22 ja ilmaisemalla osamäärä prosentteina:

- soluun D2 kirjoitetaan laskukaava: $= \mathrm{B2}/22 \cdot 100$
- kaava kopioidaan sarakkeen D seuraaville riveille
- muutetaan sarakkeen D solujen pyöristystarkkuus yhden desimaalin tarkkuuteen

Lasketaan suhteelliset summafrekvenssit sarakkeeseen E. Suhteellinen summafrekvenssi lasketaan summaamalla siihenastiset suhteelliset frekvenssit:

- soluun E2 kirjoitetaan ensimmäinen suhteellinen frekvenssi eli laskukaava: $= \mathrm{D2}$
- soluun E3 kirjoitetaan laskukaava: $= \mathrm{E2} + \mathrm{D3}$
- kaava kopioidaan sarakkeen E seuraaville riveille
- muutetaan sarakkeen E solujen pyöristystarkkuus yhden desimaalin tarkkuuteen

	A	B	C	D	E
1	**Bongausten lukumäärä**	**f**	**sf**	**f%**	**sf%**
2	1	4	4	18,2	18,2
3	2	3	7	13,6	31,8
4	3	5	12	22,7	54,5
5	4	3	15	13,6	68,2
6	5	4	19	18,2	86,4
7	6	3	22	13,6	100,0

51.

a) Laaditaan frekvenssijakauma taulukkolaskentaohjelmalla.

Taulukon ensimmäiselle riville kirjoitetaan sarakkeiden otsikot. Sarake A nimetään muuttujan mukaan: Tähtien lukumäärä. Muuttujan arvo, eli tähtien lukumäärä, vaihtelee välillä 1–5. Kirjoitetaan sarakkeeseen A muuttujan arvot 1–5.

Lasketaan kunkin muuttujan arvon frekvenssi, eli esiintymiskertojen lukumäärä. Kirjoitetaan frekvenssi sarakkeeseen B:

	A	B
1	**Tähtien lukumäärä**	**f**
2	1	7
3	2	8
4	3	7
5	4	6
6	5	5

Aloitetaan frekvenssitaulukon täyttäminen laskemalla summafrekvenssit sarakkeeseen C. Summafrekvenssi lasketaan summaamalla siihenastiset frekvenssit:

- soluun C2 kirjoitetaan ensimmäisen muuttujan arvon frekvenssi eli luku 7
- soluun C3 kirjoitetaan laskukaava: $= C2 + B3$
- kaava kopioidaan sarakkeen C seuraaville riveille

Viimeisen muuttujan arvon 5 summafrekvenssi kertoo havaintoyksiköiden lukumäärän. Solusta C6 nähdään, että havaintoyksiköiden lukumäärä on 33.

Lasketaan suhteelliset frekvenssit sarakkeeseen D. Suhteellinen frekvenssi lasketaan jakamalla muuttujan frekvenssi havaintoyksiköiden lukumäärällä 33 ja ilmaisemalla osamäärä prosentteina:

- soluun D2 kirjoitetaan laskukaava: $= B2/33 \cdot 100$
- kaava kopioidaan sarakkeen D seuraaville riveille
- muutetaan sarakkeen D solujen pyöristystarkkuus prosentin tarkkuuteen

Lasketaan suhteelliset summafrekvenssit sarakkeeseen E. Suhteellinen summafrekvenssi lasketaan summaamalla siihenastiset suhteelliset frekvenssit:

- soluun E2 kirjoitetaan ensimmäinen suhteellinen frekvenssi eli laskukaava: $= D2$
- soluun E3 kirjoitetaan laskukaava: $= E2 + D3$
- kaava kopioidaan sarakkeen E seuraaville riveille
- muutetaan sarakkeen E solujen pyöristystarkkuus prosentin tarkkuuteen

	A	B	C	D	E
1	**Tähtien lukumäärä**	**f**	**sf**	**f%**	**sf%**
2	1	7	7	21	21
3	2	8	15	24	45
4	3	7	22	21	67
5	4	6	28	18	85
6	5	5	33	15	100

b) Vähintään neljä tähteä oli arvioissa, joissa oli annettu joko 4 tai 5 tähteä. Näiden prosenttiosuus on kahden viimeisen muuttujan arvon suhteellisten frekvenssien summa:

$$18\ \% + 15\ \% = 33\ \%.$$

Prosenttiosuus voidaan laskea myös summafrekvenssien avulla vähentämällä kaikista arvioista, eli sadasta prosentista, niiden arvioiden määrä, joissa oli annettu korkeintaan kolme tähteä:

$$100\ \% + 67\ \% = 33\ \%.$$

c) Korkeintaan kolmen tähden arvioiden lukumäärä saadaan muuttujan arvon 3 summafrekvenssistä, joka on 22 arviota.

52.

a) Laaditaan frekvenssijakauma taulukkolaskentaohjelmalla.

Taulukon ensimmäiselle riville kirjoitetaan sarakkeiden otsikot. Sarake A nimetään muuttujan mukaan: Käyntikerrat (lkm). Muuttujan arvo, eli käyntikertojen lukumäärä, vaihtelee välillä 0–4. Kirjoitetaan sarakkeeseen A muuttujan arvot 0–4.

Opiskelijoiden lukumäärä kertoo käyntikertojen frekvenssin. Kirjoitetaan sarakkeeseen B kunkin muuttujan arvon frekvenssi.

Lasketaan summafrekvenssit sarakkeeseen C. Summafrekvenssi lasketaan summaamalla siihenastiset frekvenssit:

- soluun C2 kirjoitetaan ensimmäisen muuttujan arvon frekvenssi eli luku 5
- soluun C3 kirjoitetaan laskukaava: $= \mathrm{C2} + \mathrm{B3}$
- kaava kopioidaan sarakkeen C seuraaville riveille

Viimeisen muuttujan arvon 4 summafrekvenssi kertoo havaintoyksiköiden lukumäärän. Solusta C6 nähdään, että havaintoyksiköiden lukumäärä on 55.

Lasketaan suhteelliset frekvenssit sarakkeeseen D. Suhteellinen frekvenssi lasketaan jakamalla muuttujan frekvenssi havaintoyksiköiden lukumäärällä 55 ja ilmaisemalla osamäärä prosentteina:

- soluun D2 kirjoitetaan laskukaava: $= \mathrm{B2}/55 \cdot 100$
- kaava kopioidaan sarakkeen D seuraaville riveille
- muutetaan sarakkeen D solujen pyöristystarkkuus prosentin tarkkuuteen

Lasketaan suhteelliset summafrekvenssit sarakkeeseen E. Suhteellinen summafrekvenssi lasketaan summaamalla siihenastiset suhteelliset frekvenssit:

- soluun E2 kirjoitetaan ensimmäinen suhteellinen frekvenssi eli laskukaava: = D2
- soluun E3 kirjoitetaan laskukaava: = E2 + D3
- kaava kopioidaan sarakkeen E seuraaville riveille
- muutetaan sarakkeen E solujen pyöristystarkkuus prosentin tarkkuuteen

	A	B	C	D	E
1	**Käyntikerrat (lkm)**	**f**	**sf**	**f%**	**sf%**
2	0	5	5	9	9
3	1	24	29	44	53
4	2	18	47	33	85
5	3	6	53	11	96
6	4	2	55	4	100

b) Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Käyntikertojen moodi on Mo = 1 (käyntikerta). Mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää arvon 50 %. Käyntikertojen mediaani on Md = 1 (käyntikerta).

53.

a) Laaditaan frekvenssijakauma taulukkolaskentaohjelmalla.

Taulukon ensimmäiselle riville kirjoitetaan sarakkeiden otsikot. Sarake A nimetään muuttujan mukaan: Lomamatkojen määrä vuodessa. Muuttujan arvo, eli käyntikertojen lukumäärä, vaihtelee välillä 0–5. Kirjoitetaan sarakkeeseen A muuttujan arvot 0–5.

Kirjoitetaan sarakkeeseen B kunkin muuttujan arvon frekvenssi.

Lasketaan summafrekvenssit sarakkeeseen C. Summafrekvenssi lasketaan summaamalla siihenastiset frekvenssit:

- soluun C2 kirjoitetaan ensimmäisen muuttujan arvon frekvenssi eli luku 3
- soluun C3 kirjoitetaan laskukaava: $= \text{C2} + \text{B3}$
- kaava kopioidaan sarakkeen C seuraaville riveille

Viimeisen muuttujan arvon 5 summafrekvenssi kertoo havaintoyksiköiden lukumäärän. Solusta C7 nähdään, että havaintoyksiköiden lukumäärä on 168.

Lasketaan suhteelliset frekvenssit sarakkeeseen D. Suhteellinen frekvenssi lasketaan jakamalla muuttujan frekvenssi havaintoyksiköiden lukumäärällä 168 ja ilmaisemalla osamäärä prosentteina:

- soluun D2 kirjoitetaan laskukaava: $= \text{B2}/168 \cdot 100$
- kaava kopioidaan sarakkeen D seuraaville riveille
- muutetaan sarakkeen D solujen pyöristystarkkuus prosentin tarkkuuteen

Lasketaan suhteelliset summafrekvenssit sarakkeeseen E. Suhteellinen summafrekvenssi lasketaan summaamalla siihenastiset suhteelliset frekvenssit:

- soluun E2 kirjoitetaan ensimmäinen suhteellinen frekvenssi eli laskukaava: = D2
- soluun E3 kirjoitetaan laskukaava: = E2 + D3
- kaava kopioidaan sarakkeen E seuraaville riveille
- muutetaan sarakkeen E solujen pyöristystarkkuus prosentin tarkkuuteen

	A	B	C	D	E
1	**Lomamatkojen määrä vuodessa**	**f**	**sf**	**f%**	**sf%**
2	0	3	3	2	2
3	1	56	59	33	35
4	2	42	101	25	60
5	3	34	135	20	80
6	4	21	156	13	93
7	5	12	168	7	100

b) Viimeisen muuttujan arvon 5 summafrekvenssi kertoo havaintoyksiköiden lukumäärän. Havaintoyksiköiden lukumäärä on 168, joten kyselyyn osallistui 168 asiakasta.

c) Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Lomamatkojen määrän moodi on Mo = 1 (lomamatka). Mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää arvon 50 %. Lomamatkojen määrän mediaani on Md = 2 (lomamatkaa).

54.

a) Laaditaan frekvenssijakauma taulukkolaskentaohjelmalla.

Taulukon ensimmäiselle riville kirjoitetaan sarakkeiden otsikot. Nimetään sarake A muuttujan mukaan: Kotitalouden koko (henkilömäärä). Muuttujan arvo, eli henkilömäärä, on luku välillä 1–4 tai ”yli 5”. Kirjoitetaan muuttujan arvot sarakkeeseen A.

Sarakkeeseen B kirjoitetaan kunkin muuttujan arvon frekvenssi.

Lasketaan summafrekvenssit sarakkeeseen C. Summafrekvenssi lasketaan summaamalla siihenastiset frekvenssit:

- soluun C2 kirjoitetaan ensimmäisen muuttujan arvon frekvenssi: frekvenssi voidaan kirjoittaa lukuna tai voidaan viitata soluun B2 kirjoittamalla laskukaava: $= \text{B2}$
- soluun C3 kirjoitetaan laskukaava: $= \text{C2} + \text{B3}$
- kopioidaan kaava sarakkeen C seuraaville riveille

Viimeisen muuttujan arvon ”yli 5” summafrekvenssi kertoo havaintoyksiköiden lukumäärän. Havaintoyksiköiden lukumäärä nähdään solusta C6. Lukumäärä on 2 594 981.

Lasketaan suhteelliset frekvenssit sarakkeeseen D. Suhteellinen frekvenssi lasketaan jakamalla muuttujan frekvenssi havaintoyksiköiden lukumäärällä 2 594 981 ja ilmaisemalla osamäärä prosentteina. Laskukaavassa voidaan tehdä soluviittaus soluun C6:

- soluun D2 kirjoitetaan laskukaava: = B2/C6 · 100
  Laskukaavassa esiintyvät dollarimerkit ($) solun C6 sarakekirjaimen ja rivinumeron edessä tarkoittavat, että jakajana on laskukaavassa aina solussa C6 oleva luku, eikä luku muutu, kun kaavaa kopioidaan seuraaville riveille. Havaintoyksiköiden lukumäärä voidaan kirjoittaa laskukaavaan myös lukuna 2 594 981 ilman soluviittausta.
- kopioidaan kaava sarakkeen D seuraaville riveille.
- muutetaan sarakkeen D solujen pyöristystarkkuus yhden desimaalin tarkkuuteen

Lasketaan suhteelliset summafrekvenssit sarakkeeseen E. Suhteellinen summafrekvenssi lasketaan summaamalla siihenastiset suhteelliset frekvenssit:

- soluun E2 kirjoitetaan ensimmäinen suhteellinen frekvenssi eli laskukaava: = D2
- soluun E3 kirjoitetaan laskukaava: = E2 + D3
- kaava kopioidaan sarakkeen E seuraaville riveille
- muutetaan sarakkeen E solujen pyöristystarkkuus yhden desimaalin tarkkuuteen

	A	B	C	D	E
1	**Kotitalouden koko (henkilömäärä)**	**f**	**sf**	**f%**	**sf%**
2	1	1 044 957	1 044 957	40,3	40,3
3	2	909 139	1 954 096	35,0	75,3
4	3	271 416	2 225 512	10,5	85,8
5	4	244 778	2 470 290	9,4	95,2
6	yli 5	124 691	2 594 981	4,8	100,0

b) Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Henkilömäärän moodi on Mo = 1 (henkilö). Mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää arvon 50 %. Henkilömäärän mediaani on Md = 2 (henkilöä).

c) Sinkkutaloudet vastaavat muuttujan arvoa 1, jonka suhteellinen frekvenssi on 40 %. Sinkkutalouksia oli siis 40 %.

55.

a) Lasketaan suhteelliset frekvenssit sekä ilman taulukkolaskentaohjelmaa että taulukkolaskentaohjelmaa apuvälineenä käyttäen.

Suhteellisten frekvenssien laskeminen ilman taulukkolaskentaohjelmaa.

Lasketaan havaintoyksiköiden lukumäärä summaamalla frekvenssit:

$$15 + 6 + 13 + 10 + 12 + 23 = 79.$$

Muuttujan arvon suhteellinen frekvenssi lasketaan jakamalla muuttujan frekvenssi havaintoyksiköiden lukumäärällä ja ilmaisemalla osamäärä prosentteina:

Ykkössijalla oleva asia	$\boldsymbol{f}$	$\boldsymbol{f\,\%}$
perhe	15	$\frac{15}{79} \cdot 100 = 18{,}987 \ldots \approx \mathbf{19}$
raha	6	$\frac{6}{79} \cdot 100 = 7{,}594 \ldots \approx \mathbf{8}$
harrastukset	13	$\frac{13}{79} \cdot 100 = 16{,}455 \ldots \approx \mathbf{16}$
vapaa-aika	10	$\frac{10}{79} \cdot 100 = 12{,}658 \ldots \approx \mathbf{13}$
menestyminen	12	$\frac{12}{79} \cdot 100 = 15{,}189 \ldots \approx \mathbf{15}$
ystävät	23	$\frac{23}{79} \cdot 100 = 29{,}113 \ldots \approx \mathbf{29}$

Suhteellisten frekvenssien laskeminen taulukkolaskentaohjelmalla:

Kirjoitetaan taulukon ensimmäiselle riville sarakkeiden otsikot. Nimetään sarake A muuttujan mukaan: Ykkössijalla oleva asia. Muuttuja saa arvoja ”perhe”, ”raha”, ”harrastukset”, ”vapaa-aika”, ”menestyminen” ja ”ystävät”. Kirjoitetaan muuttujan arvot sarakkeeseen A.

Kirjoitetaan sarakkeeseen B kunkin muuttujan arvon frekvenssi.

Lasketaan havaintoyksiköiden lukumäärä soluun B8 summaamalla soluissa B2–B7 olevat frekvenssit. Summan laskentatapa riippuu käytettävästä ohjelmasta: selvitä miten solujen summa lasketaan käyttämälläsi ohjelmalla.

Havaintoyksiköiden lukumäärä nähdään nyt solussa B8. Lukumäärä on 79.

Lasketaan suhteelliset frekvenssit sarakkeeseen C. Suhteellinen frekvenssi lasketaan jakamalla muuttujan frekvenssi havaintoyksiköiden lukumäärällä 79 ja ilmaisemalla osamäärä prosentteina:

- soluun C2 kirjoitetaan laskukaava: $= \mathrm{B}2/79 \cdot 100$
- laskukaava kopioidaan sarakkeen C seuraaville riveille
- muutetaan sarakkeen C solujen pyöristystarkkuus prosentin tarkkuuteen

	A	B	C
1	**Ykkössijalla oleva asia**	**f**	**f%**
2	perhe	15	19
3	raha	6	8
4	harrastukset	13	16
5	vapaa-aika	10	13
6	menestyminen	12	15
7	ystävät	23	29
8	**yhteensä:**	**79**	100

b) Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Ykkössijalla olevan asian moodi on Mo = ystävät. Se tarkoittaa, että tässä aineistossa suurin osa vastaajista pitää ystäviä tärkeimpänä asiana.

56.

Laaditaan frekvenssijakauma taulukkolaskentaohjelmalla.

Kirjoitetaan taulukon ensimmäiselle riville sarakkeiden otsikot. Nimetään sarake A muuttujan mukaan: Koulutustyyppi. Muuttuja saa arvoja ”lukio”, ”ammatillinen”, ”valmistava” ja ”10-luokka” sekä ”ei jatkanut”. Kirjoitetaan muuttujan arvot sarakkeeseen A.

Kirjoitetaan sarakkeeseen B kunkin muuttujan arvon frekvenssi. Opiskelijoista 2537 jäi sijoittumatta eli ei heti jatkanut koulutuksissa: arvon ”ei jatkopaikkaa” frekvenssi on 2 537.

Lasketaan havaintoyksiköiden lukumäärä, eli kaikkien 9. luokan päättäneiden opiskelijoiden lukumäärä, soluun B7 summaamalla soluissa B2–B6 olevat frekvenssit. Summan laskentatapa riippuu käytettävästä ohjelmasta: selvitä miten solujen summa lasketaan käyttämälläsi ohjelmalla.

Havaintoyksiköiden lukumäärä nähdään nyt solussa B7. Lukumäärä on 57 853.

Lasketaan suhteelliset frekvenssit sarakkeeseen C:

- soluun C2 kirjoitetaan laskukaava: $= \text{B2}/57853 \cdot 100$
- kopioidaan kaava sarakkeen C seuraaville riveille
- muutetaan sarakkeen C solujen pyöristystarkkuus prosentin tarkkuuteen

	A	B	C
1	**Koulutustyyppi**	**f**	**f%**
2	lukio	29 857	52
3	ammatillinen	24 357	42
4	valmistava	414	1
5	10-luokka	688	1
6	ei jatkopaikkaa	2 537	4
7	**yhteensä:**	**57 853**	

a) Peruskoulun vuonna 2014 päättäneiden lukumäärä, eli kaikkien havaintoyksiköiden lukumäärä laskettiin soluun B7: se on 57 853.

b) Lukiokoulutukseen jatkaneiden prosenttiosuus nähdään muuttujan arvon ”lukio” suhteellisesta frekvenssistä, joka on 52 %.

c) Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Koulutustyypin moodi on Mo = lukio.

57.

Tässä aineistossa tutkittavana ominaisuutena eli muuttujana on lyöntien lukumäärä per peli. Muuttujan arvo on välillä 40–46. Pelien lukumäärä ilmaisee kunkin muuttujan arvon frekvenssin.

Laaditaan frekvenssijakauma taulukkolaskentaohjelmalla.

	A	B	C	D	E
1	**Lyöntejä/peli**	**f**	**sf**	**f%**	**sf%**
2	40	19	19	28	28
3	41	2	21	3	31
4	42	14	35	21	51
5	43	5	40	7	59
6	44	9	49	13	72
7	45	13	62	19	91
8	46	6	68	9	100

a) Alle 45 lyönnin pelejä ovat ne pelit, joissa lyöntejä on 40–44. Näiden pelien lukumäärä saadaan muuttujan arvon 44 summafrekvenssistä, joka on 49 (peliä).

b) Korkeintaan 42 lyönnin pelejä ovat ne pelit, joissa lyöntejä on 40–42. Näiden pelien lukumäärä saadaan muuttujan arvon 42 summafrekvenssistä, joka on 35.

c) Yli 43 lyönnin pelejä ovat ne pelit, joissa lyöntejä on 44–46. Näiden pelien prosenttiosuus saadaan vähentämällä sadasta prosentista muuttujan arvon 43 suhteellinen summafrekvenssi:

$$100\ \% - 59\ \% = 41\ \%.$$

Tämä voidaan laskea myös summaamalla muuttujan arvojen 44–46 suhteelliset frekvenssit:

$$13\ \% + 19\ \% + 9\ \% = 41\ \%.$$

58.

a) Laaditaan jakaumat taulukkolaskentaohjelmalla.

	A	B	C	D	E
1	**Pistemäärä**	**f**	**sf**	**f%**	**sf%**
2	1	2	2	2	2
3	2	10	12	12	14
4	3	9	21	11	25
5	4	**21**	**42**	25	49
6	5	19	61	22	**72**
7	6	11	72	13	85
8	7	7	79	8	93
9	8	6	**85**	7	100

b) Viimeisen muuttujan arvon summafrekvenssi kertoo kaikkien osanottajien lukumäärän. Partiotaito-kilpailuun osallistui **85** osanottajaa.

c) Moodi on se pistemäärä, jonka frekvenssi on suurin. Suurin frekvenssi (**21**) oli pistemäärällä 4, joten Mo = 4 (pistettä).

Mediaani on se pistemäärä, jonka suhteellinen summafrekvenssi ylittää ensimmäisen kerran arvon 50 %. Pistemäärän 5 suhteellinen summafrekvenssi on **72** %, joten Md = 5.

d) Korkeintaan neljä pistettä saivat ne osallistujat, joiden pistemäärä oli 1–4. Näiden osallistujien lukumäärä saadaan muuttujan arvon 4 summafrekvenssistä, joka on **42** (osallistujaa).

e) Kaikkien osallistujien määrä prosentteina on 100 %. Osallistujia, jotka saivat korkeintaan 5 pistettä, oli **72** %. Osallistujia, jotka saivat vähintään 6 pistettä, on tällöin

$$100\ \% - 72\ \% = 28\ \%.$$

Tämä voidaan laskea myös summaamalla pistemäärän 6–8 saaneiden prosenttiosuudet yhteen:

$$13\ \% + 8\ \% + 7\ \% = 28\ \%.$$

59.

a) Laaditaan jakaumat taulukkolaskentaohjelmalla.

	A	B	C	D	E
1	**Tyytyväisyys-aste**	**f**	**sf**	**f%**	**sf%**
2	1	9	9	4,5	4,5
3	2	18	27	9,0	13,5
4	3	**65**	92	32,5	46,0
5	4	61	153	30,5	**76,5**
6	5	47	**200**	23,5	100,0

b) Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Tyytyväisyysasteen moodi on Mo = 3.

Mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää 50 %. Tyytyväisyysasteen mediaani on Md = 4.

c) Kaikkien kyselyyn osallistuneiden asiakkaiden määrä nähdään alimman muuttujan arvon summafrekvenssistä, joka on **200**.

d) ”Tyytyväinen” vastaa muuttujan arvoa 4. Vähintään tyytyväisiä ovat ne asiakkaat, jotka ovat antaneet arvion 4 tai 5. Näiden prosenttiosuus saadaan laskemalla yhteen muuttujan arvojen 4 ja 5 suhteelliset frekvenssit:

$$30{,}5\ \% + 23{,}5\ \% = 54{,}0\ \%.$$

Tämä voidaan laskea myös suhteellisten summafrekvenssin avulla: kaikkien kyselyyn osallistuneiden asiakkaiden määrä prosentteina on 100,0 %. Asiakkaat jotka ovat vähemmän kuin tyytyväisiä, ovat antaneet arvosanan 1–3, ja näiden prosenttiosuus on 46,0 %. Vähintään tyytyväisiä on siis

$$100{,}0\ \% - 46{,}0\ \% = 54{,}0\ \%.$$

e) ”Korkeintaan tyytymätön” vastaa muuttujan arvoja 1 ja 2. Näiden lukumäärä on

$$9 + 18 = 27.$$

60.

a) Käytetään ratkaisussa taulukkolaskentaohjelmaa. Kunkin luokan luokkakeskus saadaan laskemalla luokan ala- ja ylärajan keskiarvo:

- soluun D2 kirjoitetaan laskukaava $= (B2 + C2)/2$
- kopioidaan kaava sarakkeen D seuraaville riveille
- muutetaan sarakkeen D solujen pyöristystarkkuus yhden desimaalin tarkkuuteen

	A	B	C	D
1	**Luokka**	**alaraja**	**yläraja**	**luokkakeskus**
2	1-5	1	5	3
3	6-10	6	10	8
4	11-15	11	15	13
5	16-20	16	20	18
6	21-25	21	25	23

b) Laaditaan tarvittavat frekvenssijakaumat taulukkolaskentaohjelmistolla.

	A	B	C	D	E	F	G
1	**Luokka**	**alaraja**	**yläraja**	**luokkakeskus**	**f**	**sf**	**sf%**
2	1-5	1	5	3	3	3	9,7
3	6-10	6	10	8	5	8	25,8
4	11-15	11	15	13	10	18	58,1
5	16-20	16	20	18	12	30	96,8
6	21-25	21	25	23	1	31	100,0

Moodi on sen luokan luokkakeskus, jonka frekvenssi on suurin. Asiakasvirran moodi on Mo = 18 (asiakasta). Mediaani on sen luokan luokkakeskus, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää arvon 50 %. Asiakasvirran mediaani on Md = 13 (asiakasta).

c) Kunkin luokan frekvenssi ilmaisee seurantapäivien lukumäärää, joten frekvenssien summa ilmaisee seurantapäivien kokonaismäärän. Havaintoyksiköiden lukumäärä, eli seurantapäivien lukumäärä, nähdään viimeisen luokan summafrekvenssistä, joka on 31. Asiakasvirtaa seurattiin siis 31 päivänä.

d) Niiden päivien lukumäärä, jolloin kävi korkeintaan 15 asiakasta, saadaan luokan 11–15 summafrekvenssistä, joka on 18. Korkeintaan 15 asiakasta kävi 18 seurantapäivänä.

61.

Muuttujana on tässä aineistossa konkurssiin haetun rakennusalan yrityksen henkilökunnan määrä. Muuttuja saa arvot välillä 1–20. Muuttujan arvot on luokiteltu viiteen luokkaan.

a) Käytetään ratkaisussa taulukkolaskentaohjelmaa. Kunkin luokan luokkakeskus saadaan laskemalla luokan ala- ja ylärajan keskiarvo:

- soluun D2 kirjoitetaan laskukaava = (B2 + C2)/2
- kopioidaan kaava sarakkeen D seuraaville riveille
- muutetaan sarakkeen D solujen pyöristystarkkuus yhden desimaalin tarkkuuteen

	A	B	C	D
1	**Luokka**	**alaraja**	**yläraja**	**luokkakeskus**
2	1-4	1	4	2,5
3	5-8	5	8	6,5
4	9-12	9	12	10,5
5	13-16	13	16	14,5
6	17-20	17	20	18,5

b) Laaditaan tarvittavat frekvenssijakaumat taulukkolaskentaohjelmalla.

	A	B	C	D	E	F	G
1	**Luokka**	**alaraja**	**yläraja**	**luokkakeskus**	**f**	**sf**	**sf%**
2	1-4	1	4	2,5	46	46	48,9
3	5-8	5	8	6,5	23	69	73,4
4	9-12	9	12	10,5	17	86	91,5
5	13-16	13	16	14,5	7	93	98,9
6	17-20	17	20	18,5	1	94	100,0

Moodi on sen luokan luokkakeskus, jonka frekvenssi on suurin. Henkilökunnan määrän moodi on Mo = 2,5 (työntekijää). Mediaani on sen luokan luokkakeskus, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää arvon 50 %. Henkilökunnan määrän mediaani on Md = 6,5 (työntekijää).

c) Korkeintaan 12 henkilön yrityksien prosenttiosuus saadaan luokan 9-12 suhteellisesta summafrekvenssistä, joka on 91,5 % %. Konkurssiin haetuista yrityksistä 91,5 % oli henkilökunnan määrältään korkeintaan 12 henkilön yrityksiä.

62.

a) Alimman luokan luokkavälin pituus on $12 - 9 = 3$. Aineisto luokitellaan tasavälisiin luokkiin, joten jokaisen luokan luokkavälin pituus on 3. Luokan yläraja saadaan summaamalla luokan alarajaan luokkavälin pituus, eli luku 3.

Luokat ovat: 9–12, 13–16, 17–20, 21–24, 25–28.

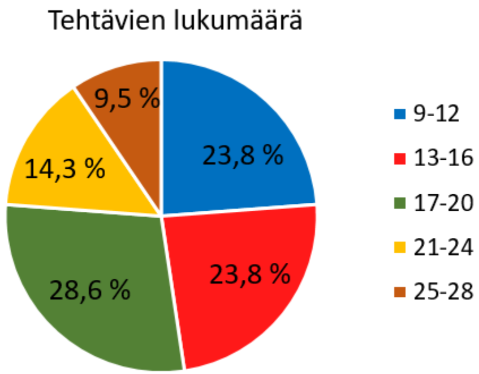
Suurin tehtävämäärä on 28, joten luku 28 kuuluu ylimpään luokkaan. Se on nyt myös ylimmän luokan yläraja.

Laaditaan frekvenssijakaumat taulukkolaskentaohjelmalla.

	A	B	C	D	E	F	G	H
1	**Luokka**	**alaraja**	**yläraja**	**luokkakeskus**	**f**	**sf**	**f%**	**sf%**
2	9-12	9	12	10,5	5	5	23,8	23,8
3	13-16	13	16	14,5	5	10	23,8	47,6
4	17-20	17	20	18,5	6	16	28,6	76,2
5	21-24	21	24	22,5	3	19	14,3	90,5
6	25-28	25	28	26,5	2	21	9,5	100,0

b) Moodi on sen luokan luokkakeskus, jonka frekvenssi on suurin. Tehtävien lukumäärän moodi on Mo = 18,5 tehtävää. Mediaani on sen luokan luokkakeskus, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää arvon 50 %. Tehtävien lukumäärän mediaani on Md = 18,5 tehtävää.

c) Sektoridiagrammi luodaan luokkien (sarake A) ja suhteellisten frekvenssien (sarake G) mukaan taulukkolaskentaohjelmalla.



63.

a) Alimman luokan luokkavälin pituus on $3 - 1 = 2$. Aineisto luokitellaan tasavälisiin luokkiin, joten jokaisen luokan luokkavälin pituus on 2. Luokan yläraja saadaan summaamalla luokan alarajaan luokkavälin pituus, eli luku 2.

Luokat ovat: 1–3, 4–6, 7–9, 10–12.

Suurin sijoitus on 10, ja se kuuluu ylimpään luokkaan. Se on nyt myös ylimmän luokan alaraja.

Laaditaan frekvenssijakaumat taulukkolaskentaohjelmalla.

	A	B	C	D	E	F	G	H
1	**Luokka**	**alaraja**	**yläraja**	**luokkakeskus**	**f**	**sf**	**f%**	**sf%**
2	1-3	1	3	2	3	3	19	19
3	4-6	4	6	5	6	9	38	56
4	7-9	7	9	8	5	14	31	88
5	10-12	10	12	11	2	16	13	100

b) Moodi on sen luokan luokkakeskus, jonka frekvenssi on suurin. Sijoituksen moodi on Mo = 5. sija. Mediaani on sen luokan luokkakeskus, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää arvon 50 %. Sijoituksen mediaani on Md = 5. sija.

c) Kolmen parhaan sijoituksen joukossa ovat ne sijoitukset, jotka kuuluvat alimpaan luokkaan 1 – 3. Alimman luokan prosenttiosuus nähdään sen suhteellisesta frekvenssistä, joka on 19 %.

64.

a) Alimman luokan luokkavälin pituus on $19 - 10 = 9$. Aineisto luokitellaan tasavälisiin luokkiin, joten jokaisen luokan luokkavälin pituus on 9. Luokan yläraja saadaan summaamalla luokan alarajaan luokkavälin pituus, eli luku 9.

Luokat ovat: 10–19, 20–29, 30–39, 40–49.

	A	B	C	D	E	F	G	H
1	**Luokka**	**alaraja**	**yläraja**	**luokkakeskus**	**f**	**sf**	**f%**	**sf%**
2	10-19	10	19	14,5	8	8	36	36
3	20-29	20	29	24,5	7	15	32	68
4	30-39	30	39	34,5	4	19	18	86
5	40-49	40	49	44,5	3	22	14	100

Laaditaan frekvenssijakaumat taulukkolaskentaohjelmalla.

b) Moodi on sen luokan luokkakeskus, jonka frekvenssi on suurin. Pelargonian päivittäisen menekin moodi on Mo = 14,5 (kappaletta). Mediaani on sen luokan luokkakeskus, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää arvon 50 %. Pelargonian päivittäisen menekin mediaani on Md = 24,5 (kappaletta).

c) Kunkin luokan frekvenssi ilmaisee päivien lukumäärää, joten frekvenssien summa ilmaisee päivien kokonaismäärän. Havaintoyksiköiden lukumäärä, eli päivien lukumäärä, nähdään viimeisen luokan summafrekvenssistä, joka on 22. Puutarhuri tilastoi pelargonian menekkiä siis 22 päivänä.

Tämä saadaan myös laskemalla havaintoyksiköiden lukumäärä aineistossa: frekvenssien summa on 22 eli päivien lukumäärä on 22.

65.

a) Laaditaan jakaumat taulukkolaskentaohjelmalla. Ilmaistaan suhteelliset frekvenssit prosentin kymmenesosan tarkkuudella.

	A	B	C	D	E
1	Tietokoneiden määrä	f	sf	f%	sf%
2	0	8	8	0,9	0,9
3	1	191	199	22,2	23,1
4	2	462	661	53,7	76,8
5	3	136	797	15,8	92,6
6	4	58	855	6,7	99,3
7	5	6	861	0,7	100,0

Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Suurin frekvenssi (462) oli arvolla 2, joten tietokoneiden määrän moodi on Mo = 2 (tietokonetta).

Mediaani on se muuttujan arvo, jonka suhteellinen summafrekvenssi ylittää ensimmäisen kerran arvon 50 %. Arvon 2 suhteellinen summafrekvenssi 76,8 % on ensimmäisen kerran yli 50 %, joten tietokoneiden määrän mediaani on Md = 2 (tietokonetta).

b) Korkeintaan kaksi tietokonetta on niissä kotitalouksissa, joissa määrä on 0–2. Näiden kotitalouksien lukumäärä saadaan muuttujan arvon 2 summafrekvenssistä, joka on 661.

c) ”Vähintään neljä tietokonetta” vastaa muuttujan arvoja 4 ja 5. Summataan näiden arvojen suhteelliset frekvenssit:

$$6{,}7\ \% + 0{,}7\ \% = 7{,}4\ \%.$$

Pyöristetään vastaus prosentin tarkkuuteen: niitä kotitalouksia, joissa on vähintään 4 tietokonetta, on 7 %.

Tämä voidaan laskea myös kertymien avulla. Kaikkien tutkimukseen osallistujien kotitalouksien määrä prosentteina on 100,0 %. Niitä kotitalouksia, joissa on korkeintaan kolme tietokonetta, on 92,6 %, joten niitä kotitalouksia joissa on vähintään neljä tietokonetta, on

$$100{,}0\ \% - 92{,}6\ \% = 7{,}4\ \%.$$

Niitä kotitalouksia, joissa on vähintään 4 tietokonetta, on 7 %.

66.

a) Alimman luokan luokkavälin pituus on $169 - 150 = 19$. Aineisto luokitellaan tasavälisiin luokkiin, joten jokaisen luokan luokkavälin pituus on 19. Luokan yläraja saadaan summaamalla luokan alarajaan luokkavälin pituus, eli luku 19.

Luokat ovat: 150–169, 170–189, 190–209, 210–229, 230–249.

Lasketaan luokkakeskukset taulukkolaskentaohjelmalla.

	A	B	C	D
1	**Luokka**	**alaraja**	**yläraja**	**luokkakeskus**
2	150-169	150	169	159,5
3	170-189	170	189	179,5
4	190-209	190	209	199,5
5	210-229	210	229	219,5
6	230-249	230	249	239,5

b) Laaditaan frekvenssijakaumat taulukkolaskentaohjelmalla.

	A	B	C	D	E	F	G	H
1	**Sanamäärä**	**alaraja**	**yläraja**	**luokkakeskus**	**f**	**sf**	**f%**	**sf%**
2	150-169	150	169	159,5	4	4	20	20
3	170-189	170	189	179,5	3	7	15	35
4	190-209	190	209	199,5	4	11	20	55
5	210-229	210	229	219,5	4	15	20	75
6	230-249	230	249	239,5	5	20	25	100

Moodi on sen luokan luokkakeskus, jonka frekvenssi on suurin. Ylimmällä luokalla 230–249 on suurin frekvenssi. Sanamäärän moodi on Mo = 239,5 sanaa. Mediaani on sen luokan luokkakeskus, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää arvon 50 %. Luokan 190–229 suhteellinen summafrekvenssi on 55 %, ja se on ensimmäinen luokka, jolla arvo on yli 50 %. Sanamäärän mediaani on Md = 199,5 sanaa.

c) Pituudeltaan vähintään 210 sanaa olevat aineet kuuluvat kahteen ylimpään luokkaan. Näiden prosenttiosuus saadaan summaamalla kahden ylimmän luokan suhteelliset frekvenssit:

$$20\ \% + 25\ \% = 45\ \%.$$

Tämä voidaan laskea myös kertymien avulla. Kaikkien aineiden osuus prosentteina on 100 %. Pituudeltaan korkeintaan 209 sanaa olevien aineiden prosenttiosuus saadaan luokan 190 – 209 suhteellisesta summafrekvenssistä, joka on 55 %. Pituudeltaan vähintään 210 sanaa olevia aineita oli siis

$$100\ \% - 55\ \% = 45\ \%.$$

67.

Laaditaan tarvittavat frekvenssijakaumat taulukkolaskentaohjelmalla.

	A	B	C	D	E	F	G
1	**Huoneiden lukumäärä**	**alaraja**	**yläraja**	**luokkakeskus**	**f**	**sf**	**sf%**
2	0-29	0	29	14,5	20	20	11
3	30-59	30	59	44,5	31	51	28
4	60-89	60	89	74,5	42	93	51
5	90-119	90	119	104,5	56	149	82
6	120-149	120	149	134,5	21	170	93
7	150-179	150	179	164,5	10	180	99
8	180-209	180	209	194,5	2	182	100

a) Moodi on sen luokan luokkakeskus, jonka frekvenssi on suurin. Luokan 90–119 frekvenssi on suurin, joten huonemäärän moodi on Mo = 104,5 (huonetta). Mediaani on sen luokan luokkakeskus, jonka suhteellinen summafrekvenssi ensimmäisen kerran ylittää arvon 50 %. Tässä aineistossa se on luokka 60–89, joten huonemäärän mediaani on Md = 74,5 (huonetta).

b) Kunkin luokan frekvenssi ilmaisee majoitusliikkeiden lukumäärää, joten frekvenssien summa ilmaisee tutkimukseen osallistuneiden majoitusliikkeiden kokonaismäärän. Havaintoyksiköiden lukumäärä, eli majoitusliikkeiden lukumäärä, nähdään viimeisen luokan summafrekvenssistä, joka on 182. Alueella on siis yhteensä 182 majoitusliikettä.

c) Alle 120 huoneen majoitusliikkeiden lukumäärä saadaan luokan 90–119 summafrekvenssistä, joka on 149. Alle 120 huonetta on 149 majoitusliikkeessä.

d) Kaikkien alueen majoitusliikkeiden osuus prosentteina on 100 %. Alle 90 huoneen majoitusliikkeiden prosenttiosuus saadaan luokan 60–89 suhteellisesta summafrekvenssistä, joka on 51 %. Vähintään 90 huoneen majoitusliikkeitä on siis

$$100\ \% - 51\ \% = 49\ \%.$$

Vähintään 90 huonetta on 49 prosentissa alueen majoitusliikkeistä.

68.

Sektoridiagrammi luodaan muuttujan arvojen ja suhteellisten frekvenssien mukaan. Laaditaan frekvenssijakauma taulukkolaskentaohjelmalla.

Kirjoitetaan taulukon ensimmäiselle riville sarakkeiden otsikot. Muuttujana on pistemäärä, jonka arvo on välillä 0–6. Nimetään sarake A muuttujan mukaan: Pistemäärä. Kirjoitetaan muuttujan arvot sarakkeeseen A.

Pylvään korkeus kertoo kunkin muuttujan arvon frekvenssin. Kirjoitetaan frekvenssit sarakkeeseen B.

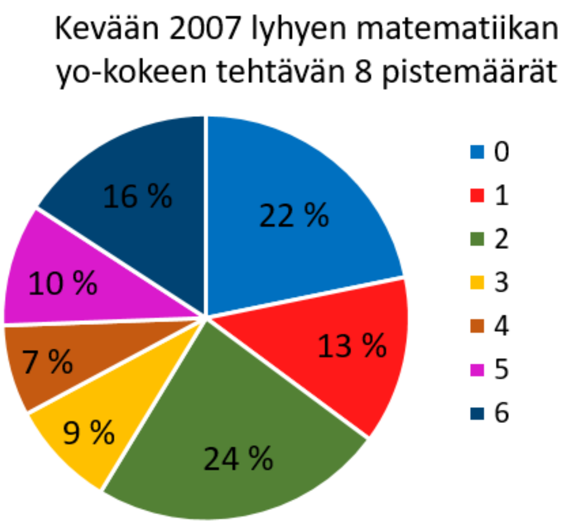
Lasketaan havaintoyksiköiden lukumäärä soluun B9 summaamalla soluissa B2–B8 olevat frekvenssit. Havaintoyksiköiden lukumäärä nähdään nyt solussa B9: lukumäärä on 9300. Kokeeseen osallistui siis 9300 kokelasta.

Lasketaan suhteelliset frekvenssit sarakkeeseen C. Suhteellinen frekvenssi lasketaan jakamalla muuttujan frekvenssi havaintoyksiköiden lukumäärällä 9300 ja ilmaisemalla osamäärä prosentteina:

- soluun C2 kirjoitetaan laskukaava: $= \mathrm{B2}/9300 \cdot 100$
- kopioidaan kaava sarakkeen C seuraaville riveille
- muutetaan sarakkeen C solujen pyöristystarkkuus prosentin tarkkuuteen

	A	B	C
1	**Luokka**	**f**	**f%**
2	0	2025	22
3	1	1245	13
4	2	2186	24
5	3	792	9
6	4	673	7
7	5	908	10
8	6	1471	16
9	**yhteensä:**	**9300**	

Sektoridiagrammi luodaan luokkien (sarake A) ja suhteellisten frekvenssien (sarake C) mukaan. Sektorin keskuskulma määräytyy suhteellisen frekvenssin mukaan. Muokataan otsikko aineistoon sopivaksi.



1 Tilastot

1.4 Tunnuslukujen laskeminen

69.

Kirjoitetaan muuttujan arvot taulukkolaskentaohjelman tai laskinsovelluksen taulukkoon. Kirjoitetaan jokainen arvo omalle riville.

Määritetään tunnusluvut tilastotoimintojen avulla:

- mediaani on
  Md = 115 479,5 ≈ 115 480 (autoa)
- keskiarvo on
  $\bar{x} = 119\,720{,}25\ldots \approx 119\,720$ (autoa)
- keskihajonta on
  $s_n = 17\,084{,}74\ldots \approx 17\,085$ (autoa)
- otoskeskihajonta on
  $s_{n-1} = 17\,844{,}43\ldots \approx 17\,844$ (autoa)

	A
1	148161
2	145700
3	125608
4	139669
5	90574
6	111968
7	126123
8	111251
9	103450
10	106236
11	108912
12	118991

70.

a) Kirjoitetaan Anssin heittotulokset taulukkolaskentaohjelman tai laskinsovelluksen taulukkoon. Lukujen järjestyksellä taulukossa ei ole väliä, mutta kirjoitetaan jokainen arvo omalle rivilleen.

	A
1	8
2	6
3	6
4	8
5	9
6	9
7	9
8	6
9	8
10	5
11	10
12	3
13	7
14	4
15	5
16	9
17	10
18	9

Määritetään Anssin heittotulosten keskiarvo tilastotoimintojen avulla:

$$\bar{x} = 7{,}277 \ldots \approx 7{,}3.$$

Akun heittotulosten keskiarvo 8,1 on suurempi, joten tämän perusteella Aku voittaa kilpailun.

b) Määritetään Anssin heittotulosten keskihajonta tilastotoimintojen avulla. Kyseessä on kokonaisaineisto (ei otos), joten käytetään keskihajontaa s_n.

Anssin heittotulosten keskihajonta on

$$s_n = 2{,}049 \ldots \approx 2{,}0.$$

Keskihajonta kuvaa muuttujan arvojen hajaantumista keskiarvon ympärille. Akun heittotulosten keskihajonta 1,5 on pienempi, joten Aku heitti tasaisemman heittosarjan.

71.

a) Lasketaan yrityksen liikevoiton keskiarvo ja keskihajonta. Kirjoitetaan vuosittaiset liikevoitot taulukkolaskentaohjelman tai laskinsovelluksen taulukkoon ja määritetään tunnusluvut tilastotoimintojen avulla:

- keskiarvo on $\bar{x} = 2{,}552 \approx 2{,}55$ (milj. euroa)
- keskihajonta on $s_n = 0{,}2505\ldots \approx 0{,}25$ (milj. euroa)
- otoskeskihajonta on $s_{n-1} = 0{,}2801\ldots \approx 0{,}28$ (milj. euroa)

Tutkimusaineisto on kokonaisaineisto, joten käytetään keskihajontaa s_n.

Yrityksen keskimääräinen liikevoitto vuosina 2010–2014 on 2,55 miljoonaa euroa ja liikevoiton keskihajonta on 0,25 miljoonaa euroa.

b) Täydennetään a-kohdan taulukkolaskentaohjelman tai laskinsovelluksen taulukkoa vuosien 2015 ja 2016 liikevoitoilla. Lasketaan keskiarvo aikavälille 2010–2017. Uusi keskiarvo on $\bar{x} = 2{,}675\ldots$ miljoonaa euroa.

Verrataan uutta keskiarvoa vanhaan keskiarvoon ja ilmaistaan osamäärä prosentteina. Käytetään laskussa tarkkoja arvoja eli tehdään lasku taulukkolaskennan soluviittauksilla tai laskinsovelluksessa kopioimalla määritettyjen keskiarvojen tarkat arvot laskukenttään.

$$\frac{2{,}675\ldots}{2{,}552} \cdot 100 = 104{,}847\ldots \approx 104{,}8.$$

Yrityksen keskimääräinen liikevoitto aikavälillä 2010–2017 on siis $104{,}8\ \% - 100\ \% = 4{,}8\ \%$ suurempi kuin aikaisemmin laskettu keskimääräinen liikevoitto.

c) Määritetään liikevoiton keskihajonta aikavälillä 2010–2017 tilastotoimintojen avulla. Uusi keskihajonta on $s_n = 0{,}2888\ldots$ (milj. euroa).

Verrataan uutta keskihajontaa vanhaan keskihajontaan ja ilmaistaan osamäärä prosentteina. Käytetään laskussa tarkkoja arvoja eli tehdään lasku taulukkolaskennan soluviittauksilla tai laskinsovelluksessa kopioimalla määritettyjen keskihajontojen tarkat arvot laskukenttään.

$$\frac{0{,}28889}{0{,}25055} \cdot 100 = 115{,}302\ldots \approx 115{,}3.$$

Liikevoiton keskihajonta suurenee siis $115{,}3\ \% - 100\ \% = 15{,}3\ \%$. Ilmaistaan vastaus kahden merkitsevän numeron tarkkuudella: liikevoiton keskihajonta suurenee 15 %.

72.

Kopioidaan salamatiheydet taulukkolaskentaohjelman tai laskinsovelluksen taulukkoon tai kopioidaan valmis aineisto työpöydän kautta.

a) Määritetään keskiarvo tilastotoimintojen avulla: maasalamatiheyden keskiarvo on $\bar{x} = 10{,}6$ salamaa/100 km^2.

b) Muuttujan vaihteluväli kertoo, mille välille havainnot asettuvat. Se ilmaistaan ilmoittamalla muuttujan pienin arvo ja suurin arvo. Useimmat tilastotoiminnot kertovat havaintoyksiköiden suurimman ja pienimmän arvon samassa yhteydessä muiden tilastollisten tunnuslukujen kanssa. Maasalamatiheyksien

- pienin arvo on 3,
- suurin arvo on 21.

Maasalamatiheyden vaihteluväli on 3–21 salamaa/100 km^2.

c) Määritetään keskihajonta tilastotoimintojen avulla: keskihajonta on $s_n = 4{,}454\ldots$

Muuttujan arvo poikkeaa merkitsevästi keskiarvosta, jos se on yli kahden keskihajonnan päässä keskiarvosta. Tällöin normitetun arvon itseisarvo on suurempi kuin kaksi.

Pienin maasalamatiheys on 3. Lasketaan pienintä arvoa $x = 3$ vastaava normitettu arvo. Käytetään laskussa tarkkoja arvoja: tehdään lasku taulukkolaskennan soluviittauksilla tai laskinsovelluksessa kopioimalla määritettyjen tunnuslukujen tarkat arvot laskukenttään.

Normitettu arvo on

$$z = \frac{x - \bar{x}}{s_n} = \frac{3 - 10{,}6}{4{,}454\ldots} = -1{,}706\ldots$$

Normitetun arvon itseisarvo on $|z| = 1{,}706\ldots < 2$, joten aineiston pienin arvo ei poikkea merkitsevästi keskiarvosta.

73.

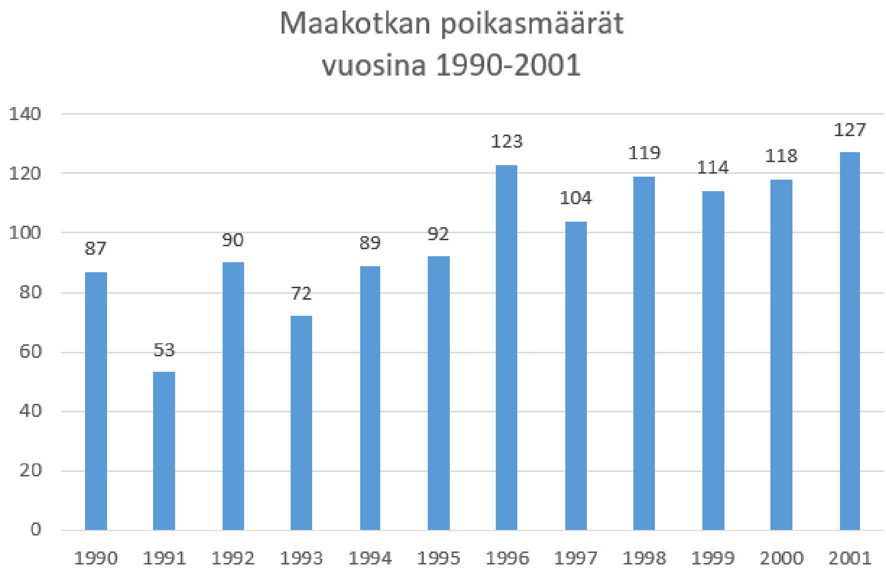
Kirjoitetaan taulukkolaskentaohjelman taulukkoon sarakkeeseen A vuosiluvut ja sarakkeeseen B muuttujan arvot, eli vuotuiset poikasmäärät.

	A	B
1	1990	87
2	1991	53
3	1992	90
4	1993	72
5	1994	89
6	1995	92
7	1996	123
8	1997	104
9	1998	119
10	1999	114
11	2000	118
12	2001	127

Lasketaan poikasmäärän keskiarvo ja keskihajonta tilastotoiminnoilla:

- keskiarvo on $\bar{x} = 99$ (poikasta/vuosi)
- keskihajonta on $s_n = 21{,}583 \ldots \approx 21{,}6$ (poikasta/vuosi)

Pylväsdiagrammi piirretään vuosilukusarakkeen ja lukumäärien mukaan taulukkolaskentaohjelmalla.



Muuttujan arvo poikkeaa merkitsevästi keskiarvosta, jos se on yli kahden keskihajonnan päässä keskiarvosta. Tällöin normitetun arvon itseisarvo on suurempi kuin kaksi.

Lasketaan vuoden 1991 poikasmäärän $x = 53$ normitettu arvo. Käytetään laskussa tarkkoja arvoja eli tehdään laskut taulukkolaskennan soluviittauksilla tai laskinsovelluksessa kopioimalla määritettyjen tunnuslukujen tarkat arvot laskukenttään. Normitettu arvo on

$$z_{1991} = \frac{x - \bar{x}}{s_n} = \frac{53 - 99}{21{,}583\ldots} = -2{,}131\ldots$$

Normitetun arvon itseisarvo on $|z| = 2{,}131\ldots > 2$, joten vuoden 1991 poikasmäärä poikkeaa merkitsevästi keskiarvosta.

Lasketaan vuoden 2001 poikasmäärän $x = 127$ normitettu arvo. Käytetään laskussa tarkkoja arvoja eli tehdään lasku taulukkolaskennan soluviittauksilla tai laskinsovelluksessa kopioimalla määritettyjen tunnuslukujen tarkat arvot laskukenttään. Normitettu arvo on

$$z_{2001} = \frac{x - \bar{x}}{s_n} = \frac{127 - 99}{21{,}583\ldots} = 1{,}297\ldots$$

Normitetun arvon itseisarvo on $|z| = 1{,}297\ldots < 2$, joten vuoden 2001 poikasmäärä ei poikkea merkitsevästi keskiarvosta.

Tehtävä voidaan ratkaista myös laskemalla ne poikasmäärät, jotka poikkeavat tasan kahden keskihajonnan verran keskiarvosta $\bar{x} = 99$ (poikasta):

- keskiarvosta kaksi keskihajontaa alaspäin oleva poikasmäärä on:

$$\bar{x} - 2s_n = 99 - 2 \cdot 21{,}583\ldots = 55{,}833\ldots \approx 56.$$

- keskiarvosta kaksi keskihajontaa ylöspäin oleva poikasmäärä on:

$$\bar{x} + 2s_n = 99 + 2 \cdot 21{,}583\ldots = 142{,}166\ldots \approx 142.$$

Vuoden 1991 poikasmäärä on $53 < 56$, eli se on yli kahden keskihajonnan päässä keskiarvosta. Vuoden 1991 poikasmäärä poikkeaa merkitsevästi poikasmäärän keskiarvosta.

Vuoden 2001 poikasmäärä 127 on välillä 56–142, eli se on alle kahden keskihajonnan päässä keskiarvosta. Vuoden 2001 poikasmäärä ei poikkea merkitsevästi poikasmäärän keskiarvosta.

74.

Kopioidaan annettu valmis aineisto taulukkolaskentaohjelman tai laskinsovelluksen taulukkoon. Kopioidaan vuoden 2013 aineisto omaan sarakkeeseen ja vuoden 2014 aineisto omaan sarakkeeseen.

	A	B
1	vuosi 2013	vuosi 2014
2	113284	25324
3	2266	27576
4	91983	85913
5	99323	22217
6	33360	3700
7	59250	3156
8	15410	23833
9	192092	6366
10	2514	5083
11	4685	838

Tunnusluvut saadaan tilastotoiminnoilla.

a) Määritetään katsojamäärien keskiarvo kummallekin vuodelle:

- vuoden 2013 katsojamäärien keskiarvo on $\bar{x}_{2013} = 37\ 843$ (katsojaa)
- vuoden 2014 katsojamäärien keskiarvo on $\bar{x}_{2014} = 33\ 361$ (katsojaa)

Vuoden 2013 keskiarvo on suurempi, joten ensi-iltaelokuvien keskimääräinen katsojamäärä oli suurempi vuonna 2013.

b) Määritetään katsojamäärien mediaani kummallekin vuodelle:

- vuoden 2013 katsojamäärien mediaani on
  $\mathrm{Md} = 37\ 843$ (katsojaa)
- vuoden 2014 katsojamäärien mediaani on
  $\mathrm{Md} = 10\ 962$ (katsojaa)

Vuoden 2013 mediaani on suurempi.

c) Määritetään vuoden 2013 pienin katsojamäärä: Min = 0 (katsojaa).

Muuttujan arvo poikkeaa merkitsevästi keskiarvosta, jos se on yli kahden keskihajonnan päässä keskiarvosta. Tällöin normitetun arvon itseisarvo on suurempi kuin kaksi.

Lasketaan minimiarvoa vastaava normitettu arvo. Laskua varten tarvitaan vuoden 2013 katsojamäärien keskiarvo $\bar{x} = 37\ 843$ sekä keskihajontaa:

- vuoden 2013 katsojamäärien keskihajonta on
  $s_n = 60\ 619{,}39\ \ldots$ (katsojaa)

Käytetään normitetun arvon laskemisessa tarkkoja arvoja eli tehdään lasku taulukkolaskennan soluviittauksilla tai laskinsovelluksessa kopioimalla määritettyjen tunnuslukujen tarkat arvot laskukenttään.

Miniarvon $x = 0$ normitettu arvo on

$$z = \frac{x - \bar{x}}{s_n} = \frac{0 - 37\ 843}{60\ 619{,}39\ \ldots} = -0{,}624\ \ldots$$

Normitetun arvon itseisarvo on $|z| = 0{,}624\ \ldots < 2$, joten vuoden 2013 pienin katsojamäärä ei poikkea merkitsevästi keskiarvosta.

d) Määritetään vuoden 2014 suurin katsojamäärä: Max = 457 334 (katsojaa).

Muuttujan arvo poikkeaa merkitsevästi keskiarvosta, jos se on yli kahden keskihajonnan päässä keskiarvosta. Tällöin normitetun arvon itseisarvo on suurempi kuin kaksi.

Lasketaan maksimiarvoa vastaava normitettu arvo. Laskua varten tarvitaan vuoden 2014 katsojamäärien keskiarvo $\bar{x} = 33\ 361$ sekä keskihajontaa:

- vuoden 2014 katsojamäärien keskihajonta on $s_n = 56\ 440{,}04\ \ldots$ (katsojaa)

Käytetään normitetun arvon laskemisessa tarkkoja arvoja eli tehdään lasku taulukkolaskennan soluviittauksilla tai laskinsovelluksessa kopioimalla määritettyjen tunnuslukujen tarkat arvot laskukenttään.

Maksimiarvon $x = 457\ 334$ normitettu arvo on

$$z = \frac{x - \bar{x}}{s_n} = \frac{457\ 334\ - 33\ 361}{56\ 440{,}04\ \ldots} = 7{,}511\ \ldots$$

Normitetun arvon itseisarvo on $|z| = 7{,}511\ \ldots > 2$, joten vuoden 2014 suurin katsojamäärä poikkeaa merkitsevästi keskiarvosta.

75.

a) Merkitään Lämärin viimeisessä ottelussa tekemien maalien lukumäärää kirjaimella *x.* Lämärin kaikissa kuudessa ottelussa tekemien maalien summa on tällöin

$$5+3+8+6+4+x=x+26$$

ja maalien keskiarvo on $\frac{x+26}{6}$.

Tiedetään, että keskiarvo on 5,5 joten saadaan yhtälö:

$$\frac{x+26}{6}=5{,}5.$$

Ratkaistaan yhtälöstä tuntematon lukumäärä *x:*

$$\begin{aligned}\frac{x+26}{6}&=5{,}5 \quad |\cdot 6\\ x+26&=33\\ x&=33-26=7.\end{aligned}$$

Lämäri teki viimeisessä ottelussa 7 maalia.

b) Merkitään Vedon kahdessa viimeisessä ottelussa tekemien maalien lukumäärää kirjaimella *x.* Vedon kaikissa kuudessa ottelussa tekemien maalien summa on tällöin

$$5 + 3 + 8 + 6 + x + x = 2x + 22$$

ja maalien keskiarvo on $\frac{2x+22}{6}$.

Tiedetään, että keskiarvo on 5,0 joten saadaan yhtälö:

$$\frac{2x + 22}{6} = 5{,}0.$$

Ratkaistaan yhtälöstä tuntematon lukumäärä *x:*

$$\begin{aligned} \frac{2x + 22}{6} &= 5{,}0 \quad |\cdot 6 \\ 2x + 22 &= 30 \\ 2x &= 8 \\ x &= \frac{8}{2} = 4. \end{aligned}$$

Veto teki kummassakin ottelussa 4 maalia.

c) Määritetään joukkueiden turnauksessa tekemien maalien keskihajonnat laskimen tai taulukkolaskentaohjelman tilastotoiminnolla. Kirjoitetaan sarakkeeseen A Lämärin tekemien maalien lukumäärät ja sarakkeeseen B Vedon tekemien maalien lukumäärät.

	A	B
1	Lämäri	Veto
2	5	5
3	3	3
4	8	8
5	6	6
6	4	4
7	7	4

Määritetään Lämärin tekemien maalien keskihajonta: keskihajonta on $s_n = 1{,}707\ \ldots$
Määritetään Vedon tekemien maalien keskihajonta: keskihajonta on $s_n = 1{,}632\ \ldots$

Verrataan Vedon tekemien maalien keskihajontaa Lämärin tekemien maalien keskihajontaan ja ilmaistaan osamäärä prosentteina. Käytetään laskussa tarkkoja arvoja eli tehdään lasku taulukkolaskennan soluviittauksilla tai laskinsovelluksessa kopioimalla määritettyjen keskihajontojen tarkat arvot laskukenttään.

$$\frac{1{,}632\ \ldots}{1{,}707\ \ldots} \cdot 100\ \% = 95{,}618\ \ldots\ \% \approx 95{,}6\ \%.$$

Vedon tekemien maalien keskihajonta oli siis $100\ \% - 95{,}6\ \% = 4{,}4\ \%$ pienempi Lämärin tekemien maalien keskihajontaan verrattuna.

76.

a) Merkitään kymmenen sentin kolikon massaa kirjaimella x (grammaa).

Kaikkien kolikoiden massojen summa on tällöin

$8{,}50 + 7{,}50 + 7{,}80 + x = x + 23{,}80$ (grammaa) ja massojen keskiarvo on $\frac{x+23{,}80}{4}$ (grammaa).

Tiedetään, että keskiarvo on 6,975 grammaa, joten saadaan yhtälö:

$$\frac{x + 23{,}80}{4} = 6{,}975.$$

Ratkaistaan yhtälöstä tuntematon massa *x:*

$$\begin{aligned} \frac{x + 23{,}80}{4} &= 6{,}975 \quad \mid \cdot 4 \\ x + 23{,}80 &= 27{,}90 \\ x &= 27{,}90 - 23{,}80 = 4{,}10. \end{aligned}$$

Kymmenen sentin kolikon massa on siis 4,10 grammaa.

b) Merkitään kukkaroon lisätyn kolikon massaa kirjaimella x (grammaa).

Kaikkien kolikoiden massojen summa on tällöin

$8{,}50 + 7{,}50 + 7{,}80 + x + x = 2x + 23{,}80$ (grammaa) ja massojen keskiarvo on $\frac{2x+23{,}80}{5}$ (grammaa).

Tiedetään, että keskiarvo on 7,76 grammaa, joten saadaan yhtälö:

$$\frac{2x + 23{,}80}{5} = 7{,}76.$$

Ratkaistaan yhtälöstä tuntematon massa *x:*

$$\begin{aligned} \frac{2x + 23{,}80}{5} &= 7{,}76 \quad \mid \cdot 5 \\ 2x + 23{,}80 &= 38{,}80 \\ 2x &= 15 \\ x &= \frac{15}{2} = 7{,}50. \end{aligned}$$

Kukkaroon lisätyn kolikon massa on 7,50 grammaa, joten kyseessä on euron kolikko.

Kukkaroon lisättiin siis kaksi euron kolikkoa. Kukkarossa on lisäyksen jälkeen yksi kahden euron kolikko, kolme euron kolikkoa sekä yksi 50 sentin kolikko. Kukkarossa on rahaa yhteensä

$$2 \text{ eur} + 3 \cdot 1 \text{ eur} + 0{,}50 \text{ eur} = 5{,}50 \text{ eur}.$$

77.

a) ”Vain kuljettaja” vastaa muuttujan arvoa 0, jonka frekvenssi on 85. Vain kuljettaja on 85 autossa.

b) Moodi on se muuttujan arvo, jonka frekvenssi on suurin. Moodi on Mo = 0 (matkustajaa).

Muut tilastolliset tunnusluvut saadaan myös taulukkolaskentaohjelman tai laskinsovelluksen tilastotoiminnoilla. Käytetään tilastotoimintoja tunnustulukujen määrittämiseen.

Syötetään muuttujan arvot 0–5 sekä näiden frekvenssit laskentataulukkoon. Määritetään tunnusluvut:

- mediaani on Md = 1 (matkustaja)
- keskiarvo on $\bar{x} = 1{,}386\ldots \approx 1{,}4$ (matkustajaa)
- keskihajonta on $s_n = 1{,}275\ldots \approx 1{,}3$ (matkustajaa)
- otoskeskihajonta on $s_{n-1} = 1{,}277\ldots \approx 1{,}3$ (matkustajaa)

c) Kaikkien tutkimuksessa mukana olleiden henkilöautojen lukumäärä on sama kuin aineiston havaintoyksiköiden lukumäärä. Useimmat tilastotoiminnot kertovat havaintoyksiköiden lukumäärän samassa yhteydessä tilastollisten tunnuslukujen kanssa. Havaintoyksiköiden lukumäärä on usein merkitty kirjaimella *n.*

Tutkimuksessa oli mukana yhteensä $n = 272$ henkilöautoa.

Havaintoyksiköiden lukumäärä saadaan myös laskemalla frekvenssien summa:

$$85 + 71 + 65 + 35 + 9 + 7 = 272.$$

78.

a) Keskiarvo ja keskihajonta saadaan taulukkolaskentaohjelman tai laskinsovelluksen tilastotoiminnoilla.

Syötetään muuttujan arvot 10–4 sekä näiden frekvenssit laskentataulukkoon. Määritetään tunnusluvut:

- keskiarvo on $\bar{x} = 6{,}207\ldots \approx 6{,}21$
- keskihajonta on $s_n = 1{,}373\ldots \approx 1{,}37$

b) Ylimpien arvosanojen 10 ja 9 frekvenssit ovat nollia, joten kukaan opiskelija ei saanut ylimpiä arvosanoja. Koulun A paras arvosana oli 8.

c) Koulun B arvosanojen keskiarvo $\bar{x}_B = 6{,}44$ on suurempi kuin koulun A arvosanojen keskiarvo $\bar{x}_A = 6{,}21$, joten koulu B menestyi kokeessa keskimääräisesti paremmin.

d) Keskihajonta kuvaa muuttujan arvojen hajaantumista keskiarvon ympärille. Tulokset ovat tasaisimmat silloin, kun keskihajonta on pienin. Koulun A arvosanojen keskihajonta $s_n = 1{,}37$ on pienempi kuin koulun B arvosanojen keskihajonta $s_n = 2{,}15$, joten koulun A tulokset ovat tasaisemmat.

e) Kaikkia arvosanoja nostetaan kahdella numerolla, joten koulun A uudet tulokset ovat:

Koulu A	
Arvosana	***f***
10	20
9	32
8	14
7	19
6	16
5	0
4	0

Määritetään uudet tunnusluvut tilastotoiminnoilla:

- keskiarvo on $\bar{x} = 8{,}207 \ldots \approx 8{,}21$
- keskihajonta on $s_n = 1{,}373 \ldots \approx 1{,}37$

Siis myös keskiarvo nousee kahdella numerolla. Keskihajonta kuvaa muuttujan arvojen hajaantumista keskiarvon ympärille, joten keskihajonta ei muutu.

79.

Valmis aineisto on annettu frekvenssitaulukkoina vuosittain. Vuosien 1990 ja 2015 tiedot ovat seuraavat:

	Vuosi 1990	**Vuosi 2015**
Henkilöiden lukumäärä	**Frekvenssi**	**Frekvenssi**
1	646 229	1 112 342
2	597 928	874 880
3	332 295	285 433
4	300 429	234 939
5	112 714	87 506
6	30 596	24 101

Tunnusluvut voidaan määrittää joko tehtävänannossa annetun aineiston Excel-taulukossa tai kopioida frekvenssitaulukot toisen taulukkolaskentaohjelman tai laskinsovelluksen taulukkoon ja määrittää tunnusluvut tilastotoiminnoilla.

Excel-ohjelmassa tunnusluvut määritetään sopivilla laskukaavakomennoilla ja soluviittauksilla. Komennot vaihtelevat Excelin kielen ja ohjelmistoversion mukaan: jos käytät Exceliä, selvitä käyttämäsi version toiminnot.

a) Määritetään asuntokunnan koon keskiarvo kummallekin vuodelle frekvenssitaulukoiden avulla:

- Vuonna 1990 asuntokunnan koon keskiarvo oli $\bar{x}_{1990} = 2{,}3699\ldots \approx 2{,}4$ henkilöä.
- Vuonna 2015 asuntokunnan koon keskiarvo oli $\bar{x}_{2015} = 2{,}0007\ldots \approx 2{,}0$ henkilöä.

b) Määritetään asuntokunnan koon keskihajonta kummallekin vuodelle frekvenssitaulukoiden avulla:

- vuonna 1995 asuntokunnan koon keskihajonta oli $s_n = 1{,}298\ldots \approx 1{,}3$ henkilöä
- vuonna 1995 asuntokunnan koon keskihajonta oli $s_n = 1{,}156\ldots \approx 1{,}2$ henkilöä

c) Yhden henkilön asuntokuntien keskiarvo aikavälillä 1990–2015 saadaan laskemalla valmiin aineiston solujen B3–B28 keskiarvo ja keskihajonta. Tämä voidaan tehdä aineiston Excel-laskentataulukossa soluviittauksilla tai kopioimalla solut laskinsovellukseen ja käyttämällä tilastotoimintoja.

Yhden henkilön asuntokuntien

- keskiarvoksi on $\bar{x} = 901\,029$ kappaletta
- keskihajonta on $s_n = 140\,873{,}856\ldots \approx 140\,874$ kappaletta
- otoskeskihajonta on $s_{n-1} = 143\,663{,}709\ldots \approx 143\,664$ kappaletta

Koska kyseessä on otos, käytetään otoskeskihajontaa.

Yhden henkilön asuntokuntien määrän keskiarvo on 901 029 kappaletta ja määrän otoskeskihajonta on 143 664 kappaletta.

80.

a) Lasketaan luokkakeskukset luokkarajojen avulla: luokkakeskus on luokan ala- ja ylärajan keskiarvo.

Kirjaimien määrä sanassa	**Luokkakeskus**
1–5	$\frac{1+5}{2}=3$
6–10	$\frac{6+10}{2}=8$
11–15	$\frac{11+15}{2}=13$
16–20	$\frac{16+20}{2}=18$

b) Luokitellussa aineistossa tunnuslukujen laskemiseen käytetään luokkakeskuksia. Kirjoitetaan luokkakeskukset ja frekvenssit taulukkolaskentaohjelman tai laskinsovelluksen taulukkoon ja määritetään tunnusluvut tilastotoimintojen avulla:

	A	B
1	3	84
2	8	234
3	13	43
4	18	7

- keskiarvo on $\bar{x} = 7{,}633\ldots \approx 7{,}6$ kirjainta
- keskihajonta on $s_n = 3{,}224\ldots \approx 3{,}2$ kirjainta

c) Muuttujan arvo ei poikkea merkitsevästi keskiarvosta, jos se on korkeintaan kahden keskihajonnan päässä keskiarvosta. Lasketaan pisin mahdollinen sana, joka toteuttaa tämän ehdon, eli lasketaan keskiarvosta kaksi keskihajontaa ylöspäin oleva lukuarvo kirjainten määrälle. Käytetään laskussa tarkkoja arvoja: tehdään lasku taulukkolaskennan soluviittauksilla tai laskinsovelluksessa kopioimalla määritettyjen tunnuslukujen tarkat arvot laskukenttään.

$$\bar{x} + 2s_n = 7{,}633\ldots + 2 \cdot 3{,}224\ldots = 14{,}081\ldots \approx 14.$$

Pisin sana, joka ei poikkea merkitsevästi keskiarvosta on 14 kirjainta pitkä sana.

81.

a) Lasketaan luokkakeskukset luokkarajojen avulla: luokkakeskus on luokan ala- ja ylärajan keskiarvo.

Pituus (cm)	**Luokkakeskus**
10–14	$\frac{10+14}{2}=12$
15–19	$\frac{15+19}{2}=17$
20–24	$\frac{20+24}{2}=22$
25–29	$\frac{25+29}{2}=27$
30–34	$\frac{30+34}{2}=32$

b) Mitattujen heinien lukumäärä, eli havaintoyksiköiden lukumäärä, saadaan summaamalla frekvenssit:

$$10+6+18+22+6=62.$$

Ryhmä mittasi 62 heinän pituuden.

c) Luokitellussa aineistossa tunnuslukujen laskemiseen käytetään luokkakeskuksia. Kirjoitetaan luokkakeskukset ja frekvenssit taulukkolaskentaohjelman tai laskinsovelluksen taulukkoon ja määritetään tunnusluvut tilastotoimintojen avulla:

	A	B
1	12	10
2	17	6
3	22	18
4	27	22
5	32	6

- keskiarvo on $\bar{x} = 22{,}645\ldots \approx 22{,}65$ cm
- keskihajonta on $s_n = 6{,}056\ldots \approx 6{,}06$ cm

d) Puuttuvat mittaustulokset kasvattavat frekvenssejä seuraavasti:

- luokan 15–19 cm frekvenssi kasvaa kuudella, ja uusi frekvenssi on 6 + 6 = 12
- luokan 25–29 cm frekvenssi kasvaa yhdellä, ja uusi frekvenssi on 22 + 1 =23
- muiden luokkien frekvenssit pysyvät ennallaan

Päivitetään uudet frekvenssit laskentataulukkoon ja määritetään tunnusluvut uudestaan. Uuden aineiston tunnusluvut ovat:

- keskiarvo on $\bar{x} = 22{,}217\ldots \approx 22{,}22$ cm
- keskihajonta on $s_n = 5{,}985\ldots \approx 5{,}99$ cm

e) Muutoksen suunnan ja suuruuden laskemiseksi uutta arvo verrataan vanhaan. Käytetään laskussa tarkkoja arvoja eli tehdään lasku taulukkolaskennan soluviittauksilla tai laskinsovelluksessa kopioimalla määritettyjen tunnuslukujen tarkat arvot laskukenttään.

Lasketaan uuden keskiarvon suhde vanhaan ja ilmaistaan osamäärä prosentteina:

$$\frac{22{,}218\ldots}{22{,}645\ldots} \cdot 100\ \% = 98{,}110\ldots\ \% \approx 98{,}1\ \%.$$

Keskiarvo pieneni 100,0 % – 98,1 % = 1,9 %.

Lasketaan uuden keskihajonnan suhde vanhaan ja ilmaistaan osamäärä prosentteina:

$$\frac{5{,}985\ldots}{6{,}056\ldots} \cdot 100\ \% = 98{,}823\ldots\ \% \approx 98{,}8\ \%.$$

Keskihajonta pieneni 100,0 % – 98,8 % = 1,2 %.

82.

Kirjoitetaan elokuvateattereiden istumapaikkojen lukumäärät taulukkolaskentaohjelman tai laskinsovelluksen taulukkoon tai kopioidaan annettu valmis aineisto työpöydän kautta.

a) Vaihteluvälin pituus lasketaan vähentämällä suurimmasta muuttujan arvosta pienin muuttujan arvo. Vaihteluvälin pituuden laskemiseksi pitää määrittää istumapaikkojen lukumäärän pienin ja suurin arvo. Useimmat tilastotoiminnot kertovat nämä samassa yhteydessä muiden tilastollisten tunnuslukujen kanssa.

- istumapaikkojen suurin arvo on 703
- istumapaikkojen pienin arvo on 57

Istumapaikkojen lukumäärän vaihteluvälin pituus on $703 - 57 = 646$.

b) Määritetään tunnusluvut tilastotoimintojen avulla:

- mediaani on Md = 121 paikkaa
- keskiarvo on $\bar{x} = 162{,}727\ldots \approx 163$ paikkaa
- keskihajonta on $s_n = 126{,}468\ldots \approx 126$ paikkaa

Istumapaikkojen lukumäärän mediaani on siis Md = 121 paikkaa, keskiarvo on $\bar{x} \approx 163$ paikkaa ja keskihajonta on $s_n \approx 126$ paikkaa.

83.

a) Kirjoitetaan kunnan vuosittaiset työntekijöiden lukumäärät taulukkolaskentaohjelman tai laskinsovelluksen taulukkoon ja määritetään tunnusluvut tilastotoimintojen avulla:

- keskiarvo on $\bar{x} = 567$ (työntekijää)
- keskihajonta on $s_n = 23{,}433\ldots \approx 23{,}4$ (työntekijää)

b) Täydennetään a-kohdan taulukkoa vuoden 2015 työntekijämäärällä 504 ja lasketaan keskiarvo aikavälille 2006–2015.

Uusi keskiarvo on $\bar{x} = 560{,}7 \approx 561$ työntekijää.

c) Merkitään kunnan työntekijöiden lukumäärää vuonna 2016 kirjaimella *x.*

Kunnan työntekijöiden lukumäärän summa vuosina 2006–2016 on tällöin

$$546 + 581 + 602 + 594 + 573 + 562 + 567 + 558 + 520 + 504 + x = x + 5607.$$

Työntekijöiden keskiarvo aikavälillä 2006–2016 on $\frac{x+5607}{11}$. Tiedetään, että keskiarvo on 556, joten saadaan yhtälö:

$$\frac{x + 5607}{11} = 556.$$

Ratkaistaan yhtälöstä tuntematon lukumäärä *x:*

$$\frac{x + 5607}{11} = 556 \quad | \cdot 11$$

$$x + 5607 = 6116$$

$$x = 6116 - 5607 = 509.$$

Kunnan työntekijöin määrä vuonna 2016 oli 509 työntekijää.

84.

Tunnusluvut voidaan määrittää joko tehtävänannossa annetun aineiston Excel-taulukossa tai kopioida aineisto toisen taulukkolaskentaohjelman tai laskinsovelluksen taulukkoon ja määrittää tunnusluvut tilastotoiminnoilla.

Excel-ohjelmassa tunnusluvut määritetään sopivilla laskukaavakomennoilla ja soluviittauksilla. Komennot vaihtelevat Excelin kielen ja ohjelmistoversion mukaan: jos käytät Exceliä, selvitä käyttämäsi version toiminnot.

a) Määritetään lämpötilojen keskiarvot sopivalla laskukaavakomennolla ja soluviittauksilla, tai laskinsovelluksen tilastotoiminnoilla:

- Helsingissä talven keskilämpötilojen keskiarvo on
  $\bar{x} = -3{,}953\ldots \approx -4{,}0\ °\text{C}$

- Jyväskylässä talven keskilämpötilojen keskiarvo on
  $\bar{x} = -8{,}020\ldots \approx -8{,}0\ °\text{C}$

- Sodankylässä talven keskilämpötilojen keskiarvo on
  $\bar{x} = -12{,}844\ldots \approx -12{,}8\ °\text{C}$

b) Määritetään lämpötilojen keskihajonnat sopivalla laskukaavakomennolla ja soluviittauksilla, tai laskinsovelluksen tilastotoiminnoilla:

- Helsingissä talven keskilämpötilojen keskihajonta on $s_n = 2{,}684 \ldots \approx 2{,}7$ °C
- Jyväskylässä talven keskilämpötilojen keskihajonta on $s_n = 3{,}079 \ldots \approx 3{,}1$ °C
- Sodankylässä talven keskilämpötilojen keskihajonta on $s_n = 3{,}019 \ldots \approx 3{,}0$ °C

Helsingissä talven keskilämpötilojen keskihajonta on pienin, joten Helsingin keskilämpötilat poikkeavat vähiten keskiarvosta.

c) Vaihteluvälin pituus lasketaan vähentämällä muuttujan suurimmasta arvosta muuttujan pienin arvo. Määritetään lämpötilojen pienin ja suurin arvo sopivilla laskukaavakomennoilla ja soluviittauksilla, tai laskinsovelluksen tilastotoiminnoilla.

- Helsingissä talven keskilämpötilojen pienin arvo on $-9{,}5$ °C ja suurin arvo on $1{,}4$ °C. Keskilämpötilan vaihteluvälin pituus on $1{,}4 - (-9{,}5) = 1{,}4 + 9{,}5 = 10{,}9$ °C.

- Jyväskylässä talven keskilämpötilojen pienin arvo on $-14{,}8$ °C ja suurin arvo on $-2{,}0$ °C. Keskilämpötilan vaihteluvälin pituus on $-2{,}0 - (-14{,}8) = -2{,}0 + 14{,}8 = 12{,}8$ °C.

- Sodankylässä talven keskilämpötilojen pienin arvo on $-20{,}0$ °C ja suurin arvo on $-7{,}4$ °C. Keskilämpötilan vaihteluvälin pituus on $-7{,}4 - (-20{,}0) = -7{,}4 + 20{,}0 = 12{,}6$ °C.

d) Lasketaan kullekin kaupungille keskihajonnan suhde vaihteluvälin pituuteen. Käytetään laskussa keskihajonnan tarkkoja arvoja eli tehdään laskut taulukkolaskennan soluviittauksilla tai laskinsovelluksessa kopioimalla määritettyjen keskihajontojen tarkat arvot laskukenttään.

- Helsingissä suhde on $\frac{2{,}684\ldots}{10{,}9} = 0{,}246\ldots$

- Jyväskylässä suhde on $\frac{3{,}079\ldots}{12{,}8} = 0{,}240\ldots$

- Sodankylässä suhde on $\frac{3{,}019\ldots}{12{,}6} = 0{,}239\ldots$

Suhde on suurin Helsingissä, joten Helsingin talven keskilämpötilojen keskihajonta on suurin suhteessa vaihteluvälin pituuteen. Huomaa ero a-kohtaan.

85.

a) Syötetään arvosanat ja kummankin koulun arvosanafrekvenssit taulukkolaskentaohjelman tai laskinsovelluksen taulukkoon. Määritetään kummankin koulun arvosanojen keskiarvo ja keskihajonta tilastotoimintojen avulla:

Koulu A:

- keskiarvo on $\bar{x}_A = 7{,}358\ldots \approx 7{,}36$
- keskihajonta on $s_n = 1{,}136\ldots \approx 1{,}14$

Koulu B:

- keskiarvo on $\bar{x}_B = 7{,}928\ldots \approx 7{,}93$
- keskihajonta on $s_n = 1{,}640\ldots \approx 1{,}64$

b) Koulussa B matematiikan arvosanojen keskiarvo on suurempi, joten koulu B pärjäsi matematiikan kokeessa keskimääräisesti paremmin.

c) Opiskelijoiden lukumäärä on sama kuin aineiston havaintoyksiköiden lukumäärä, joka saadaan laskemalla frekvenssit yhteen. Koulussa A frekvenssien summa on

$$0 + 12 + 27 + 29 + 4 + 9 = 81.$$

Useimmat tilastotoiminnot antavat havaintoyksiköiden lukumäärän samassa yhteydessä tilastollisten tunnuslukujen kanssa. Havaintoyksiköiden lukumäärä on usein merkitty kirjaimella n eli koulun A tuloksissa $n = 81$.

Tällä luokkatasolla oli koulussa A yhteensä 81 opiskelijaa.

d) Koulussa A ylimmän arvosanojan 10 frekvenssi on nolla, joten kukaan opiskelija ei saanut ylintä arvosanaa. Koulun A paras saatu arvosana oli 9.

Koulussa B alimman arvosanan 4 frekvenssi on nolla, joten kukaan opiskelija ei saanut alinta arvosanaa. Koulun B huonoin saatu arvosana oli 5.

e) Muuttujan arvo poikkeaa merkitsevästi keskiarvosta, jos se on yli kahden keskihajonnan päässä keskiarvosta. Tällöin normitetun arvon itseisarvo on suurempi kuin kaksi.

Koulussa A huonoin arvosana on 5. Lasketaan arvosanaa $x = 5$ vastaava normitettu arvo. Käytetään laskussa tarkkoja arvoja eli tehdään lasku taulukkolaskennan soluviittauksilla tai laskinsovelluksessa kopioimalla määritettyjen tunnuslukujen tarkat arvot laskukenttään.

Arvosanan 5 normitettu arvo on

$$z_A = \frac{x - \bar{x}_A}{s_n} = \frac{5 - 7{,}358\ldots}{1{,}136\ldots} = -2{,}074\ldots$$

Normitetun arvon itseisarvo on $|z_A| = 2{,}074\ldots > 2$, joten koulun A huonoin arvosana poikkeaa merkitsevästi keskiarvosta.

Koulussa B huonoin arvosana on 4. Lasketaan arvosanaa $x = 4$ vastaava normitettu arvo. Käytetään laskussa tarkkoja arvoja: tehdään lasku taulukkolaskennan soluviittauksilla tai laskinsovelluksessa kopioimalla määritettyjen tunnuslukujen tarkat arvot laskukenttään.

Arvosanan 4 normitettu arvo on

$$z_B = \frac{x - \bar{x}_B}{s_n} = \frac{4 - 7{,}928\ldots}{1{,}640\ldots} = -2{,}394\ldots$$

Normitetun arvon itseisarvo on $|z_B| = 2{,}394\ldots > 2$, joten myös koulun B huonoin arvosana poikkeaa merkitsevästi keskiarvosta.

86.

a) Lasketaan luokkakeskukset luokkarajojen avulla: luokkakeskus on luokan ala- ja ylärajan keskiarvo, eli välin keskikohta.

Pisteet	**Luokkakeskus**
50–54	$\frac{50+54}{2}=52$
45–49	47
40–44	42
35–39	37
30–34	32
25–29	27

	A	B
1	52	2
2	47	1
3	42	1
4	37	12
5	32	14
6	27	20

b) Luokitellussa aineistossa tunnuslukujen laskemiseen käytetään luokkakeskuksia. Kirjoitetaan luokkakeskukset ja frekvenssit taulukkolaskentaohjelman tai laskinsovelluksen taulukkoon ja määritetään tunnusluvut tilastotoimintojen avulla:

- keskiarvo on $\bar{x} = 32{,}5$ pistettä
- keskihajonta on $s_n = 6{,}184\ldots \approx 6{,}2$ pistettä

c) Kahdessa eri jakaumassa (SM-liiga ja NHL-liiga) normitetut arvot vastaavat toisiaan.

Normitetaan McDavidin NHL-liigassa saama pistemäärä $x = 63$. Normitettu arvo on

$$\frac{x - \bar{x}}{s_n} = \frac{63 - 48{,}4}{6{,}1}.$$

Merkitään kysyttyä SM-liigan pistesaalista kirjaimella *x.* Normitettu arvo on

$$\frac{x - \bar{x}}{s_n} = \frac{x - 32{,}5}{6{,}184\ \ldots}.$$

Merkitään normitetut arvot yhtä suuriksi:

$$\frac{63 - 48{,}4}{6{,}1} = \frac{x - 32{,}5}{6{,}184\ \ldots}$$

Ratkaistaan yhtälöstä tuntematon *x.* Tehdään se tarkoilla arvoilla kopioimalla taulukkolaskentaohjelmassa tai laskinsovelluksessa määritetyt tunnuslukujen tarkat arvot yhtälönratkaisun laskukenttään.

Yhtälön ratkaisu on $x = 47{,}302\ \ldots \approx 47$. McDavidin NHL-liigan pistemäärä 63 pistettä vastaa siis SM-liigassa pistemäärää 47.

d) Merkitään kysyttyä NHL-liigan pistemäärää kirjaimella *x.* Merkitään normitetut arvot yhtä suuriksi:

$$\frac{52 - 32{,}5}{6{,}184 \ldots} = \frac{x - 48{,}4}{6{,}1}$$

Yhtälön ratkaisu on $x = 67{,}633 \ldots \approx 68$. Haapalan SM-liigan pistemäärä 52 pistettä vastaa siis NHL-liigassa pistemäärää 68.

1 Tilastot

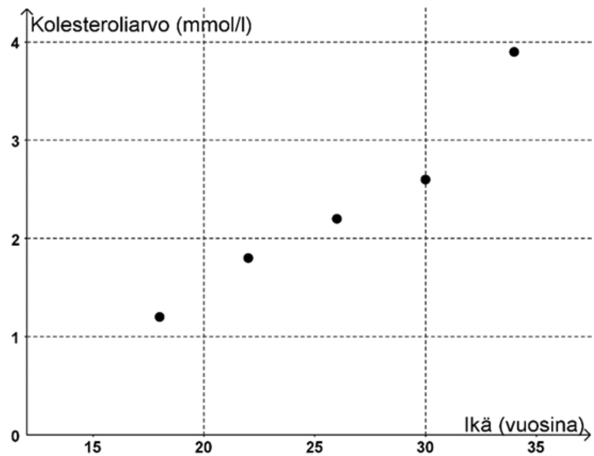
1.5 Kahden muuttujan yhteisjakauma

87.

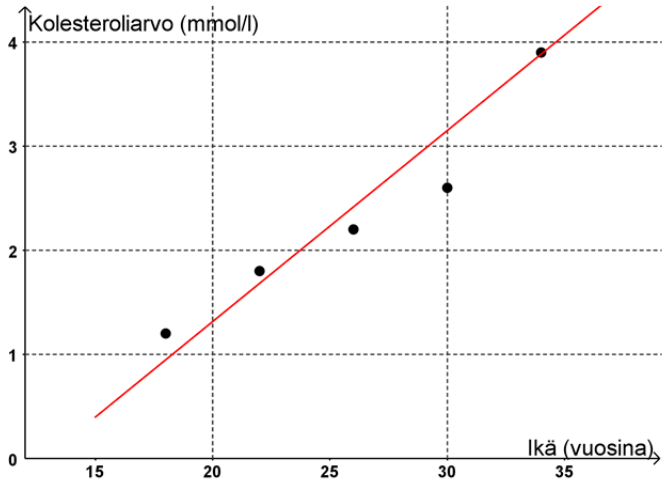
a) Kolesteroliarvo usein nousee henkilön ikääntyessä. Selittävä muuttuja *x* on siis ikä ja selitettävä muuttuja *y* on kolesteroliarvo.

b) Tutkitaan riippuvuutta sopivan taulukkolaskentaohjelman avulla. Kirjoitetaan potilaan iät ja niitä vastaavat kolesteroliarvot taulukkoon ja piirretään taulukon avulla sirontakuvaaja: selittävän muuttujan (iän) arvot ovat *x*-akselilla ja selitettävän muuttujan (kolesteroliarvon) arvot ovat *y*-akselilla.

	A	B
1	**Ikä**	**Kolesteroliarvo**
2	18	1.2
3	22	1.8
4	26	2.2
5	30	2.6
6	34	3.9
7	38	5



c) Sovitetaan pistejoukkoon suora.



Riippuvuutta kuvaavan suoran yhtälö on

$$y = 0{,}184 \ldots x - 2{,}357 \ldots \approx 0{,}18x - 2{,}4.$$

Huomaa, että kerroin ja vakio pyöristetään kahden merkitsevän numeron tarkkuuteen.

d) Korrelaatiokertoimen arvo on

$$r = 0{,}970 \ldots \approx 0{,}97.$$

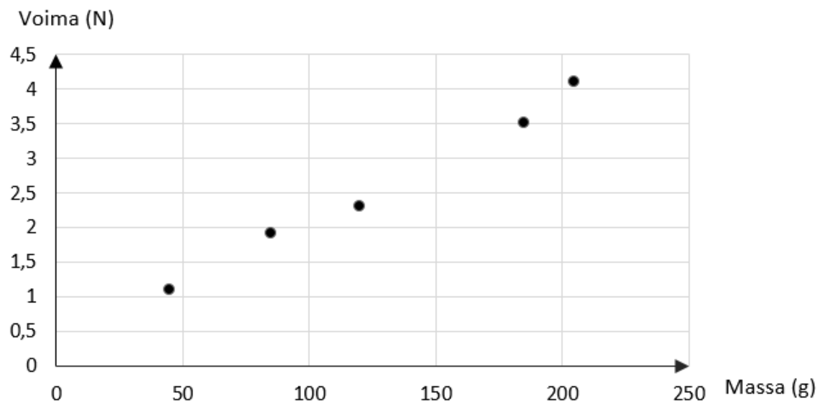
Koska korrelaatiokertoimen itseisarvo $|r| = 0{,}97 > 0{,}8$, niin lineaarinen riippuvuus on voimakas.

88.

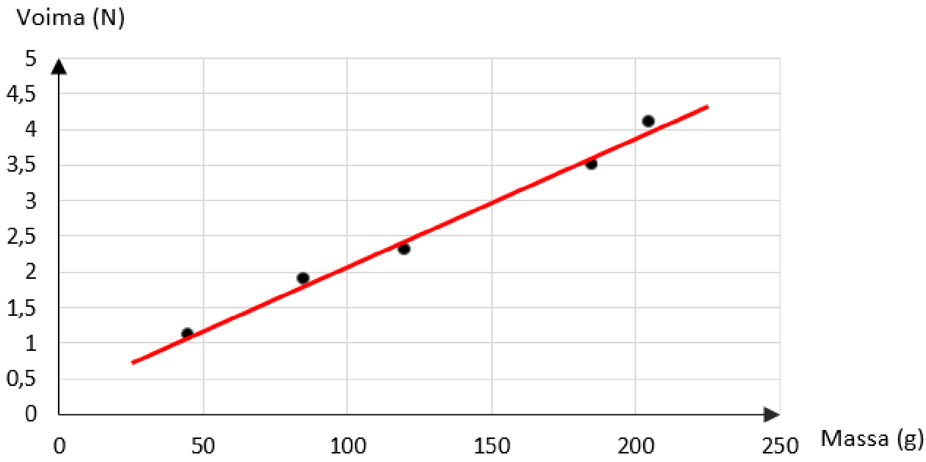
a) Punnuksen vetämiseen tarvittava voima riippuu punnuksen massasta. Selittävä muuttuja on siis massa ja selitettävä muuttuja on voima.

b) Tutkitaan riippuvuutta sopivan taulukkolaskentaohjelman avulla. Kirjoitetaan punnusten massat ja niitä vastaavat voimat taulukkoon. Piirretään taulukon avulla sirontakuvaaja: selittävän muuttujan eli massan arvot ovat *x*-akselilla ja selitettävän muuttujan eli voiman arvot ovat *y*-akselilla.

	A	B
1	**Massa (g)**	**Voima (N)**
2	45	1,1
3	85	1,9
4	120	2,3
5	185	3,5
6	205	4,1



Sovitetaan pistejoukkoon suora.



Riippuvuutta kuvaavan suoran yhtälö on

$$y = 0{,}0180\ldots x + 0{,}274\ldots \approx 0{,}018x + 0{,}27.$$

Huomaa, että kerroin ja vakio pyöristetään kahden merkitsevän numeron tarkkuuteen.

c) Korrelaatiokertoimen arvo kolmen merkitsevän numeron tarkkuudella on

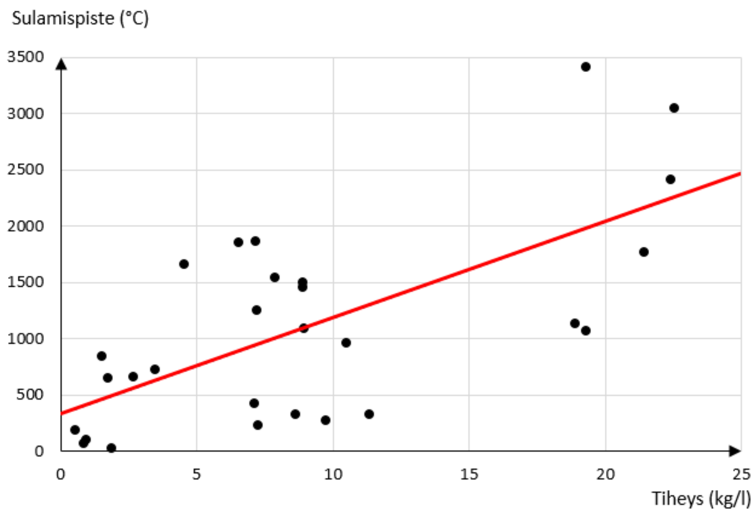
$$r = 0{,}9951\ldots \approx 0{,}995.$$

Koska korrelaatiokertoimen itseisarvo $|r| = 0{,}995$ on lähes 1, lineaarinen riippuvuus on voimakas.

89.

a) Tutkitaan tiheyden vaikutusta sulamispisteeseen. Piirretään tehtävänannossa annetussa tiedostossa sarakkeissa D ja E olevien lukuparien sirontakuvaaja taulukkolaskentaohjelman avulla: selittävän muuttujan eli tiheyden arvot tulevat x-akselille ja selitettävän muuttujan eli sulamispisteen arvot tulevat y-akselilla.

Sovitetaan pistejoukkoon (x, y) regressiosuora.



Muuttujien välistä riippuvuutta kuvaavan regressiosuoran yhtälö on

$$y = 85{,}425\ldots x + 329{,}533\ldots \approx 85x + 330.$$

Huomaa, että kerroin ja vakio pyöristetään kahden merkitsevän numeron tarkkuuteen.

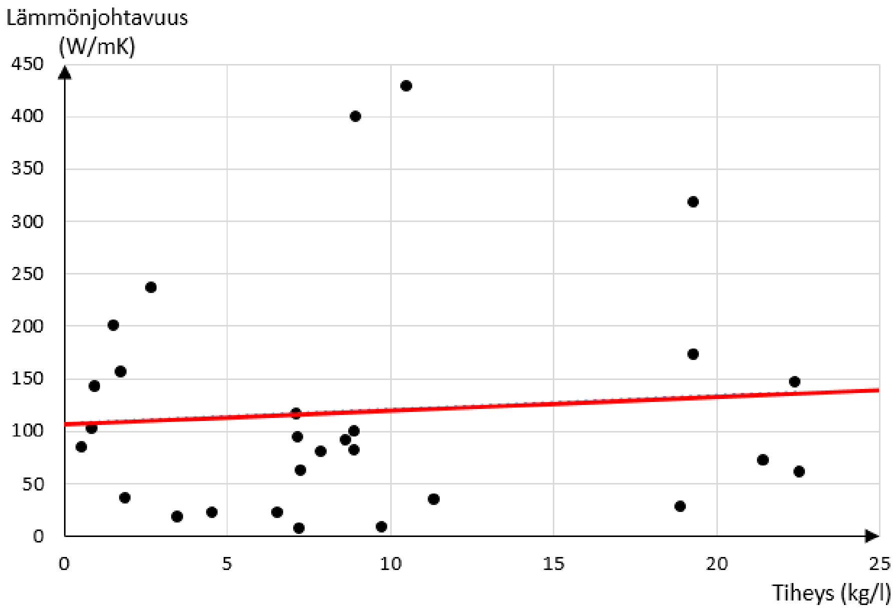
Korrelaatiokertoimen arvo on

$$r = 0{,}673\ldots \approx 0{,}67.$$

Korrelaatiokertoimen itseisarvo $|r| = 0{,}67$ on välillä 0,6–0,8, joten muuttujien välinen lineaarinen riippuvuus on huomattava.

b) Tutkitaan tiheyden vaikutusta lämmönjohtavuuteen. Piirretään annetussa tiedostossa sarakkeissa D ja F olevien lukuparien sirontakuvio taulukkolaskentaohjelman avulla: selittävän muuttujan eli tiheyden arvot tulevat x-akselille ja selitettävän muuttujan eli lämmönjohtavuuden arvot tulevat y-akselilla.

Sovitetaan pistejoukkoon (x, y) regressiosuora.



Muuttujien välistä riippuvuutta kuvaavan regressiosuoran yhtälö on

$$y = 1{,}309\ldots x + 107{,}026\ldots \approx 1{,}3x + 110.$$

Huomaa, että kerroin ja vakio pyöristetään kahden merkitsevän numeron tarkkuuteen.

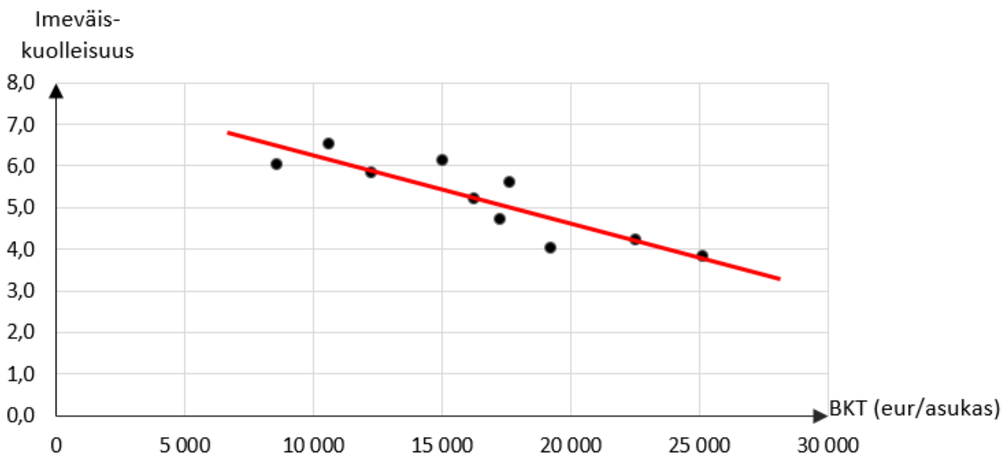
Korrelaatiokertoimen arvo on

$$r = 0{,}0822\ldots \approx 0{,}082.$$

Korrelaatiokertoimen itseisarvo on $|r| = 0{,}082 < 0{,}3$, joten muuttujien välinen lineaarinen riippuvuus on merkityksetön. Tämä nähdään jo sirontakuvaajasta: pistejoukon pisteet eivät edes likimain sijoitu suoralle.

90.

Kirjoitetaan bruttokansantuotteet x ja imeväiskuolleisuudet y taulukkolaskentaohjelman taulukkoon. Piirretään sirontakuvaaja ja sovitetaan pistejoukkoon regressiosuora.



a) Regressiosuoran yhtälö on

$$y = -0{,}000165\ldots x + 7{,}910\ldots \approx -0{,}00017x + 7{,}9.$$

Huomaa, että kerroin ja vakio pyöristetään kahden merkitsevän numeron tarkkuuteen.

b) Imeväiskuolleisuus vuonna 2002 lasketaan sijoittamalla $x = 26\,863$ regressiosuoran yhtälöön. Käytetään laskussa mallia, jossa regressiosuoran kulmakerroin ja vakiotermi on pyöristetty kahden merkitsevän numeron tarkkuuteen:

$$y = -0{,}00017 \cdot 26\,863 + 7{,}9 \approx 3{,}3.$$

Mallin mukaan imeväiskuolleisuus vuonna 2002 oli 3,3 alle 1-vuotiasta lasta jokaista tuhatta syntynyttä lasta kohden.

Huomaa, että laskussa käytetään regressiosuoran kertoimelle ja vakiolle pyöristettyjä arvoja. Jos lasku tehdään tarkoilla arvoilla esimerkiksi taulukkolaskentaohjelman laskukentässä, niin imeväiskuolleisuudeksi saadaan

$$y = -0{,}000165\ldots \cdot 26\,863 + 7{,}910\ldots = 3{,}477\ldots \approx 3{,}5$$

eli hivenen suurempi luku kuin pyöristetyillä kertoimilla laskettaessa.

c) Vuonna 2015 imeväiskuolleisuus oli $y = 1{,}7$. Muodostetaan regressiosuoran avulla yhtälö ja ratkaistaan yhtälöstä bruttokansantuote *x:*

$$1{,}7 = -0{,}00017x + 7{,}9$$

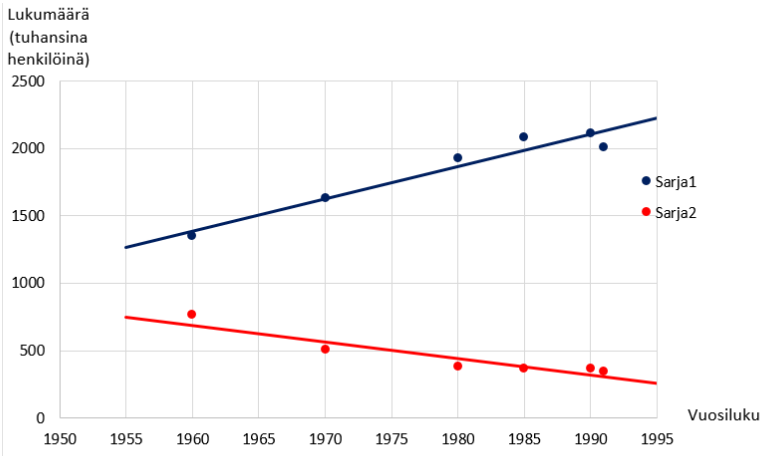
Yhtälö voidaan ratkaista symbolisen laskennan ohjelmalla:
$x = 36\,470{,}5\ldots$

Pyöristetään vastaus kahden merkitsevä numeron tarkkuuteen: mallin mukaan bruttokansantuote vuonna 2015 oli 36 000 eur/asukas.

91.

Kirjoitetaan vuosiluvut x ja niitä vastaavat lukumäärät y taulukkolaskentaohjelman taulukkoon. Piirretään sirontakuvaajat ja sovitetaan pistejoukkoihin regressiosuorat.

	A	
1	**Vuosi**	**Pal**
2	1960	
3	1970	
4	1980	
5	1985	
6	1990	
7	1991	



Palkansaajien (Sarja1) regressiosuoran yhtälö on

$$y \approx 23{,}94x - 45\,530.$$

Yrittäjien (Sarja2) regressiosuoran yhtälö on

$$y \approx -12{,}41x + 25\,013.$$

Sijoitetaan $x = 1994$ regressiosuoran yhtälöön: palkansaajien mallissa

$$y \approx 23{,}94 \cdot 1994 - 45\,530 \approx 2200.$$

Muuttujan y arvo on ilmaistu tuhansina henkilöinä, joten palkansaajia oli

$$2200 \cdot 1000 = 2\,200\,000$$

eli palkansaajia oli vuonna 1994 noin 2,2 miljoonaa.

Yrittäjien mallissa

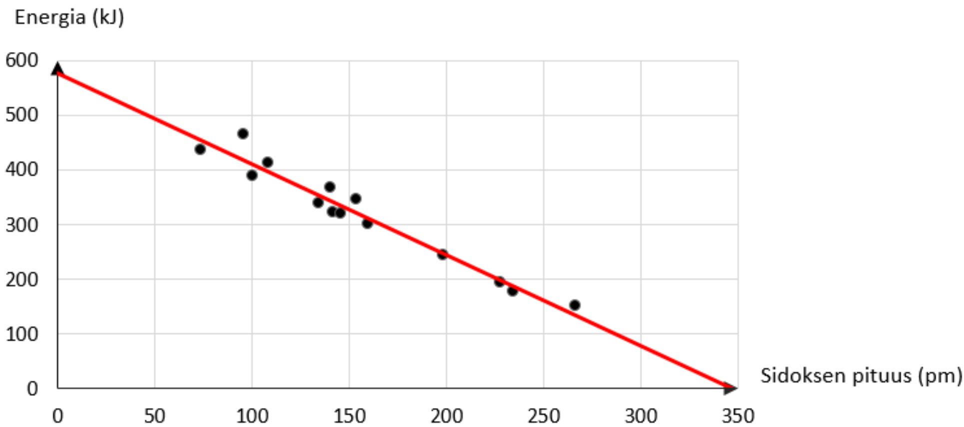
$$y \approx -12{,}41 \cdot 1994 + 25\,013 \approx 270.$$

Yrittäjiä oli vuonna 1994 noin 270 000.

92.

a) Kirjoitetaan sidoksen pituudet *x* ja niitä vastaavat energiamäärät *y* taulukkolaskentaohjelman taulukkoon. Piirretään sirontakuvaaja ja sovitetaan pistejoukkoon regressiosuora.

	A	B
1	**Sidoksen pituus (pm)**	**Energia (kJ)**
2	74	436
3	96	463
4	101	388
5	109	412
6	135	338
7	141	366
8	142	322
9	146	318
10	154	346
11	160	299
12	199	242
13	228	193
14	235	176
15	267	151



Regressiosuoran yhtälö on

$$y = -1{,}662x + 577{,}36 \ldots \approx -1{,}7x + 580.$$

Huomaa, että kerroin ja vakio pyöristetään kahden merkitsevän numeron tarkkuuteen.

b) Kysytty energiamäärä lasketaan sijoittamalla $x = 128$ regressiosuoran yhtälöön:

$$y = -1{,}7 \cdot 128 + 580 = 362{,}4 \approx 360.$$

Huomaa, että laskussa käytetään regressiosuoran kertoimelle ja vakiolle pyöristettyjä arvoja. Mallin mukaan kloorin ja vedyn väliseen sidokseen sitoutuu energiaa 360 kJ.

c) Kahden typen välisessä sidoksessa energiaa on $y = 163$ (kJ). Muodostetaan regressiosuoran avulla yhtälö ja ratkaistaan yhtälöstä sidoksen pituus *x:*

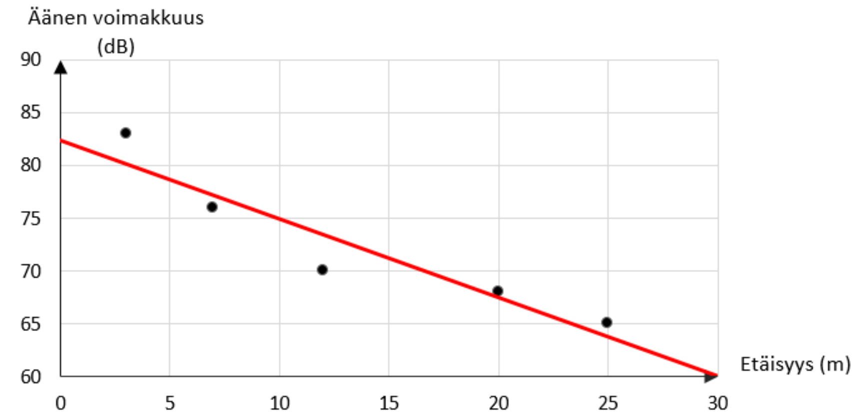
$$163 = -1{,}7x + 580$$

Yhtälö voidaan ratkaista symbolisen laskennan ohjelmalla:
$x = 245{,}2 \ldots \approx 250.$

Mallin mukaan kahden typen välisen N–N-sidoksen pituus on 250 pm.

93.

a) Kirjoitetaan etäisyydet x ja niitä vastaavat voimakkuudet y taulukkolaskentaohjelman taulukkoon. Piirretään pistejoukon (x, y) kuvaaja ja sovitetaan siihen regressiosuora.



	A	B
1	**Etäisyys (m)**	**Voimakkuus (dB)**
2	25,0	65
3	20,0	68
4	12,0	70
5	7,0	76
6	3,0	83

Regressiosuoran yhtälö on

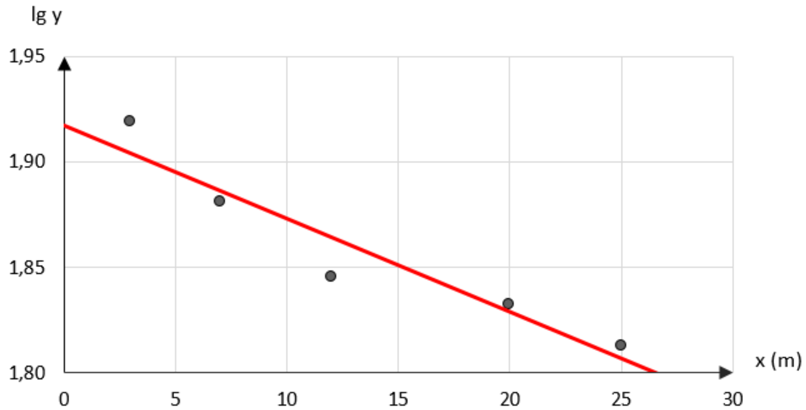
$$y = -0{,}743\ldots x + 82{,}364\ldots \approx -0{,}74x + 82.$$

Muuttujien arvot muuttuvat eri suuntiin: muuttujan x arvon kasvaessa muuttujan y arvo pienenee, eli korrelaatiosuora on laskeva. Korrelaatiokerroin on siis negatiivinen luku.

Korrelaatiokerroin on $r = -0{,}941\ldots \approx -0{,}94$.

b) Lasketaan taulukkoon 10-kantaiset logaritmit äänenvoimakkuuksista. Piirretään pistejoukon $(x, \lg y)$ kuvaaja ja sovitetaan siihen regressiosuora.

	A	B	C
1	**Etäisyys (m)**	**Voimakkuus (dB)**	**lg y**
2	25,0	65	1,81291
3	20,0	68	1,83251
4	12,0	70	1,84510
5	7,0	76	1,88081
6	3,0	83	1,91908



Regressiosuoran yhtälö on

$$y_{\text{muunnos}} = -0{,}00441\ldots x + 1{,}917\ldots \approx -0{,}0044x + 1{,}9.$$

Muuttujien arvot muuttuvat eri suuntiin: muuttujan x arvon kasvaessa muuttujan y arvo pienenee, eli regressiosuora on laskeva. Korrelaatiokerroin on siis negatiivinen luku.

Korrelaatiokerroin on $r = -0{,}951\ldots \approx -0{,}95$.

Korrelaatiokerroin kuvaa sitä, miten hyvin sirontakuvaajan pisteet toteuttavat regressiosuoran yhtälön: mitä lähempänä korrelaatiokertoimen itseisarvo $|r|$ on lukua 1, sitä paremmin malli kuvaa tilannetta.

a-kohdan mallille $y = -0{,}74x + 82$ on $|r| = |-0{,}94| = 0{,}94$.

b-kohdan mallille
$y_{\text{muunnos}} = -0{,}0044x + 1{,}9$ on $|r| = |-0{,}95| = 0{,}95 > 0{,}94$.

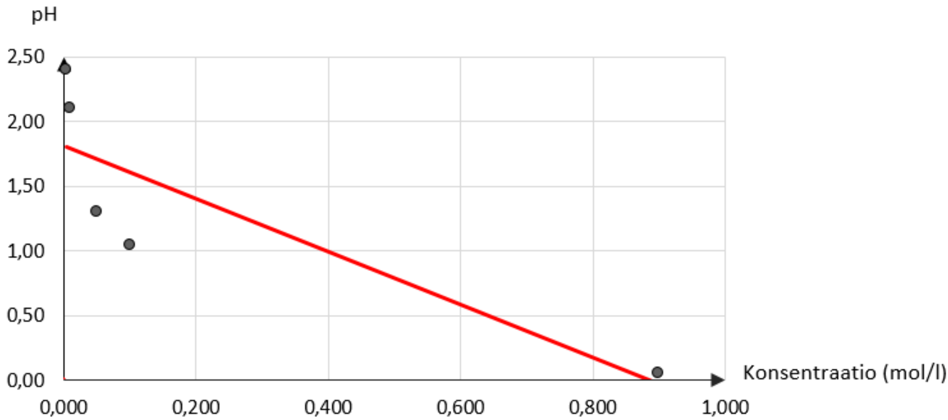
b-kohdan malli, eli suora $y = -0{,}0044x + 1{,}9$ kuvaa pistejoukkoa hieman paremmin, koska muunnoksella saadun mallin korrelaatiokertoimen itseisarvo on hieman suurempi kuin a-kohdan mallin.

b-kohdan suora $= -0{,}0044x + 1{,}9$ on hieman parempi malli.

94.

a) Kirjoitetaan konsentraatiot x ja niitä vastaavat pH-arvot y taulukkolaskentaohjelman taulukkoon. Piirretään pistejoukon (x, y) kuvaaja ja sovitetaan siihen regressiosuora.

	A	B
1	**Konsentraatio (mol/l)**	**pH**
2	0,900	0,05
3	0,100	1,04
4	0,050	1,30
5	0,010	2,10
6	0,005	2,40



Regressiosuoran yhtälö riippuu käytettävästä ohjelmasta:

$$y \approx -2{,}0x + 1{,}8 \quad \text{tai} \quad y \approx -2{,}1x + 1{,}8.$$

Muuttujien arvot muuttuvat eri suuntiin: muuttujan x arvon kasvaessa muuttujan y arvo pienenee, eli regressiosuora on laskeva. Korrelaatiokerroin on siis negatiivinen luku.

Korrelaatiokerroin on $r = -0{,}851\ldots \approx -0{,}85$.

b) Lasketaan taulukkoon 10-kantaiset logaritmit äänenvoimakkuuksista. Piirretään pistejoukon $(x, \lg y)$ kuvaaja ja sovitetaan siihen regressiosuora.

	A	B	C
1	**Konsentraatio (mol/l)**	**pH**	**lg y**
2	0,900	0,05	-1,301
3	0,100	1,04	0,017
4	0,050	1,30	0,114
5	0,010	2,10	0,322
6	0,005	2,40	0,380

Regressiosuoran yhtälö on

$$y_{\text{muunnos}} \approx -1{,}8x + 0{,}29.$$

Muuttujien arvot muuttuvat eri suuntiin: muuttujan x arvon kasvaessa muuttujan y arvo pienenee. Korrelaatiokerroin on siis negatiivinen luku.

Korrelaatiokerroin on $r = -0{,}992\ldots \approx -0{,}99$.

Korrelaatiokerroin kuvaa sitä, miten hyvin sirontakuvaajan pisteet toteuttavat regressiosuoran yhtälön: mitä lähempänä korrelaatiokertoimen itseisarvo on $|r|$ on luku 1, sitä paremmin malli kuvaa tilannetta.

a-kohdan mallin korrelaatiokertoimelle on $|r| = |-0{,}85| = 0{,}85$

b-kohdan mallin korrelaatiokertoimelle pätee
$|r| = |-0{,}99| = 0{,}99 > 0{,}85$

b-kohdan malli, eli suora $y = -1{,}8x + 0{,}29$ kuvaa pistejoukkoa paremmin, koska muunnoksella saadun pistejoukon korrelaatiokertoimen itseisarvo on suurempi kuin a-kohdan mallin.

b-kohdan suora $= -1{,}8x + 0{,}29$ on parempi malli.

95.

Kirjoitetaan muuttujan x arvot ja niitä vastaavat muuttujan y arvot taulukkolaskentaohjelman taulukkoon.

	A	B
1	**Muuttuja x**	**Muuttuja y**
2	1	1,6
3	2	1,7
4	3	3,0
5	4	5,8
6	5	8,5
7	6	11,1
8	7	15,1

a) Piirretään pistejoukon (x, y) kuvaaja ja sovitetaan siihen eksponentiaalinen regressiomalli.

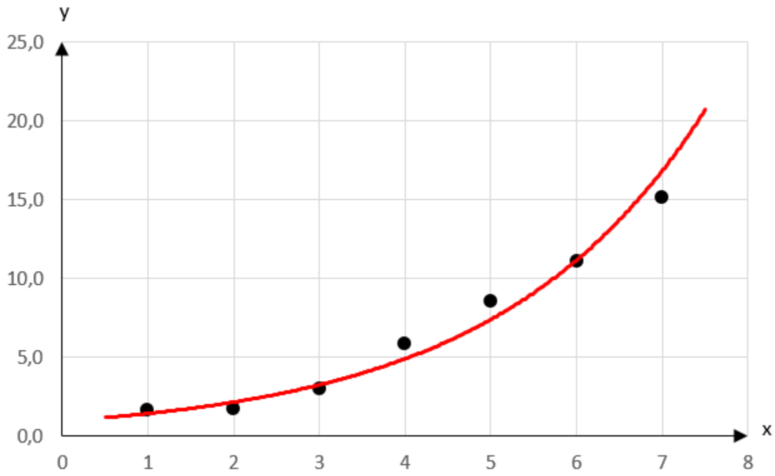
Regressiokäyrän yhtälö riippuu käytettävästä ohjelmasta:

$$y = 0{,}943\ldots e^{0{,}411\ldots x} \approx 0{,}94e^{0{,}41x} \quad \text{(kantalukuna Neperin luku e)}$$

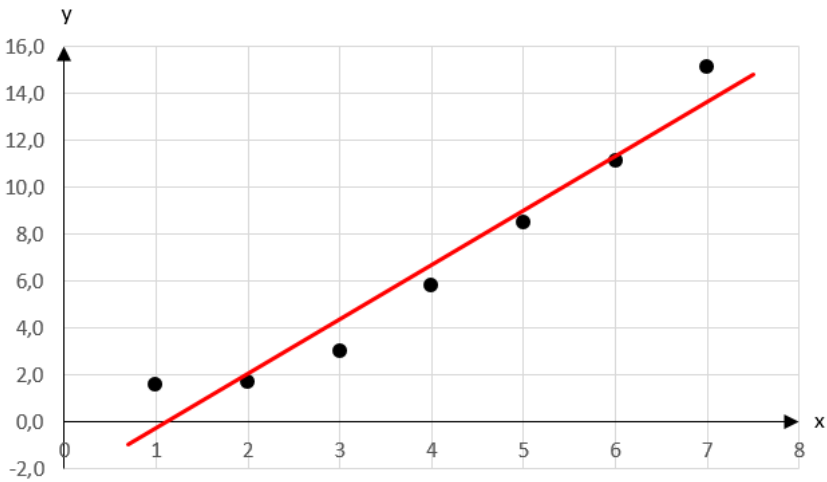
tai

$$y = 0{,}943\ldots \cdot (1{,}509\ldots)^x \approx 0{,}94 \cdot 1{,}5^x \quad \text{(kantalukuna luku 1,5)}$$

Eksponentiaalisen mallin selitysaste on $r^2 = 0{,}972\ldots \approx 0{,}97$.



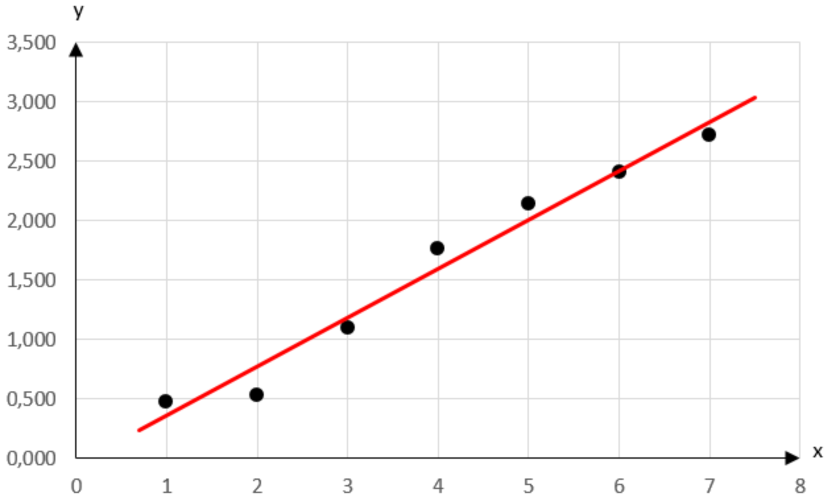
b) Piirretään pistejoukon (x, y) kuvaaja ja sovitetaan siihen lineaarinen malli.



Regressiosuoran yhtälö on $y = 2{,}314 \ldots x - 2{,}571 \ldots \approx 2{,}3x - 2{,}6$.

Lineaarisen mallin selitysaste on $r^2 = 0{,}945 \ldots \approx 0{,}95$.

c) Lasketaan taulukkoon luonnolliset logaritmit muuttujan y arvoista. Piirretään pistejoukon $(x, \ln y)$ kuvaaja ja sovitetaan siihen lineaarinen malli.



	A	B	C
1	**Muuttuja x**	**Muuttuja y**	**ln y**
2	1	1,6	0,470
3	2	1,7	0,531
4	3	3,0	1,099
5	4	5,8	1,758
6	5	8,5	2,140
7	6	11,1	2,407
8	7	15,1	2,715

Regressiosuoran yhtälö on $y = 0{,}411\ldots x - 0{,}0584\ldots \approx 0{,}41x - 0{,}058$.

Mallin selitysaste on $r^2 = 0{,}972\ldots \approx 0{,}97$.

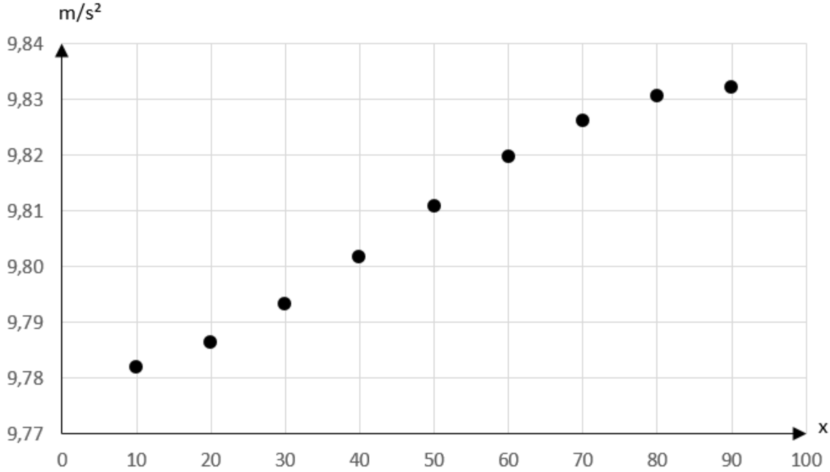
d) Selitysaste r^2 kuvaa mallin yhteensopivuutta havaintoaineiston kanssa. Kaikkien mallien selitysaste on lähellä lukua 1, joten kaikki mallit kuvaavat riippuvuutta hyvin.

b-kohdan lineaarisen mallin selitysaste on $r^2 \approx 0{,}95$, eli lähellä lukua 1, joten malli sopii tilanteeseen hyvin. a- ja c-kohtien mallien selitysasteet vielä hieman suurempia, joten a- ja c-kohtien mallit kuvaavat muuttujien x ja y riippuvuutta parhaiten.

96.

Kirjoitetaan leveyspiirin arvot x ja niitä vastaavat putoamiskiihtyvyyden arvot y taulukkolaskentaohjelman taulukkoon.

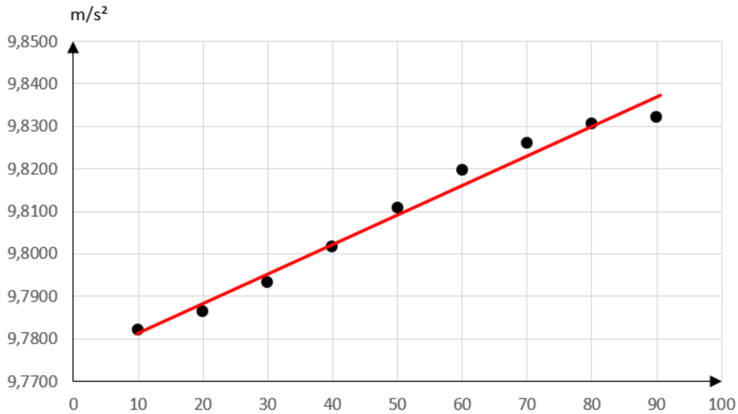
a) Piirretään pistejoukon (x, y) kuvaaja.



b) Sovitetaan pistejoukkoon regressiokäyrät. Ilmoitetaan kertoimet kahden merkitsevän numeron tarkkuudella ja vakiotermit kolmen merkitsevän numeron tarkkuudella:

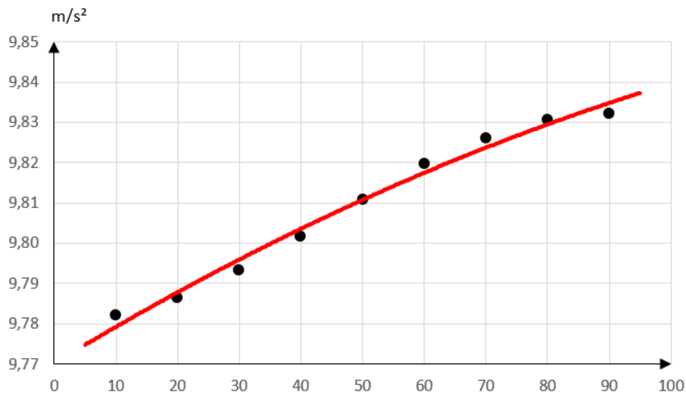
Lineaarisen mallin regressiosuoran yhtälö on $y \approx 0{,}00070x + 9{,}77$

Lineaarisen mallin selitysaste on $r^2 = 0{,}9811 \ldots \approx 0{,}981$.



Toisen asteen polynomisen mallin regressioyhtälö on
$y \approx -2{,}3 \cdot 10^{-6}x^2 + 0{,}00093x + 9{,}77$

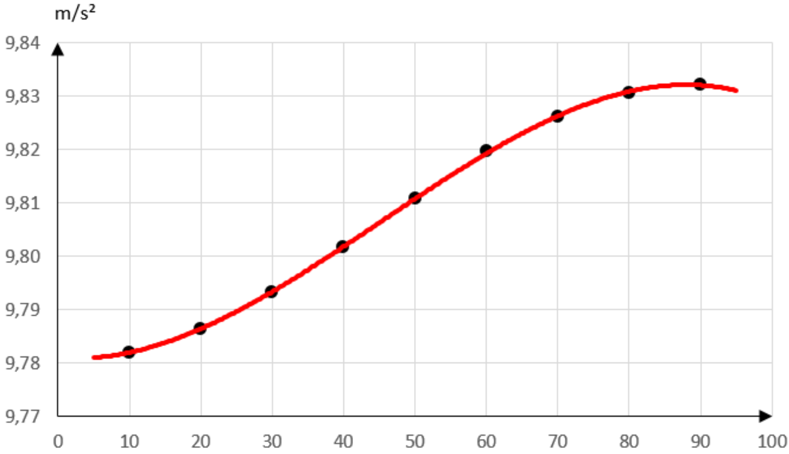
Mallin selitysaste on $r^2 = 0{,}9869 \ldots \approx 0{,}987$.



Kolmannen asteen polynomisen mallin regressioyhtälö on

$$y \approx -1{,}6 \cdot 10^{-7}x^3 + 2{,}2 \cdot 10^{-5}x^2 - 0{,}00011x + 9{,}78.$$

Mallin selitysaste on $r^2 = 0{,}9998 \ldots \approx 1{,}000$.



Selitysaste r^2 kuvaa mallin yhteensopivuutta havaintoaineiston kanssa, joten parhaiten riippuvuutta kuvaa se malli, jonka selitysaste on suurin.

Sopivin malli on siis kolmannen asteen polynominen malli, koska sen selitysaste on lähimpänä lukua 1.

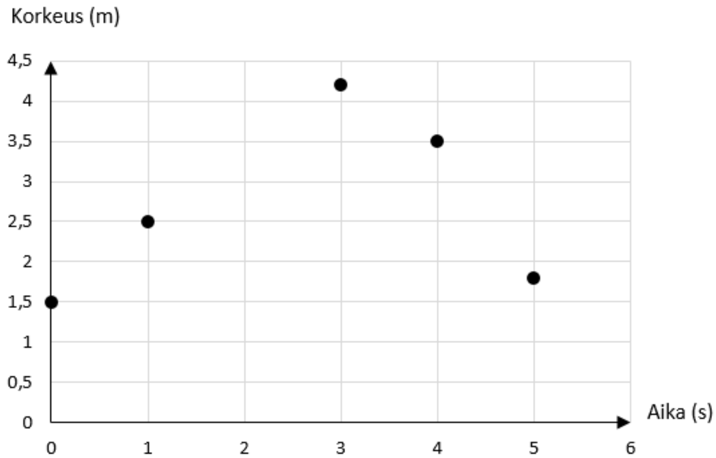
c) Sijoitetaan $x = 0$ kolmannen asteen polynomisen regressiokäyrän yhtälöön. Mallin mukaan

$$y \approx -1{,}6 \cdot 10^{-7} \cdot 0^3 + 2{,}2 \cdot 10^{-5} \cdot 0^2 - 0{,}00011 \cdot\ 0 + 9{,}78 = 9{,}78.$$

Huomaa, että laskussa käytetään regressiokäyrän kertoimille ja vakiolle pyöristettyjä arvoja. Putoamiskiihtyvyyden arvo päiväntasaajalla on siis mallin mukaan 9,78 m/s^2.

97.

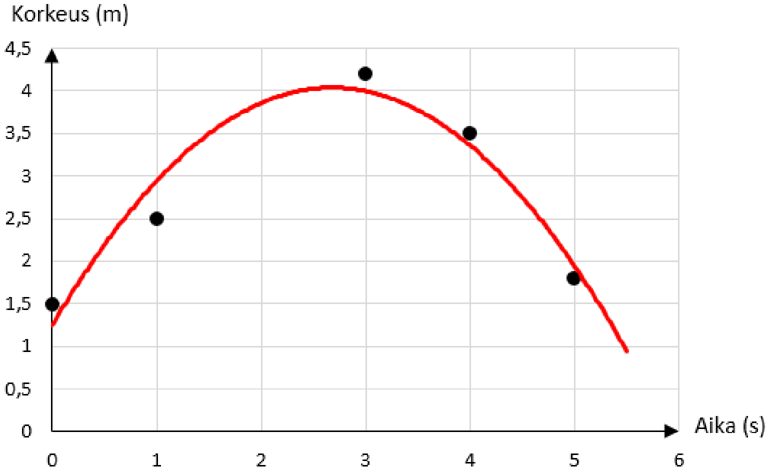
a) Kirjoitetaan ajat x ja niitä vastaavat korkeudet y taulukkolaskentaohjelman taulukkoon. Piirretään pistejoukon (x, y) kuvaaja.



Pisteiden sijoittumisesta nähdään heti, että lineaarinen malli ei sovellu kuvaamaan muuttujien välistä riippuvuutta. Myöskään eksponentiaalinen malli ei sovellu kuvaamaan riippuvuutta.

Kuvataan riippuvuutta polynomisella mallilla. Polynomisen mallin selitysaste paranee asteluvun kasvaessa. Toisen asteen polynomisen mallin selitysasteeksi saadaan $r^2 \approx 0{,}934$ joka on riittävän lähellä lukua 1 kuvatakseen riippuvuutta hyvin.

Valitaan malliksi toisen asteen polynominen malli.



b) Mallin regressioyhtälö on

$y = -0{,}388\ldots x^2 + 2{,}079\ldots x + 1{,}253\ldots \approx -0{,}39x^2 + 2{,}1x + 1{,}3.$

c) Pallon korkeus saadaan sijoittamalla $x = 2{,}0$ (s) regressiokäyrän $y = -0{,}39x^2 + 2{,}1x + 1{,}3$ yhtälöön. Mallin mukaan

$$y = -0{,}39 \cdot 2{,}0^2 + 2{,}1 \cdot 2{,}0 + 1{,}3 = 3{,}94 \approx 3{,}9.$$

Pallon korkeus ajanhetkellä 2,0 sekuntia on siis 3,9 metriä.

d) Pallo putoaa maahan hetkellä, jolloin sen korkeus on nolla, eli $y = 0$. Sijoitetaan regressiokäyrän yhtälöön $y = 0$ ja ratkaistaan yhtälöstä *x:*

$$0 = -0{,}39x^2 + 2{,}1x + 1{,}3$$

Tämä on toisen asteen yhtälö: $-0{,}39x^2 + 2{,}1x + 1{,}3 = 0$. Se voidaan ratkaista symbolisen laskennan ohjelmalla tai ratkaisukaavalla:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$ Tässä yhtälössä: $a = -0{,}39 \quad b = 2{,}1 \quad c = 1{,}3$

$$x = \frac{-2{,}1 \pm \sqrt{2{,}1^2 - 4 \cdot (-0{,}39) \cdot 1{,}3}}{2 \cdot (-0{,}39)}$$

$$= \frac{-2{,}1 \pm \sqrt{6{,}438}}{-0{,}78}$$

$$= \frac{-2{,}1 \pm 2{,}537\ldots}{-0{,}78}$$

Yhtälöllä on kaksi ratkaisua:

$$x = \frac{-2{,}1 + 2{,}537\ldots}{-0{,}78} = -0{,}560\ldots$$

tai

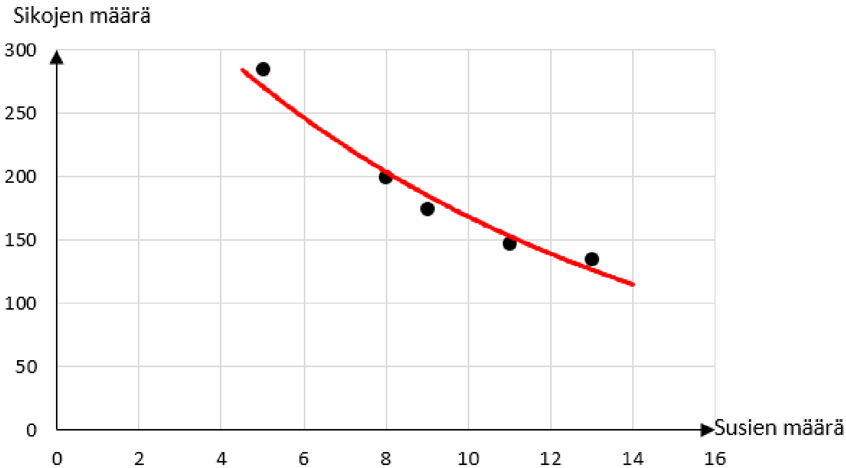
$$x = \frac{-2{,}1 - 2{,}537\ldots}{-0{,}78} = 5{,}945\ldots$$

Muuttuja *x* kuvaa ajanhetkiä lähtöhetkestä $x = 0$ eteenpäin, joten ainoastaan positiivinen ratkaisu kelpaa. Pallo putoaa maahan 5,9 sekunnin kuluttua.

98.

a) Sikojen määrä riippuu susien määrästä, joten selittävänä muuttujana on susien määrä x ja selitettävänä muuttujana on sikojen määrä y.

Kirjoitetaan susien määrät x ja niitä vastaavat sikojen määrät y taulukkolaskentaohjelman taulukkoon. Piirretään pistejoukon (x, y) kuvaaja ja sovitetaan siihen eksponentiaalinen malli.



	A	B
1	**Susien määrä**	**Sikojen määrä**
2	13	135
3	11	148
4	9	175
5	8	200
6	5	285

b) Eksponentiaalisen mallin regressikäyrä riippuu käytetystä ohjelmasta:

$$y = 436{,}028\ldots \cdot e^{-0{,}09501\ldots\cdot x} \approx 436 \cdot e^{-0{,}0950x}$$

tai

$$y = 436{,}028\ldots \cdot (0{,}9093\ldots)^x \approx 436 \cdot 0{,}909^x.$$

Huomaa, että kertoimet pyydettiin kolmen merkitsevän numeron tarkkuudella.

c) ”Susia ei ole” vastaa muuttujan arvoa $x = 0$. Lasketaan muuttujan y arvo, kun regressiokäyrän yhtälössä $x = 0$. Jos käytetään mallia $y = 436 \cdot e^{-0{,}0950x}$, niin saadaan

$$y \approx 436 \cdot e^{-0{,}0950\cdot 0} = 436 \cdot e^0 = 436 \cdot 1 = 436.$$

Sama lukuarvo saadaan, jos käytetään mallia $y = 436 \cdot 0{,}909^x$.

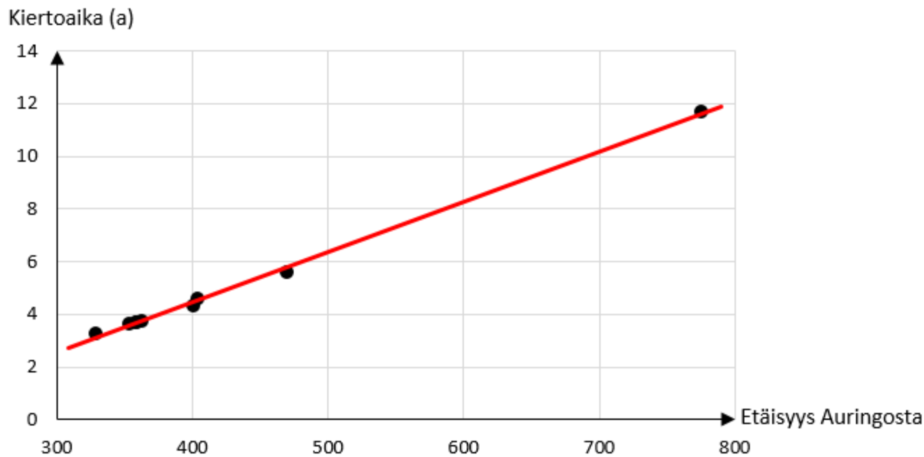
Villisikapopulaation koko on siis 436 sikaa sitten, kun susia ei enää ole.

99.

a) Asteroidin etäisyys Auringosta vaikuttaa sen kiertoaikaan Auringon ympäri. Selittävä muuttuja x on siis etäisyys ja selitettävä muuttuja y on kiertoaika.

b) Kirjoitetaan asteroidien etäisyydet x ja niitä vastaavat kiertoajat y taulukkolaskentaohjelman taulukkoon. Piirretään pistejoukon (x, y) kuvaaja ja sovitetaan pistejoukkoon regressiosuora.

	A	B
1	**Etäisyys**	**Kiertoaika (a)**
2	404	4,6
3	353	3,63
4	470	5,59
5	401	4,36
6	775	11,7
7	358	3,68
8	362	3,78
9	358	3,69
10	329	3,27



c) Regressiosuoran yhtälö on $y = 0{,}0190\ldots x - 3{,}14\ldots \approx 0{,}019x - 3{,}1$.

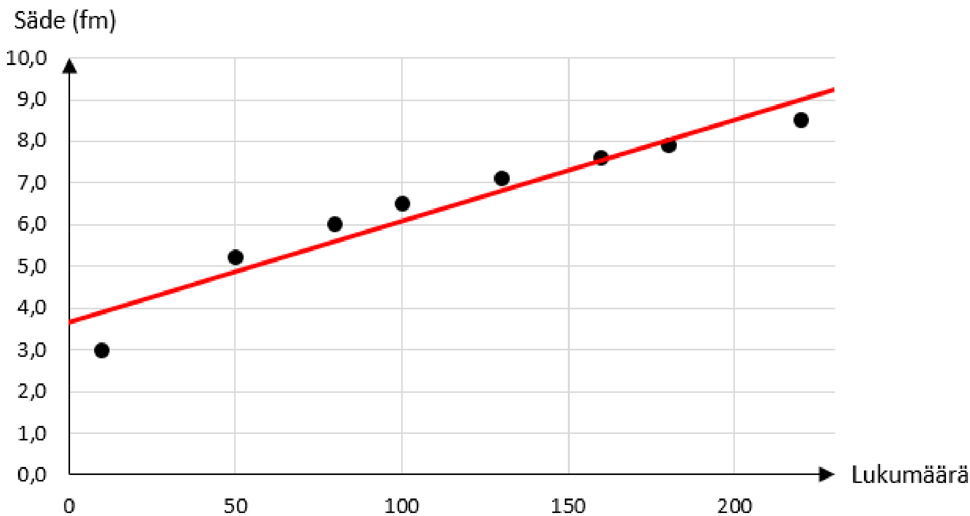
Korrelaatiokerroin on $r = 0{,}9990\ldots \approx 0{,}999$.

d) Korrelaatiokerroin on hyvin lähellä lukua 1, joten asteroidin kiertoaika riippuu voimakkaasti sen etäisyydestä. Kiertoaika voidaan siis arvioida etäisyyden avulla.

100.

a) Kirjoitetaan ytimessä olevien hiukkasten lukumäärät A ja niitä vastaavat säteet r taulukkolaskentaohjelman taulukkoon. Piirretään pistejoukon (A, r) kuvaaja ja sovitetaan pistejoukkoon regressiosuora.

	A	B
1	**Hiukkasten lukumäärä**	**Säde r (fm)**
2	10	3,0
3	50	5,2
4	80	6,0
5	100	6,5
6	130	7,1
7	160	7,6
8	180	7,9
9	220	8,5



Regressiosuoran yhtälö on

$$y = 0{,}0242\ldots x + 3{,}655\ldots \approx 0{,}024x + 3{,}7.$$

Mallissa x = lukumäärä (eli A) ja y = säde (eli r).

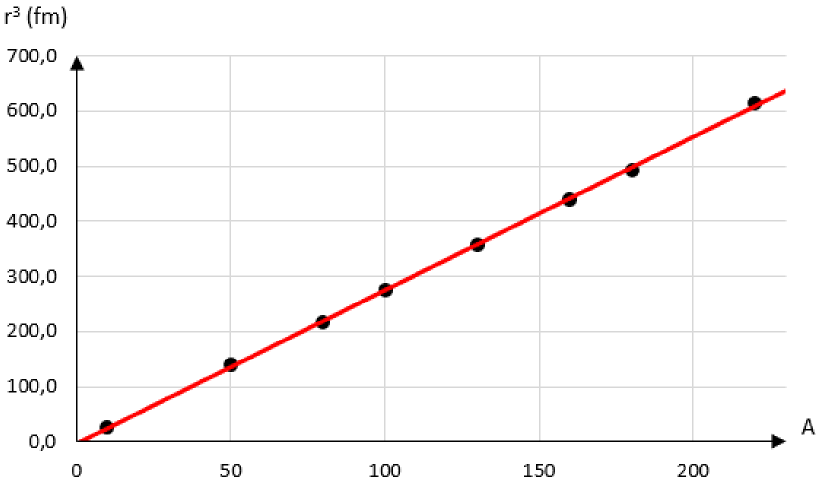
Kerroin ja vakio pyöristetään kahden merkitsevän numeron tarkkuuteen.

b) Lasketaan taulukkoon säteen pituuksien kuutiot:

- soluun C2 kirjoitetaan laskukaava: = B2^3
- kaava kopioidaan sarakkeen C seuraaville riveille

	A	B	C
1	**Hiukkasten lukumäärä**	**Säde r (fm)**	$\mathbf{r^3}$
2	10	3,0	27,0
3	50	5,2	140,6
4	80	6,0	216,0
5	100	6,5	274,6
6	130	7,1	357,9
7	160	7,6	439,0
8	180	7,9	493,0
9	220	8,5	614,1
10			

Piirretään pistejoukon (A, r^3) kuvaaja ja sovitetaan siihen regressiosuora.



Regressiosuoran yhtälö riippuu käytetystä ohjelmasta:

$$y = 2{,}77 \ldots x - 2{,}05 \ldots \approx 2{,}8x - 2{,}1$$

tai

$$y \approx 2{,}7x.$$

Mallissa x = lukumäärä (eli A) ja y = säteen kuutio (eli r^3).

c) a-kohdan suoran korrelaatiokerroin on $r = 0{,}9624\ldots \approx 0{,}96$. Muunnoksella saadun b-kohdan suoran korrelaatiokerroin on $r_{\text{muunnos}} = 0{,}9998\ldots \approx 1{,}0$.

Korrelaatiokerroin kuvaa sitä, miten hyvin pistejoukon pisteet toteuttavat regressiosuoran yhtälön. b-kohdan muunnoksella saatu suora on hieman parempi malli, koska muunnoksella saadun pistejoukon korrelaatiokerroin on hieman suurempi kuin a-kohdan pistejoukon korrelaatiokerroin.

Muunnoksella saatu suora on parempi malli.

d) Sijoitetaan lukumäärä $x = 204$ muunnoksella saatuun regressiomalliin:

$$y = 2{,}8 \cdot 204 - 2{,}1 = 569{,}1$$

tai

$$y = 2{,}7 \cdot 204 = 550{,}8.$$

Huomaa, että laskussa käytetään regressiosuoran kertoimelle ja vakiolle pyöristettyjä arvoja.

Mallissa muuttuja y ilmaisee säteen pituuden r kuution, eli $r^3 = 569{,}1$ tai $r^3 = 550{,}8$. Ratkaistaan säde r kuutiojuurella:

$$r = \sqrt[3]{569{,}1} = 8{,}286\ldots \approx 8{,}3$$

tai

$$r = \sqrt[3]{550{,}8} = 8{,}197\ldots \approx 8{,}2.$$

Mallit antavat lyijyn isotoopin Pb-204 ytimen säteen pituudeksi 8,2 fm ja 8,3 fm.

101.

Kopioidaan annetusta taulukosta 100 metrin juoksun pistetulokset ja kuulantyönnön pistetulokset taulukkolaskentaohjelmaan sarakkeisiin A ja B.

Lasketaan pistetulosten summa:

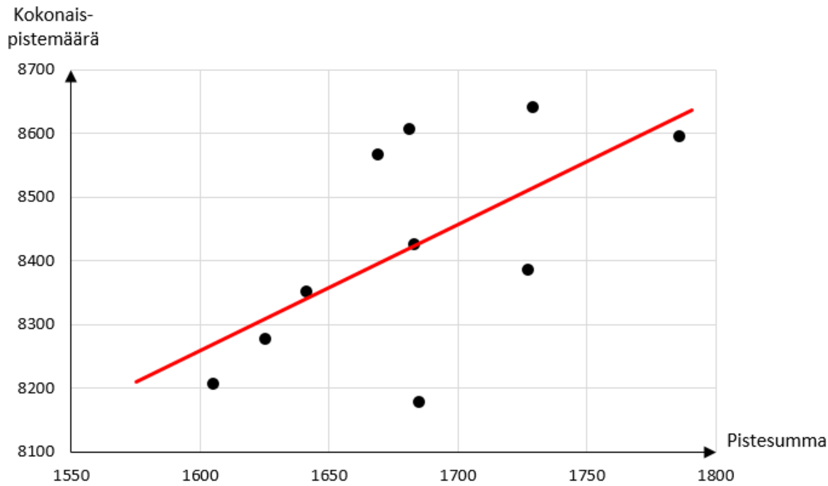
- soluun C2 kirjoitetaan laskukaava: = A2 + B2
- kaava kopioidaan sarakkeen C seuraaville riveille

Kopioidaan annetusta taulukosta kokonaispistemäärät sarakkeeseen D.

	A	B	C	D
1	**100m**	**kuula**	**summa**	**kokonais-pisteet**
2	933	796	1729	8641
3	878	803	1681	8606
4	980	806	1786	8595
5	903	766	1669	8567
6	901	782	1683	8425
7	881	846	1727	8385
8	881	760	1641	8351
9	838	787	1625	8277
10	874	731	1605	8206
11	897	788	1685	8178

Piirretään sarakkeiden C ja D avulla sirontakuvaaja: selittävän muuttujan (juoksun ja kuulantyönnön pistesumman) arvot tulevat x-akselille ja selitettävän muuttujan (kokonaispistemäärän) arvot tulevat y-akselille.

Sovitetaan pistejoukkoon suora.



Riippuvuutta kuvaavan suoran yhtälö on

$$y = 1{,}979444 \ldots x + 5091{,}48 \ldots \approx 1{,}9794x + 5091{,}5.$$

Kerroin ja vakio on pyöristetty viiden merkitsevän numeron tarkkuuteen.

Korrelaatiokertoimen arvo on

$$r = 0{,}6193 \ldots \approx 0{,}619.$$

Korrelaatiokertoimen itseisarvo $|r| = 0{,}619$ on välillä 0,6–0,8, joten lineaarinen riippuvuus on huomattava. Sadan metrin juoksun ja kuulantyönnön pistesumma ennustaa siis hyvin lopullista menestymistä kymmenottelussa.

Hämäläisen 100 metrin juoksun pistetuloksen ja kuulantyönnön pistetuloksen summa on

$$x = 858 + 732 = 1590.$$

Mallin mukaan Hämäläisen odotettavissa oleva kokonaispistemäärä on tällöin

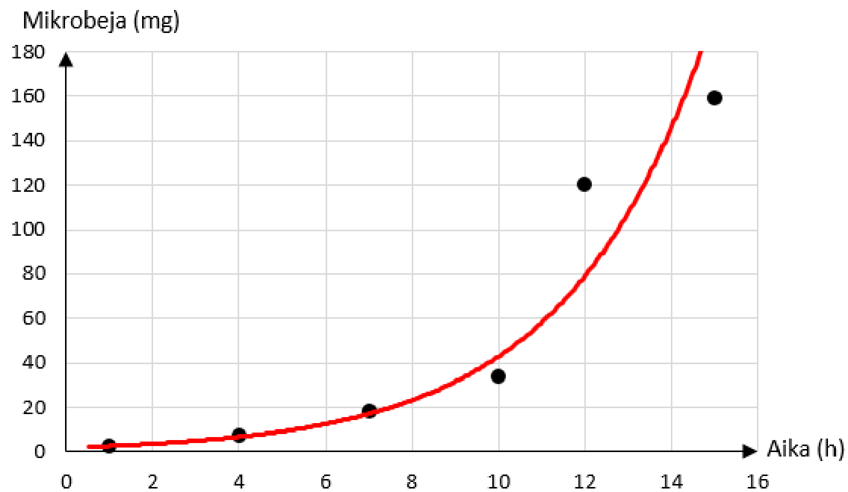
$$y = 1{,}9794 \cdot 1590 + 5091{,}5 = 8238{,}736 \approx 8239 \text{ (pistettä)}.$$

Hämäläisen odotettavissa oleva lopullinen pistemäärä olisi ollut 8239 pistettä.

102.

a) Mikrobien määrä kasvaa ajan kuluessa, joten selittävänä muuttujana on aika x ja selitettävänä muuttujana on mikrobien määrä y. Kirjoitetaan ajanhetket x ja niitä vastaavat mikrobien määrät y taulukkolaskentaohjelman taulukkoon. Piirretään pistejoukon (x, y) kuvaaja ja sovitetaan siihen eksponentiaalinen malli.

	A	B
1	**Aika (h)**	**Mikrobeja (mg)**
2	1	2,5
3	4	7,3
4	7	18,1
5	10	33,8
6	12	120,4
7	15	159,5



b) Eksponentiaalisen mallin regressiokäyrä riippuu käytetystä ohjelmasta:

$$y = 2{,}011\ldots \cdot e^{0{,}3059\ldots \cdot x} \approx 2{,}01 \cdot e^{0{,}306x}$$

tai

$$y = 2{,}011\ldots \cdot (1{,}357\ldots)^x \approx 2{,}01 \cdot 1{,}36^x.$$

Huomaa, että kertoimet pyydettiin kolmen merkitsevän numeron tarkkuudella.

c) Mikrobien määrä saadaan sijoittamalla ajanhetki $x = 20$ (tuntia) regressiokäyrän yhtälöön.

Jos käytetään mallia $y = 2{,}01 \cdot e^{0{,}306x}$, niin

$$y = 2{,}01 \cdot e^{0{,}306 \cdot 20} = 914{,}27\ldots \approx 910,$$

joten mikrobien määrä 20 tunnin kuluttua on 910 mg.

Jos käytetään mallia $y = 2{,}01 \cdot 1{,}36^x$, niin

$$y = 2{,}01 \cdot 1{,}36^{20} = 941{,}83\ldots \approx 940,$$

joten mikrobien määrä 20 tunnin kuluttua on 940 mg.

d) Kokeen alkaessa ajanhetki on $x = 0$. Lasketaan muuttujan y arvo, kun regressiokäyrän yhtälössä $x = 0$.

$$y = 2{,}01 \cdot e^{0{,}306 \cdot 0} = 2{,}01 \cdot e^{0} = 2{,}01 \cdot 1 = 2{,}01.$$

Sama lukuarvo saadaan, jos käytetään mallia $y = 2{,}01 \cdot 1{,}36^{x}$.

Mikrobien määrä jauhelihassa oli kokeen alkaessa siis 2,0 g.