

8 TILASTOT

ALOITA PERUSTEISTA

323A. Keskiarvo on pituuksien summan ja lukumäärän osamäärä, joten A ja III kuuluvat yhteen.

Keskihajonta mittaa havaintoarvojen ryhmittymistä keskiarvon ympärille, joten B ja I kuuluvat yhteen.

Mediaani on suuruusjärjestykseen asetetuista havaintoarvoista keskimäinen tai keskimmäisten keskiarvo, joten C ja II kuuluvat yhteen.

Moodi on havaintoarvoista se, jonka frekvenssi on suurin eli moodi on yleisin havaintoarvo, joten D ja IV kuuluvat yhteen.

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

324A. a) Polkupyörien määrät suuruusjärjestyksessä ovat: 0, 0, 1, 2, 2, 2, 3, 4, 5, 5.

Vastaus: 0, 0, 1, 2, 2, 2, 3, 4, 5, 5

b) Koska lukumäärää kaksi on eniten, moodi on $M_o = 2$.

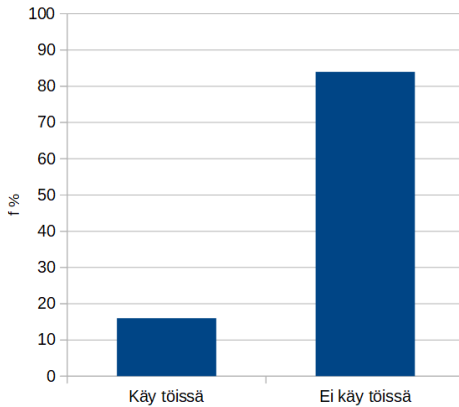
Lukuja on yhteensä 10, joten 5. ja 6. lukumäärä ovat keskimäiset. 5. lukumäärä on 2 ja 6. lukumäärä on 2. Näiden keskiarvo on 2, joten mediaani on 2.

$$\text{Keskiarvo on } \bar{x} = \frac{0+0+1+2+2+2+3+4+5+5}{10} = 2,4.$$

Vastaus: $M_o = 2$, $M_d = 2$ ja $\bar{x} = 2,4$

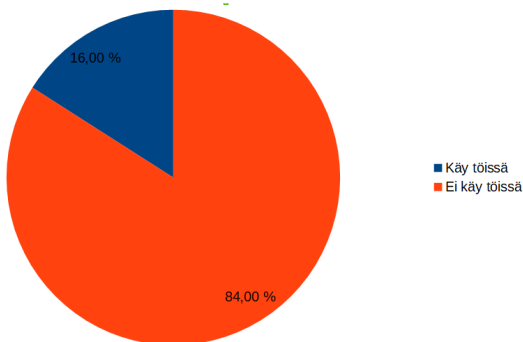
325B. a) Kopioidaan taulukon tiedot taulukkolaskentaohjelmaan ja piirretään pylväskaavio.

Vastaus:



b) Kopioidaan taulukon tiedot taulukkolaskentaohjelmaan ja piirretään ympyräkaavio.

Vastaus:



- 326A. a)** Kuvaajasta nähdään, että kun ikä on 15, suhteellinen summafrekvenssi on noin 40, joten alle 15-vuotiaita on noin 40 %.

Vastaus: n. 40 %

- b)** Kuvaajasta nähdään, että kun suhteellinen summafrekvenssi on 50, ikä on noin 20, joten mediaani-ikä on noin 20 vuotta.

Vastaus: n. 20 vuotta

- c)** Kuvaajasta nähdään, että kun suhteellinen summafrekvenssi on 90, ikä on noin 53 vuotta, joten sitä nuorempia on noin 90 % keski-afrikkalaisista.

Vastaus: n. 53 vuotta

- 327A.** Täydennetään taulukkoon summafrekvenssit sf , suhteelliset frekvenssit $f\%$ ja suhteelliset summafrekvenssit $sf\%$.

Lumitöiden tekemiskertojen määrä	f	sf	$f\%$	$sf\%$
0	19	19	$\frac{19}{31} \approx 61\%$	$\frac{19}{31} \approx 61\%$
1	6	$19 + 6 = 25$	$\frac{6}{31} \approx 19\%$	$\frac{25}{31} \approx 81\%$
2	4	$25 + 4 = 29$	$\frac{4}{31} \approx 13\%$	$\frac{29}{31} \approx 94\%$
3	2	$29 + 2 = 31$	$\frac{2}{31} \approx 6\%$	$\frac{31}{31} = 100\%$

Vastaus:

Lumitöiden tekemiskertojen määrä	f	sf	$f\%$	$sf\%$
0	19	19	61 %	61 %
1	6	25	19 %	81 %
2	4	29	13 %	94 %
3	2	31	6 %	100 %

328A. Jakauman B suurimmat frekvenssit ovat lähellä toisiaan ja pienimmät molemmiin puolin laitimmaisina, joten siinä on pienin keskihajonta.

Jakauman A suurimmat frekvenssit ovat ääri-laidoilla, joten sen keskihajonta on suurin.

Jakauman C arvot ovat jakautuneet melko tasaisesti koko vaihteluvälille, joten sen keskihajonta A:n ja C:n keskihajonnan välissä.

Jakaumien A–C keskihajonnan mukainen suuruusjärjestys pienimmästä alkaen on B, C, A.

Vastaus: B, C, A

329A. Kuvan A havaintoarvot asettuvat varsin hyvin nousevan suoran läheisyyteen, joten korrelaatio on selvästi positiivinen. A ja IV kuuluvat yhteen.

Kuvan B havaintoarvot asettuvat varsin hyvin laskevan suoran läheisyyteen, joten korrelaatio on selvästi negatiivinen. B ja I kuuluvat yhteen.

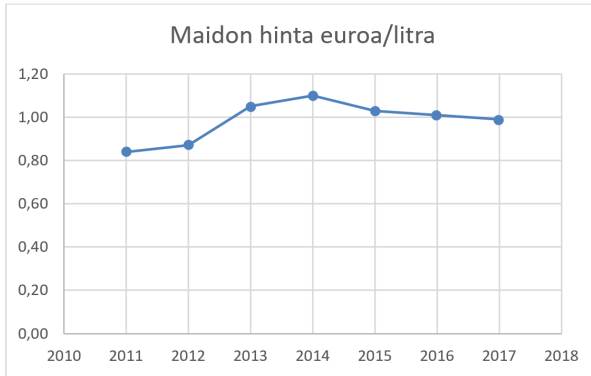
Kuvan C havaintoarvot näyttävät asettuvan paraabelille, joten muuttujien x ja y välillä ei ole lineaarista riippuvuutta. C ja III kuuluvat yhteen.

Kuvan D havaintoarvot asettuvat melko hyvin laskevan suoran läheisyyteen, joten korrelaatio on jonkin verran negatiivinen. D ja II kuuluvat yhteen.

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

330B. Piirretään aineistosta viivakaavio.

Vastaus:



331A. a) Uintikertojen keskiarvo on $\bar{x} = \frac{7+10+0+2+1+4}{6} = 4$.

Vastaus: $\bar{x} = 4$ kertaa

b) Keskihajonta on

$$s = \sqrt{\frac{(7-4)^2 + (10-4)^2 + (0-4)^2 + (2-4)^2 + (1-4)^2 + (4-4)^2}{6-1}}$$

$$= 3,847\dots$$

$$\approx 3,8$$

Vastaus: $s = 3,8$ kertaa

332A. Niiden, jotka eivät ole ollenkaan käyttäneet, suhteellinen frekvenssi on suurin, joten moodiluokka on 0.

Kahteen ensimmäiseen luokkaan kuuluu $27\% + 16\% = 43\%$ ja seuraavaan luokkaan 16% . 50% :n raja ylittyy luokan 3–5 kohdalla, joten mediaaniluokka on 3–5.

Vastaus: moodiluokka: 0, mediaaniluokka: 3–5

333A. a) Hammaslääkäriissäkäyntikertojen keskiarvo on

$$\begin{aligned}\bar{x} &= \frac{13 \cdot 0 + 22 \cdot 1 + 7 \cdot 2 + 6 \cdot 3 + 2 \cdot 4}{50} \\ &= 1,24 \\ &\approx 1,2\end{aligned}$$

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

b) Keskihajonta on

$$\begin{aligned}s &= \sqrt{\frac{13 \cdot (0 - \bar{x})^2 + 22 \cdot (1 - \bar{x})^2 + 7 \cdot (2 - \bar{x})^2 + 6 \cdot (3 - \bar{x})^2 + 2 \cdot (4 - \bar{x})^2}{50 - 1}} \\ &= 1,098\dots \\ &\approx 1,10\end{aligned}$$

Vastaus: $s = 1,10$

334B. Sopivalla ohjelmalla keskiarvoksi saadaan $\bar{x} = 4,860\dots \approx 4,86$ ja keskihajonnaksi $s = 1,257\dots \approx 1,26$.

n	86
Keskiarvo	4.8605
σ	1.2498
s	1.2571
Σx	418
Σx^2	2166
Min	2
Q1	4
Mediaani	5
Q3	6
Max	7

Vastaus: $\bar{x} = 4,86$, $s = 1,26$

335B. Täydennetään taulukkoon todelliset rajat ja luokkakeskukset.

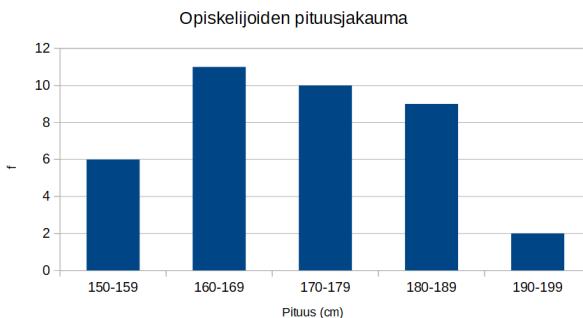
Pituus (cm)	Todellinen alaraja (cm)	Todellinen yläraja (cm)	Luokkakeskus (cm)
150–159	149,5	159,5	$\frac{149,5 + 159,5}{2} = 154,5$
160–169	159,5	169,5	$\frac{169,5 + 179,5}{2} = 164,5$
170–179	169,5	179,5	$\frac{179,5 + 189,5}{2} = 174,5$
180–189	179,5	189,5	$\frac{179,5 + 189,5}{2} = 184,5$
190–199	189,5	199,5	$\frac{189,5 + 199,5}{2} = 194,5$

Lasketaan opiskelijoiden pituuksien keskiarvo luokkakeskusten ja frekvenssien avulla.

Pituus (cm)	Luokkakeskus (cm)	f
150–159	154,5	6
160–169	164,5	11
170–179	174,5	10
180–189	184,5	9
190–199	194,5	2

$$\begin{aligned}\bar{x} &= \frac{6 \cdot 154,5 + 11 \cdot 164,5 + 10 \cdot 174,5 + 9 \cdot 184,5 + 2 \cdot 194,5}{38} \\ &= 171,868... \\ &\approx 171,9\end{aligned}$$

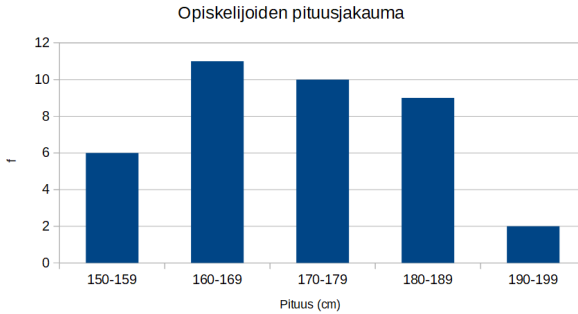
Piirretään pylväskaavio sopivalla ohjelmalla.



Vastaus:

Pituus (cm)	Luokkakeskus (cm)
150–159	154,5
160–169	164,5
170–179	174,5
180–189	184,5
190–199	194,5

$$\bar{x} = 171,9 \text{ cm}$$



336A. Täydennetään taulukkoon todelliset rajat ja luokkakeskukset.

Hinta (€)	Todellinen alaraja (€)	Todellinen yläraja (€)	Luokkakeskus (€)	f
20,00–29,99	19,995	29,995	$\frac{19,995 + 29,995}{2} = 24,995$	4
30,00–39,99	29,995	39,995	$\frac{29,995 + 39,995}{2} = 34,995$	19
40,00–49,99	39,995	49,995	$\frac{39,995 + 49,995}{2} = 44,995$	17
50,00–59,99	49,995	59,995	$\frac{49,995 + 59,995}{2} = 54,995$	6

Lasketaan joulukuusien hintojen keskiarvo.

$$\begin{aligned}\bar{x} &= \frac{4 \cdot 24,995 + 19 \cdot 34,995 + 17 \cdot 44,995 + 6 \cdot 54,995}{46} \\ &= 40,429\dots \\ &\approx 40,43\end{aligned}$$

Joulukuusien hintojen keskiarvo on 40,43 €.

Vastaus: 40,43 €

337B. a) Kopioidaan taulukon tiedot sopivaan ohjelmaan.

n	50
Keskiarvo	202.82
σ	9.3759
s	9.4711
Σx	10141
Σx^2	2061193
Min	183
Q1	196
Mediaani	204.5
Q3	211
Max	216

Keskiarvoksi saadaan $\bar{x} = 202,8$ cm ja keskihajonnaksi $s = 9,471\dots$ cm $\approx 9,47$ cm.

Vastaus: $\bar{x} = 202,8$ cm, $s = 9,47$ cm

- b) Koripalloilijoiden pituuden ja sijoituksen välinen korrelaatiokerroin on $r = 0,0682\dots \approx 0,068$ ja selitysaste $r^2 = 0,00465\dots \approx 0,005$.

KeskiarvoX	202.82
KeskiarvoY	25.5
Sx	9.4711
Sy	14.5774
r	0.0682
p	0.0806
Sxx	4395.38
VarianssiY	10412.5
Sxy	461.5

Vastaus: $r = 0,068$, $r^2 = 0,005$

- c) Kohdan b perusteella pelaajan pituuden ja sijoituksen välillä ei ole korrelaatiota, joten väite ”Mitä pidempi koripalloilija, sitä paremmin hän pelaa” ei pidä paikkaansa.

Vastaus: ei ole

VAHVISTA OSAAMISTA

- 338A. a)** Vaihteluvälin pituus saadaan laskemalla suurimman ja pienimmän arvon erotus. Lämpötilan vaihteluvälin pituus maailmassa on $57\text{ °C} - (-89\text{ °C}) = 146\text{ °C}$.

Vastaus: 146 °C

- b)** Merkitään alinta Death Valleyssä mitattua lämpötilaa kirjaimella x . Muodostetaan yhtälö ja ratkaistaan siitä lämpötila x .

$$\begin{aligned} 57 - x &= 67 \\ x &= -10 \end{aligned}$$

Alin lämpötila Death Valleyssä on -10 °C .

Vastaus: -10 °C

- c)** Merkitään korkeinta Vostokissa mitattua lämpötilaa kirjaimella x . Muodostetaan yhtälö ja ratkaistaan siitä lämpötila x .

$$\begin{aligned} x - (-89) &= 75 \\ x &= -14 \end{aligned}$$

Ylin lämpötila Vostokissa on -14 °C .

Vastaus: -14 °C

- 339A. a)** Kuvaajalta nähdään, että 60 minuutin kestoaikaa vastaa suhteellisen summafrekvenssiin arvo noin 5 %, joten noin 5 prosenttia akuista kesti korkeintaan tunnin.

Vastaus: n. 5 %

- b)** Kolme tuntia on 180 minuuttia, jota vastaava suhteellisen summafrekvenssiin arvo noin 55 %, joten likimain $100\% - 55\% = 45\%$ akuista kesti yli kolme tuntia.

Vastaus: n. 45 %

- c) Suhteellisen summafrekvenssin arvoa 50 % vastaava kestoikä on noin 170 minuuttia, joten kestoajan mediaani on noin 170 min.

Suhteellisen summafrekvenssin arvoa 25 % vastaava kestoikä on noin 120 minuuttia, joten kestoajan alakvartiili Q_1 on noin 120 min.

Suhteellisen summafrekvenssin arvoa 75 % vastaava kestoikä on noin 215 minuuttia, joten kestoajan yläkvartiili Q_3 on noin 215 min.

Vastaus: $M_d \approx 170$ min, $Q_1 \approx 120$ min, $Q_3 \approx 215$ min

- d) Suhteellisen summafrekvenssin arvoa 90 % vastaava kestoikä on noin 240 minuuttia, joten parhaat 10 % akuista kestivät noin 240 min.

Vastaus: n. 240 min

- 340A. a) Selitysaste on $r^2 = 0,59^2 = 0,3481 \approx 0,35 = 35$ %. Muuttuja A selittää 35 % muuttujan B arvon vaihtelusta.

Vastaus: 35 %

- b) Selitysaste $r^2 = 74,2$ % = 0,742 ratkaistaan korrelaatiokerroin r yhtälöstä

$$r^2 = 0,742$$

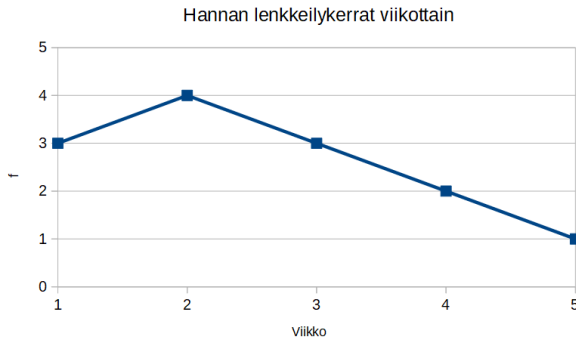
$$r = (\pm) 0,861\dots$$

$$r \approx 0,86$$

Vastaus: $r = 0,86$

341B a) Piirretään viivakaavio sopivalla ohjelmalla.

Vastaus:



b) Merkitään seuraavan viikon lenkkeilykertojen määrää kirjaimella x . Muodostetaan yhtälö ja ratkaistaan siitä x .

$$\frac{3 + 4 + 3 + 2 + 1 + x}{6} = 3$$

$$\frac{13 + x}{6} = 3 \quad || \cdot 6$$

$$13 + x = 18$$

$$x = 5$$

Hannan on käytävä seuraavalla viikolla lenkillä vähintään 5 kertaa, jotta lenkkeilykertojen keskiarvo olisi vähintään 3.

Vastaus: vähintään 5 kertaa

342A Tarkastellaan ohjelman ilmoittamia tunnuslukuja.

Havaintojen lukumäärä on $n = 50$.
 Keskiarvo on $\bar{x} = 132,0323 \approx 132$
 Keskihajonta on $s = 61,5394 \approx 62$.
 Min on pienin havaintoarvo
 Alakvartiili on $Q1 = 89,7786 \approx 90$.
 Mediaani on $Md = 133,8062 \approx 134$
 Yläkvartiili on $Q3 = 172,3969 \approx 172$.
 $\text{Max} = 290,9595 \approx 291$ on suurin havaintoarvo.

n	50
Keskiarvo	132.0323
σ	60.9209
s	61.5394
Σx	6601.6154
Σx^2	1057194.594
Min	6.1206
Q1	89.7786
Mediaani	133.8062
Q3	172.3969
Max	290.9595

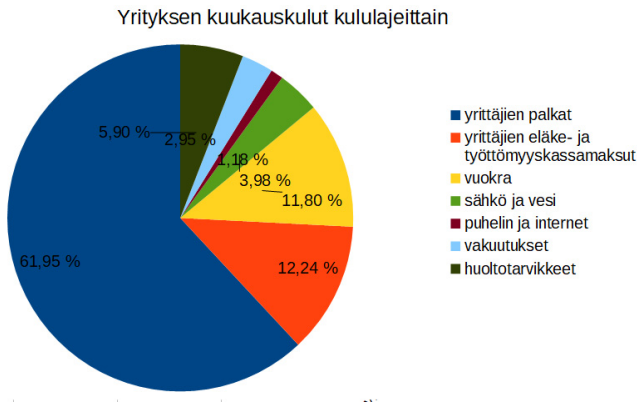
Täydennetään lauseet.

- Aikojen keskiarvo on 132 minuuttia ja mediaani 134 minuuttia.
- Aineiston vaihteluväli on 6 min–291 min ja keskihajonta 62 minuuttia.
- Aineistossa on 50 havaintoarvoa.
- Päivystyksessä kävijöistä 25 % oli terveyskeskuksessa korkeintaan 90 minuuttia.
- Päivystyksessä kävijöistä 25 % oli terveyskeskuksessa kauemmin kuin 172 minuuttia.
- Alakvartiilin ala puolella on 25 % havainnoista ja yläkvartiilin yläpuolella on 25 %, joten väliin jää 50 %. Tällöin aikavälillä [90 min, 172 min] on 50 % havaintoarvoista.

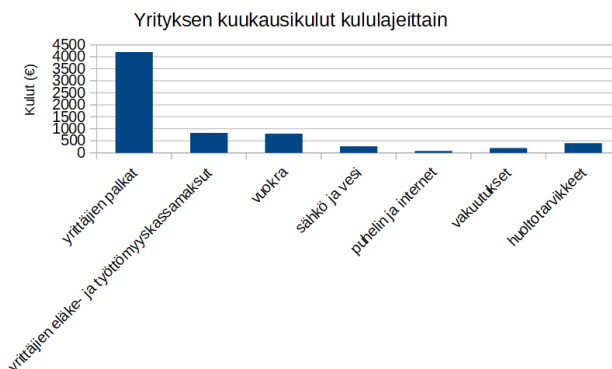
Vastaus:

- 132 min, 134 min
- 6 min–291 min, 62 min
- 50
- 90
- 172
- 50

343B. Jotta aineistoa olisi mielekästä kuvata viivakaaviolla, tulisi vaaka-akselilla olla suuruusjärjestykseen asetettuja lukuarvoja. Kululajit eivät ole sellaisia, joten yrityksen kuukausikuluja ei voi havainnollistaa viivakaaviolla.



Ympyräkaavio sopii hyvin havainnollistamaan esimerkiksi kulujen suhteellisia osuuksia. Pylväskaaviolla voi havainnollistaa euromääräisiä kuluja.



Vastaus: Ympyräkaavio ja pylväskaavio

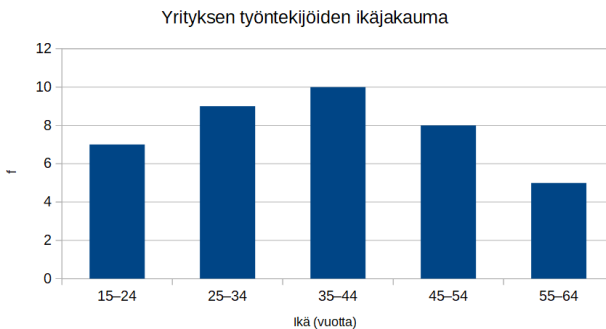
344B. Aineisto on luokiteltava viiteen tasaväliseen luokkaan. Jotta saataisiin selville luokkavälin pituus, lasketaan tilaston pienimmän arvon ja suurimman arvon erotus.

$$62 - 19 = 43$$

Jos aineisto jaettaisiin viiteen luokkaan käyttäen lukua 19 ensimmäisen luokan alarajana ja lukua 62 viimeisen luokan ylärajana, tulisi yhden luokan pituudeksi $\frac{43}{5} = 8,6$, joka ei ole kokonaisluku. Jotta taulukko olisi mahdollisimman selkeä, valitaan ensimmäisen luokan alarajaksi 15 ja luokan pituudeksi 10. Laaditaan luokitellusta aineistosta frekvenssitaulukko.

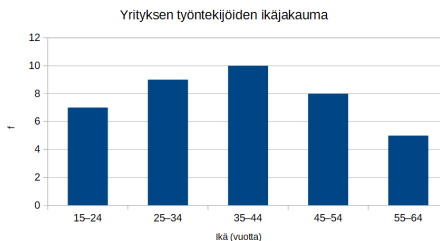
Ikä (vuotta)	f
15–24	7
25–34	9
35–44	10
45–54	8
55–64	5

Piirretään pylväskaavio.



Vastaus:

Ikä (vuotta)	f
15–24	7
25–34	9
35–44	10
45–54	8
55–64	5



345A. a) 100% vastaa täysi ympyrää, joten 15,6 % vastaavan keskuskulman suuruus on $0,156 \cdot 360^\circ = 56,16^\circ \approx 56^\circ$.

Vastaus: 56°

b) 100 % vastaa täysiympyrää. Tällöin ympyräkaavion sektorin keskuskulmaa $95,6^\circ$ vastaava suhteellinen frekvenssi on

$$\frac{95,6^\circ}{360^\circ} = 0,2655\dots \approx 26,6 \%$$

Vastaus: 26,6 %

346B. Täydennetään taulukkoon summafrekvenssit ja suhteelliset summafrekvenssit.

Puhelujen määrä	f	sf	$sf\%$
0–19	2	2	$\frac{2}{74} \approx 3 \%$
20–39	17	$2 + 17 = 19$	$\frac{19}{74} \approx 26 \%$
40–59	42	$19 + 42 = 61$	$\frac{61}{74} \approx 82 \%$
60–79	13	$61 + 13 = 74$	100 %

Luokka 20–39 on alakvartiililuokka, koska suhteellinen summafrekvenssi ylittää sen kohdalla arvon 25 %.

Luokka 40–59 on yläkvartiililuokka, koska suhteellinen summafrekvenssi ylittää sen kohdalla arvon 75 %.

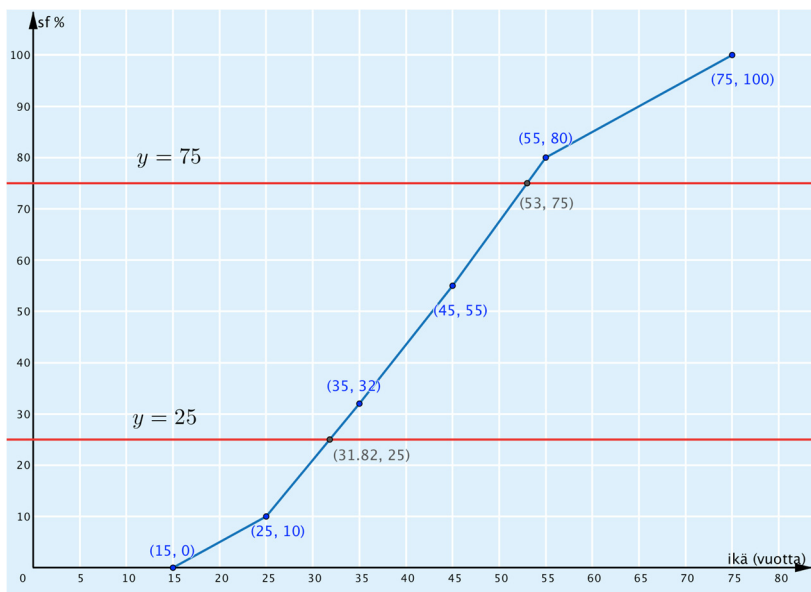
Vastaus: alakvartiililuokka 20–39, yläkvartiililuokka 40–59

347B. a) Täydennetään taulukkoon todelliset rajat ja suhteelliset summafrekvenssit. Ikä ilmoitetaan yleensä kokonaisina vuosina ja pyöristetään alaspäin. Esimerkiksi 24 vuotias lasketaan 24-vuotiaaksi, vaikka hän täyttäisi huomenna 25 vuotta. Luokan 18–24 todellinen alaraja on siis 18 vuotta ja todellinen yläraja 25 vuotta.

Ikä (vuotta)	Todellinen alaraja (vuotta)	Todellinen yläraja (vuotta)	f %	sf %
18–24	18	25	10	10
25–34	25	35	22	$10 + 22 = 32$
35–44	35	45	23	$32 + 23 = 55$
45–54	45	55	25	$55 + 25 = 80$
55–74	55	75	20	$80 + 20 = 100$

Piirretään suhteellisen summafrekvenssin kuvaaja.

Suorien $y = 25$ ja $y = 75$ ja suhteellisen summafrekvenssin kuvaajan leikkauspisteet ovat $(31,82; 25)$ ja $(53, 75)$, joten alakvartiili on noin 32 vuotta ja yläkvartiili 53 vuotta.



Vastaus: Alakvartiili n. 32 vuotta, yläkvartiili 53 vuotta.

- b) Kopioidaan taulukko taulukkolaskentaohjelmaan ja täydennetään taulukkoon summafrekvenssit ja suhteelliset summafrekvenssit.

Huoneluku	f	sf	sf %
1	274 187	274 187	9 %
2	433 415	707 602	24 %
3	633 471	1 341 073	46 %
4	619 715	1 960 788	68 %
5	517 640	2 478 428	86 %
6	282 129	2 760 557	96 %
Vähintään 7	128 817	2 889 374	100 %

Suhteellisen summafrekvenssisarakkeesta nähdään, että 25 %:n raja ylittyy huoneluvun 3 kohdalla, joten alakvartiili on 3. Yläkvartiili on 5, koska 75 %:n raja ylittyy huoneluvun 5 kohdalla.

Vastaus: alakvartiili 3, yläkvartiili 5

- 348A** Suklaiden kilohintojen keskiarvo on 31,68 €/kg ja keskihajonta 17,23 €/kg.

Suklaiden A ja B kilohinnat poikkesivat keskiarvosta 0,23 ja $-0,19$ keskihajontaa.

Suklaan A kilohinta on

$$31,68 \text{ €/kg} + 0,23 \cdot 17,23 \text{ €/kg} = 35,6429 \text{ €/kg} \approx 35,64 \text{ €/kg}.$$

Suklaan B kilohinta on

$$31,68 \text{ €/kg} - 0,19 \cdot 17,23 \text{ €/kg} = 28,4063 \text{ €/kg} \approx 28,41 \text{ €/kg}$$

Vastaus: A: 35,64 €/kg, B: 28,41 €/kg

- 349B.** Sopivalla ohjelmalla saadaan elinkustannusindeksin pisteluvun ja syöpätapausten määrän väliseksi korrelaatiokertoimeksi $r = 0,9728 \approx 0,97$ ja selitysasteeksi $r^2 = 0,9464 \approx 0,95$. Korrelaatio kertoo tilastollisesta riippuvuudesta, mutta ei syy-seuraussuhteesta, joten ei voida päätellä, että korkeat hinnat aiheuttavat syöpää.

KeskiarvoX	878.1667
KeskiarvoY	90961.3...
Sx	643.7741
Sy	38070.9...
r	0.9728
ρ	1
Sxx	455889...
VarianssiY	159433...
Sxy	262279...
R ²	0.9464
SSE	854043...

Vastaus: $r = 0,97$, $r^2 = 0,95$, ei voida päätellä

- 350B.** Käytetään prosenttiosuuksia frekvensseinä. Sopivalla ohjelmalla saadaan keskiarvoksi $\bar{x} = 3,103... \approx 3,10$ ja keskihajonnaksi $s = 1,568... \approx 1,57$.

n	100
Keskiarvo	3.1037
σ	1.5609
s	1.5687
Σx	310.37
Σx^2	1206.93
Min	0
Q1	2
Mediaani	3
Q3	4
Max	6

Vastaus: $\bar{x} = 3,10$, $s = 1,57$

351A. Täydennetään taulukkoon todelliset rajat ja luokkakeskukset. Koska viimeisen luokan ylärajaa ei ole ilmoitettu, oletetaan sen olevan 499 (€/kk).

Rahamäärä (€/kk)	Todellinen alaraja (€/kk)	Todellinen ylärajaa (€/kk)	Luokkakeskus (€/kk)	$f\%$
0–49	0	49,5	$\frac{0 + 49,5}{2} = 24,75$	72 %
50–99	49,5	99,5	$\frac{49,5 + 99,5}{2} = 74,75$	11 %
100–149	99,5	149,5	$\frac{99,5 + 149,5}{2} = 124,75$	8 %
150–199	149,5	199,5	$\frac{149,5 + 199,5}{2} = 174,5$	5 %
200–249	199,5	249,5	$\frac{199,5 + 249,5}{2} = 224,5$	3 %
250–499	249,5	499,5	$\frac{249,5 + 499,5}{2} = 374,5$	1 %

Koska vastanneiden kokonaismäärää ei ole kerrottu, merkitään sitä kirjaimella n . Tällöin luokkaan 0–49 kuuluvien määrä on $0,72n$ jne. Lasketaan liikuntaan kuluvan rahamäärän keskiarvo.

$$\begin{aligned}\bar{x} &= \frac{0,72n \cdot 24,75 + 0,11n \cdot 74,5 + \dots + 0,01n \cdot 374,5}{n} \\ &= \frac{55,18n}{n} \\ &= \frac{55,18n}{n} \\ &= 55,18 \\ &\approx 55\end{aligned}$$

Lukiolaiselta liikuntaan kuluvan rahamäärän keskiarvo on 55 €/kk.

Vastaus: 55 €/kk

352A. Muodostetaan yhtälö ja ratkaistaan siitä a .

$$\frac{a + 3a + (2a - 5) + (a + 6) + (-a)}{5} = 3,44$$

$$\frac{6a + 1}{5} = 3,44 \quad || \cdot 5$$

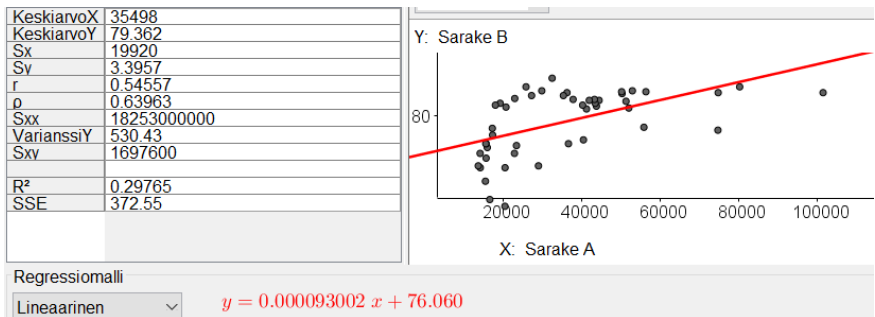
$$6a + 1 = 17,2$$

$$6a = 16,2 \quad || : 6$$

$$a = 2,7$$

Vastaus: $a = 2,7$

353B. a) Sopivalla ohjelmalla saadaan lineaariseksi malliksi
 $y = 0,00009302x + 76,060$, jossa x on bruttokansantuote henkeä kohti
dollareina ja y on eliniänodote vuosina.



Vastaus: $y = 0,00009302x + 76,060$

b) Kohdan a tunnuslukutaulukon perusteella bruttokansantuotteen ja eliniänodotteen välinen korrelaatiokerroin $r = 0,545\dots \approx 0,55$. Korrelaatio on voimakkuudeltaan kohtalainen.

Vastaus: $r = 0,55$, kohtalainen

- c) Kohdan a mallin mukaan eliniänodote on 80,710... vuotta \approx 81 vuotta, jos bruttokansantuote henkeä kohti on 50 000 dollaria.

$$y = 0.000093002 x + 76.060$$

Tarkka arvo: $x =$ $y = 80.710$

Vastaus: 81 vuotta

- d) Sopivalla ohjelmalla yhtälöstä $80,0 = 0,000093002 \dots x + 76,060 \dots$ saadaan $x = 42\,361,751 \dots$ dollaria \approx 42 000 dollaria.

Lista

lista1 = {(101450, 82.4), (80215, 83), (74735, 82.4), (74667, 78.5), (56328, 82.5), (55837, 7

Piste

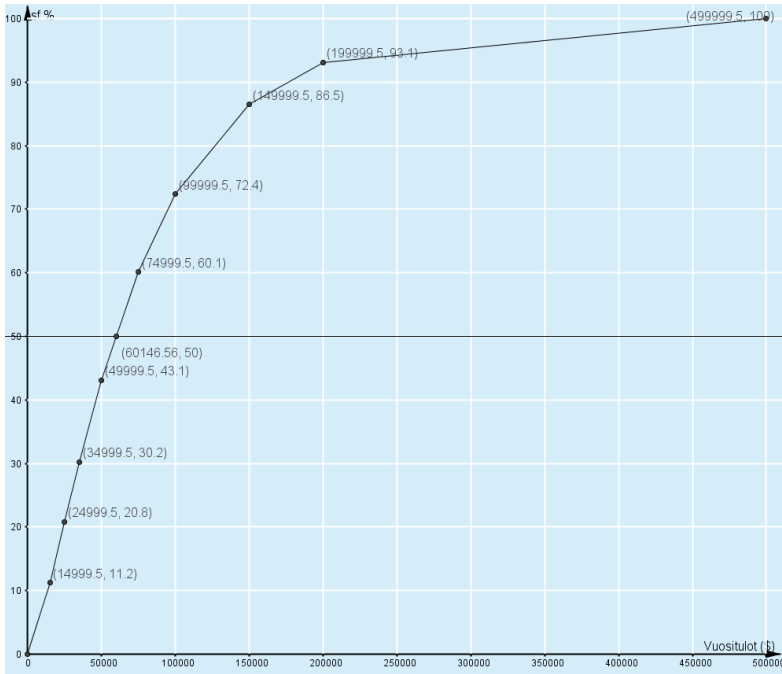
1	k:=SovitaSuora[lista1] → k := y = 0.0000930018x + 76.0602799441
2	Ratkaise[Sijoita[k, y=80],x] → $\left\{ x = \frac{118187434032487362099635}{2789956255305443049} \right\}$
3	{x = 118187434032487362099635 / 2789956255305443049} ≈ {x = 42361.75166107 }

Vastaus: 42 000 dollaria

354B. Täydennetään taulukkoon todelliset rajat, luokkakeskukset ja suhteelliset summafrekvenssit. Merkitään ylimmän tuloluokan ylärajaksi 500 000 dollaria.

Todellinen alaraja (\$)	Todellinen ylärajaja (\$)	Luokkakeskus (\$)	f %	sf %
0	14 999,5	$\frac{0 + 14999,5}{2} = 7499,75$	11,2 %	11,2 %
14 999,5	24 999,5	$\frac{14999,5 + 24999,5}{2} = 19999,5$	9,6 %	20,8 %
24 999,5	34 999,5	$\frac{24999,5 + 34999,5}{2} = 29999,5$	9,4 %	30,2 %
34 999,5	49 999,5	$\frac{34999,5 + 49999,5}{2} = 42499,5$	12,9 %	43,1 %
49 999,5	74 999,5	$\frac{49999,5 + 74999,5}{2} = 62499,5$	17 %	60,1 %
74 999,5	99 999,5	$\frac{74999,5 + 99999,5}{2} = 87499,5$	12,3 %	72,4 %
99 999,5	149 999,5	$\frac{99999,5 + 149999,5}{2} = 124999,5$	14,1 %	86,5 %
149 999,5	199 999,5	$\frac{149999,5 + 199999,5}{2} = 174999,5$	6,6 %	93,1 %
199 999,5	499 999,5	$\frac{199999,5 + 499999,5}{2} = 349999,5$	7 %	100,1 %

Piirretään suhteellisen summafrekvenssin kuvaaja. Mediaani määritetään piirtämällä suora $y = 50$. Mediaani on suoran ja suhteellisen summafrekvenssin kuvaajan leikkauspisteen x -koordinaatti $x = 60\,146,56 \approx 60\,147$. Mediaanivuositulo on noin 60 147 \$.



Määritetään vuositulojen keskiarvo ohjelmalla käyttäen frekvensseinä prosentiosuuksia $f\%$. Vuositulojen keskiarvo on 86 038,486... \$ \approx 86 038 \$.

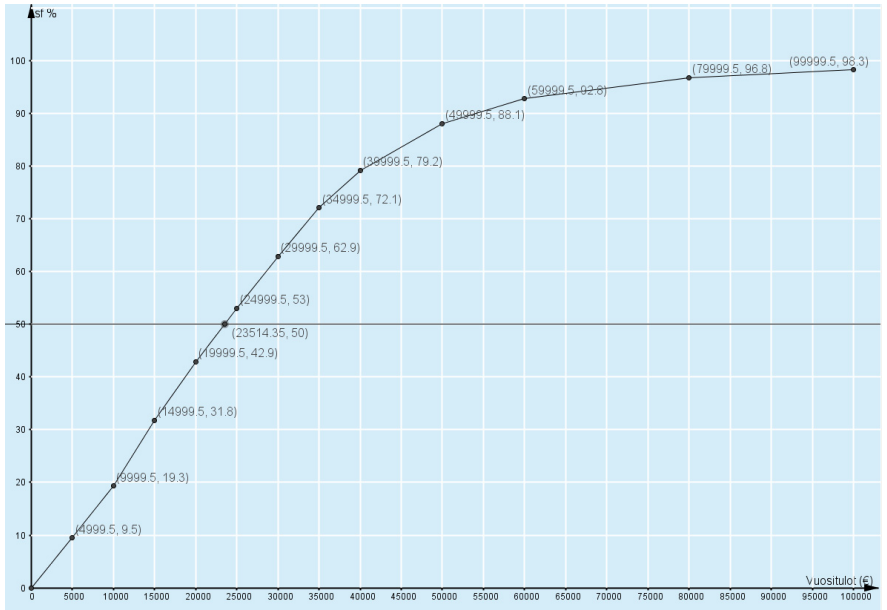
n	100.1
Keskiarvo	86038.489
σ	85732.156
s	86163.6241
Σx	8612452.75
Σx^2	147673767...
Min	7499.75
Q1	29999.5
Mediaani	62499.5
Q3	124999.5
Max	349999.5

Vastaus: Mediaanivuositulo 62 147 \$, keskimääräinen vuositulo 86 038 \$

355B. a) Täydennetään taulukkoon todelliset rajat, luokkakeskukset, summafrekvenssit ja suhteelliset summafrekvenssit. Frekvenssit ovat tuhansia. Viimeisin luokan yläraja ei ole tiedossa, mutta sitä ei tarvita mediaanin määrittämisessä.

Todellinen alaraja (€)	Todellinen yläraja (€)	Luokkakeskus (€)	f	sf	$sf\%$
0	4999,5	$\frac{0 + 4999,5}{2} = 2499,75$	442	442	$\frac{442}{4654} \approx 9,5\%$
4999,5	9999,5	$\frac{4999,5 + 9999,5}{2} = 7499,5$	457	899	$\frac{899}{4654} \approx 19,3\%$
9999,5	14 999,5	$\frac{9999,5 + 14999,5}{2} = 12 499,5$	579	1478	$\frac{1478}{4654} \approx 31,8\%$
14 999,5	19 999,5	$\frac{14999,5 + 19999,5}{2} = 17 499,5$	520	1998	$\frac{1998}{4654} \approx 42,9\%$
19 999,5	24 999,5	$\frac{19999,5 + 24999,5}{2} = 22 499,5$	468	2466	$\frac{2466}{4654} \approx 53,0\%$
24 999,5	29 999,5	$\frac{24999,5 + 29999,5}{2} = 27 499,5$	463	2929	$\frac{2929}{4654} \approx 62,9\%$
29 999,5	34 999,5	$\frac{29999,5 + 34999,5}{2} = 32 499,5$	425	3354	$\frac{3354}{4654} \approx 72,1\%$
34 999,5	39 999,5	$\frac{34999,5 + 39999,5}{2} = 37 499,5$	331	3685	$\frac{3685}{4654} \approx 79,2\%$
39 999,5	49 999,5	$\frac{39999,5 + 49999,5}{2} = 44 999,5$	414	4099	$\frac{4099}{4654} \approx 88,1\%$
49 999,5	59 999,5	$\frac{49999,5 + 59999,5}{2} = 54 999,5$	218	4317	$\frac{4317}{4654} \approx 92,8\%$
59 999,5	79 999,5	$\frac{59999,5 + 79999,5}{2} = 69 999,5$	190	4507	$\frac{4507}{4654} \approx 96,8\%$
79 999,5	99 999,5	$\frac{79999,5 + 99999,5}{2} = 89 999,5$	69	4576	$\frac{4576}{4654} \approx 98,3\%$
99 999,5			78	4654	$\frac{4654}{4654} = 100\%$

Piirretään suhteellisen summafrekvenssin kuvaaja. Mediaani määritetään piirtämällä suora $y = 50$. Mediaani on suoran ja suhteellisen summafrekvenssin kuvaajan leikkauspisteen x -koordinaatti $x = 23\,513,85 \approx 23\,514$. Mediaanivuositulo on 23 514 €.



Vastaus: 23 514 €

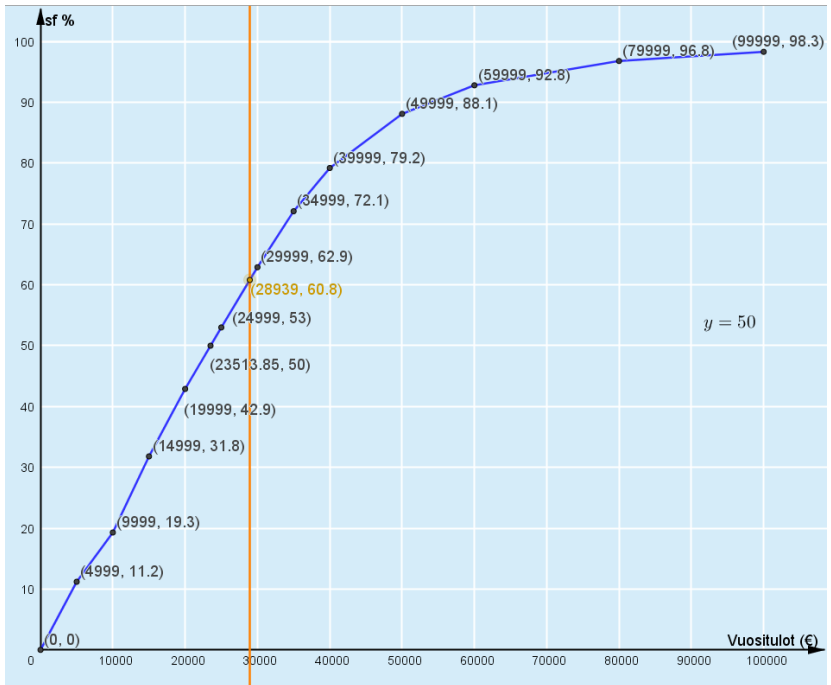
- b) Määritetään vuositulojen keskiarvo ohjelmalla käyttäen luokkakeskuksia. Viimeisen luokan keskimääräinen tulo henkeä kohden on $\frac{1419700000}{78000} = 182012,820\dots$

Käytetään viimeisen luokan edustajana arvoa 182 012,820...
 Vuositulojen keskiarvo on 28 939,046... € \approx 28 939 €.

n	4654
Keskiarvo	28939.0465
σ	27241.2598
s	27244.1869
Σx	134682322.5
Σx^2	735124732...
Min	2499.75
Q1	12499.5
Mediaani	22499.5
Q3	37499.5
Max	182012.8205

Vastaus: 28 939 €

- c) Tulojen keskiarvoa enemmän ansainneiden osuus saadaan piirtämällä suora $x = 28\,939$. Suoran ja suhteellisen summafrequenssin kuvaajan leikkauspisteen y -koordinaatti on $y = 60,8$. Tällöin keskiarvotuloja vähemmän ansaitsee $60,8\%$ ja enemmän ansaitsevia on $100\% - 60,8\% = 39,2\%$.



Vastaus: 39,2 %

- 356A. a)** Jos aineiston suurin havaintoarvo pienenee, niin keskiarvo pienenee, koska havaintoarvojen summa pienenee.

Vastaus: pienenee

- b)** Jos aineiston suurin havaintoarvo kasvaa, niin mediaani pysyy samana, koska mediaani on suuruusjärjestykseen asetetun havaintoaineiston keskimäinen arvo ja se ei muutu, vaikka viimeistä arvoa muutetaan.

Vastaus: pysyy samana

- c)** Jos aineiston pienin havaintoarvo pienenee, niin keskihajonta kasvaa, koska pienimmän luvun etäisyys keskiarvosta kasvaa.

Vastaus: kasvaa

- 357A. a)** Esimerkiksi aineiston 1, 2, 2, 4 moodi on $Mo = 2$, $Md = 2$ ja

$$\bar{x} = \frac{9}{4} = 2,25.$$

Vastaus: Esim. 1, 2, 2, 4

- b)** Esimerkiksi aineiston 0, 1, 2, 3, 3, 9 moodi on $Mo = 3$,

$$Md = \frac{2+3}{2} = 2,5 \text{ ja } \bar{x} = \frac{18}{6} = 3.$$

Vastaus: Esim. 0, 1, 2, 3, 3, 9

- c)** Esimerkiksi aineiston 0, 1, 2, 2, 3, 3, 3 moodi on $Mo = 3$, $Md = 2$ ja

$$\bar{x} = \frac{14}{7} = 2.$$

Vastaus: Esim. 0, 1, 2, 2, 3, 3, 3

358A. Uusien arvosanojen summa saadaan kertomalla keskiarvo 7,2 arvosanojen määrällä 25. Summa on $7,2 \cdot 25 = 180$. Koska opiskelijoista 7 korotti arvosanaansa yhdellä, saadaan vanhojen arvosanojen summa vähentämällä 7 uusien arvosanojen summasta. Vanhojen arvosanojen summa on $180 - 7 = 173$. Jakamalla saatu summa opiskelijoiden määrällä saadaan vanha keskiarvo.

$$\bar{x} = \frac{173}{25} = 6,92$$

Vastaus: $\bar{x} = 6,92$

359A. Merkitään arvosanan 10 määrää kirjaimella n . Tällöin arvosanan 9 määrä on $2n$ ja arvosanan 8 määrä on $3n$. Lasketaan keskiarvo.

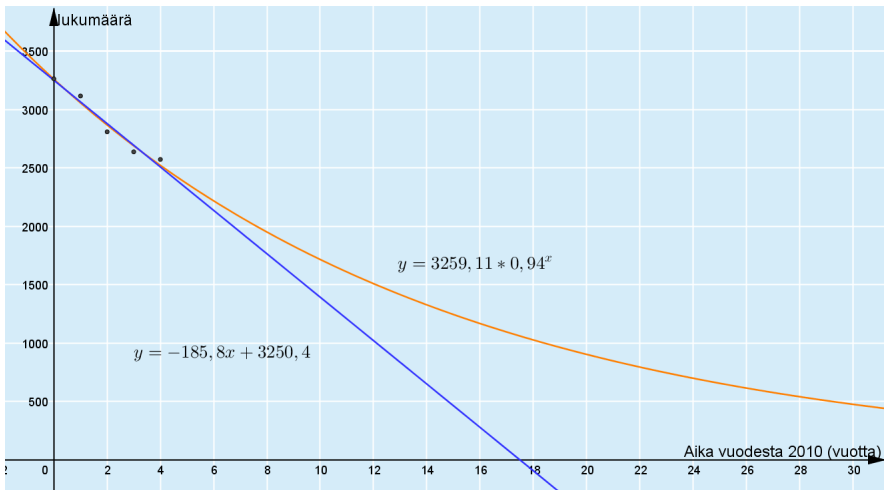
$$\begin{aligned}\bar{x} &= \frac{n \cdot 10 + 2n \cdot 9 + 3n \cdot 8}{n + 2n + 3n} \\ &= \frac{52n}{6n} \\ &= 8,666\dots \\ &\approx 8,7\end{aligned}$$

Vastaus: $\bar{x} = 8,7$

360B. a) Sopivalla ohjelmalla saadaan matkatoimistotyöntekijöiden määrän keskiarvoksi $\bar{x} = 2878,8 \approx 2879$ ja keskihajonnaksi $s = 300,0995 \approx 300$ vuosina 2010–2014.

Vastaus: $\bar{x} = 2879$, $s = 300$

b) Sopivalla ohjelmalla saadaan lineaariseksi malliksi $y = -185,8x + 3250,4$ ja eksponentiaaliseksi malliksi $y = 3259,11 \cdot 0,94^x$. Malleissa muuttuja x on aika vuodesta 2010 ja muuttuja y matkatoimistovirkailijoiden määrä.



Symbolisen laskennan toiminnolla saadaan yhtälön $-185,8x + 3250,4 = 0$ ratkaisuksi $x = 17,494\dots$. Lineaarisen mallin mukaan matkatoimistovirkailijoita ei ole enää reilun seitsemäntoista vuoden kuluttua vuodesta 2010 eli vuonna 2028.

Ratkaise($-185,8x + 3250,4 = 0$)

$$\rightarrow \left\{ x = \frac{16252}{929} \right\}$$

\$1

$$\approx \{x = 17.49\}$$

Symbolisen laskennan toiminnolla todetaan, että yhtälöllä $3259,11 \cdot 0,94^x = 0$ ei ole ratkaisua.

Ratkaise($3259,11 \cdot 0,94^x = 0$)

$$\approx \{ \}$$

Ekspontiaalisen mallin mukaan matkatoimistovirkailijoiden määrä ei ole nolla missään vaiheessa.

Vastaus: lineaarinen malli: $y = -185,8x + 3250,4$, eksponeniaalinen malli: $y = 3259,11 \cdot 0,94^x$ Lineaarisen mallin mukaan matkatoimistovirkailijoita ei enää ole Suomessa vuonna 2028, eksponeniaalisen mallin mukaan virkailijoiden on Suomessa aina.

- c) Ohjelma antaa lineaarisen mallin selitysasteeksi $0,9583 \approx 0,96$ ja eksponentiaalisen mallin selitysasteeksi $0,9665 \approx 0,97$.

KeskiarvoX	2
KeskiarvoY	2878.8
Sx	1.5811
Sy	300.0995
r	-0.9789
ρ	-1
Sxx	10
VarianssiY	360238.8
Sxy	-1858
R ²	0.9583
SSE	15022.4

KeskiarvoX	2
KeskiarvoY	2878.8
Sx	1.5811
Sy	300.0995
r	-0.9789
ρ	-1
Sxx	10
VarianssiY	360238.8
Sxy	-1858
R ²	0.9665
SSE	12065.1506

Vastaus: lineaarinen malli: 0,96, eksponentiaalinen malli: 0,97

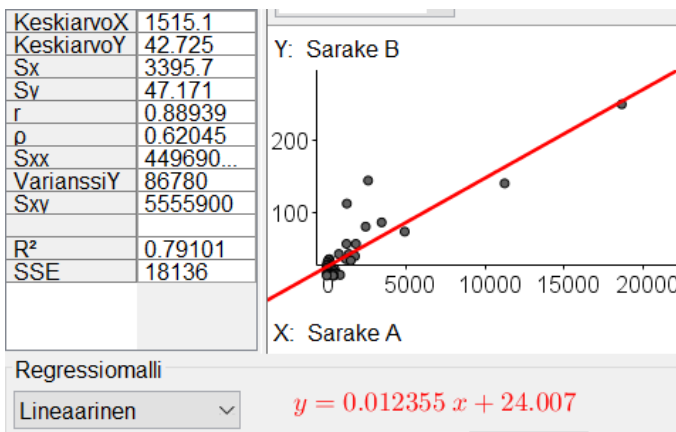
SYVENNÄ YMMÄRRYSTÄ

361B. a) Muodostetaan taulukko mitalipisteluvuista ja bruttokanstantuotteesta (GDP). Lasketaan jokainen kultamitali kolmeksi pisteeksi, hopeamitali kahdeksi ja pronssi yhdeksi pisteeksi.

	Mitalipisteluku	GDP
USA	250	18624
Britannia	144	2629
Kiina	140	11232
Venäjä	112	1283
Saksa	86	3479
Ranska	80	2466
Japani	73	4937
Italia	56	1850
Australia	56	1262
Etelä-Korea	42	1411
Hollanti	42	778
Brasilia	39	1799
Espanja	35	1233
Uusi-Seelanti	35	182
Unkari	34	126
Kanada	33	1530
Kenia	31	69
Kazakstan	28	134
Azerbaidzhan	27	38
Jamaika	26	14
Tanska	25	307
Kroatia	23	50
Kuuba	23	87
Uzbekistan	23	67
Ruotsi	21	511
Etelä-Afrikka	20	350
Ukraina	20	93

Puola	18	468
Kolumbia	16	282
Serbia	16	38
Sveitsi	15	669
Iran	15	377
Valko-Venäjä	15	49
Pohjois-Korea	14	40
Tshekki	14	193
Kreikka	13	194
Turkki	13	863
Belgia	12	467
Thaimaa	12	407
Georgia	12	14

Sopivalla ohjelmalla saadaan bruttokansantuotteen ja mitalipisteluvun väliseksi korrelaatiokertoimeksi $r = 0,889... \approx 0,89$ ja selitysasteeksi $r^2 = 0,791 \approx 0,79$ sekä regressiosuoran yhtälöksi $y = 0,012355x + 24,007$.



Vastaus: $r \approx 0,89$, $r^2 \approx 0,79$, $y = 0,012355x + 24,007$

- b) Tilastollinen riippuvuus on voimakasta. Mahdollisia selityksenä voi olla se, että bruttokansantuotteen kasvaessa on enemmän varaa panostaa valmennukseen yms. Selityksenä voi olla myös se, että mitä suurempi väestö on, sitä suurempi on maan bruttokansantuote. On todennäköisempää, että suuren väestön joukossa on useampia urheilulahjakkuuksia kuin pienemmän väestön joukossa.

Vastaus: –

- c) Sijoittamalla $x = 237$ regressiosuoran yhtälöön $y = 0,012355x + 24,007$ saadaan Suomen mitalipistemääräksi $26,935 \approx 27$. Tämä on paljon suurempi kuin Suomen saavuttama yhden pronssimitalin pistemäärä 1.

$$y = 0.012355x + 24.007$$

Tarkka arvo: $x =$ $y = 26.935$

Suomen tulos Rion vuoden 2016 olympialaisissa ei ollut a-kohdan mallin mukainen.

Vastaus: ei ole

- 362B. a)** Määritetään taulukkolaskentaohjelmalla kysytyt tunnusluvut. Havaintoarvon etäisyys keskiarvosta kannattaa laskea itseisarvofunktion avulla, jotta havaintoarvon ja keskiarvon erotus on aina suurempi tai yhtä suuri kuin nolla.

C2 fx Σ = =ITSEISARVO(B2-\$B\$12)

	A	B	C	D
1		x	Etäisyys keskiarvosta	
2		1	4,5	
3		2	3,5	
4		3	2,5	
5		4	1,5	
6		5	0,5	
7		6	0,5	
8		7	1,5	
9		8	2,5	
10		9	3,5	
11		10	4,5	
12	Keskiarvo	5,5	2,5	
13	Keskihajonta	3,027650354		

Keskiarvo on 2,5 ja se on jonkin verran pienempi kuin keskihajonta $s = 3,027\dots$

Vastaus: 2,5, pienempi kuin keskihajonta

- b)** Keskiarvo on 5 ja se on jonkin verran pienempi kuin keskihajonta $s = 5,270\dots$

C12 fx Σ = =KESKIARVO(C2:C11)

	A	B	C	D
1		x	Etäisyys keskiarvosta	
2		0	5	
3		0	5	
4		0	5	
5		0	5	
6		0	5	
7		10	5	
8		10	5	
9		10	5	
10		10	5	
11		10	5	
12	Keskiarvo	5	5	
13	Keskihajonta	5,270462767		
14				

Vastaus: 5, pienempi kuin keskihajonta

363A. Olkoon luvut 1, 2 ja 3. Näiden aritmeettinen keskiarvo on

$$\bar{x} = \frac{1+2+3}{3} = 2 \text{ ja geometrinen keskiarvo on } \sqrt[3]{1 \cdot 2 \cdot 3} = 1,817\dots$$

Aritmeettinen keskiarvo on suurempi, mutta esimerkiksi luvuilla 2, 2 ja 2.

Aritmeettinen keskiarvo on $\bar{x} = \frac{2+2+2}{3} = 2$ ja geometrinen keskiarvo

on $2\sqrt[3]{2 \cdot 2 \cdot 2} = 2$, joten väite ei pidä yleisesti paikkansa. Tällöin ei yhden tapauksen avulla voida päätellä yleistä tapausta.

Vastaus: aritmeettinen, ei voida päätellä

364B. a) Tilaston mukaan vähintään sadan vuoden ikäisiä on $701 + 2045 = 2746$ sataa tuhatta syntynyttä kohti. Tämän mukaan

$$\frac{2746}{200\,000} = 0,01373 \approx 1,4\% \text{ vastasyntyneistä elää vähintään}$$

100-vuotiaaksi.

Vastaus: 1,4 %

b) 20-vuotiaita miehiä on 99 477 ja näistä elää vähintään 90-vuotiaaksi 18 465, joten 20-vuotiaista miehistä elää vähintään 90-vuotiaaksi

$$\frac{18465}{99477} = 0,185\dots \approx 19\%.$$

Vastaus: 19 %

c) 20-vuotiaita naisia on 99 613. 60-vuotiaita naisia on 95 414 ja 90-vuotiaita naisia 34 840. Naisia, jotka elävät 60-vuotiaiksi, mutta eivät 90-vuotiaiksi on $95\,414 - 34\,840 = 60\,574$. Tällöin 20-vuotiaista naisista elää 60-vuotiaaksi, mutta ei 90-vuotiaaksi

$$\frac{60\,574}{99\,613} = 0,609\dots \approx 61\%.$$

Vastaus: 61 %

365A. Absoluuttinen virhe on $53 \text{ g} - 52,69 \text{ g} = 0,31 \text{ g}$ ja suhteellinen virhe $\frac{0,31 \text{ g}}{52,69 \text{ g}} = 0,00588\dots \approx 0,006 = 0,6 \%$.

Vastaus: absoluuttinen virhe 0,31 g, suhteellinen virhe 0,6 %

366B. a) Ylemmästä taulukosta voidaan lukea, että 25 % vastaajista oli iältään 15–19-vuotiaita.

	Ikäluokka (vuotta)				
Ryhmä	10–14	15–19	20–24	25–29	yhteensä %
Pelaaja	27	29	26	18	100
Tiedonhakija	30	20	21	29	100
Kommentoija	13	27	31	29	100
yhteensä %	23	25	26	26	100

Vastaus: 25 %

b) Alemmasta taulukosta voidaan lukea, että Pelaaja-ryhmäläisistä 27 % oli iältään 15–19-vuotiaita.

	Ikäluokka (vuotta)				
Ryhmä	10–14	15–19	20–24	25–29	yhteensä %
Pelaaja	27	27	23	16	23
Tiedonhakija	52	33	33	44	40
Kommentoija	21	40	44	40	37
yhteensä %	100	100	100	100	100

Vastaus: 27 %

- c) Alemmasta taulukosta voidaan lukea, että Tiedonhakija-ryhmään kuului 40 % vastaajista.

Ryhmä	Ikäluokka (vuotta)				yhteensä %
	10–14	15–19	20–24	25–29	
Pelaaja	27	27	23	16	23
Tiedonhakija	52	33	33	44	40
Kommentoija	21	40	44	40	37
yhteensä %	100	100	100	100	100

Vastaus: 40 %

- 367B. a) Kaikkien aineistojen keskiarvot ja keskihajonnat ovat kahden desimaalin tarkkuudella yhtä suuret.

Muuttuja x : keskiarvo 9,00 ja keskihajonta 3,32.

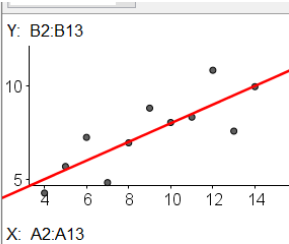
Muuttuja y : keskiarvo 7,50 ja keskihajonta 2,03.

Vastaus: Muuttuja x : $\bar{x} = 9,00$ ja $s = 3,32$, muuttuja y : $\bar{x} = 7,50$ ja $s = 2,03$, kaikkien aineistojen keskiarvot ja keskihajonnat ovat kahden desimaalin tarkkuudella yhtä suuret

- b) Piirretään aineistojen hajontakuviot ja regressiosuorat sekä määritetään kullekin aineistolle korrelaatiokerroin muuttujien x ja y väliselle riippuvuudelle.

Aineisto I

KeskiarvoX	9.0000
KeskiarvoY	7.5009
Sx	3.3166
Sy	2.0316
r	0.81642
ρ	0.81818
Sxx	110.00
VarianssiY	41.273
Sxy	55.010
R ²	0.66654
SSE	13.763

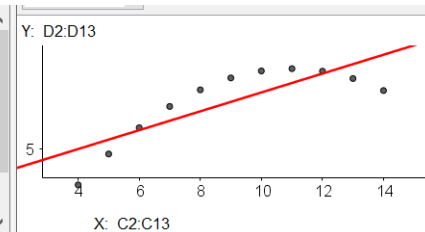


Regressiomalli

Lineaarinen $y = 0.50009x + 3.0001$

Aineisto II

KeskiarvoX	9.0000
KeskiarvoY	7.5009
Sx	3.3166
Sy	2.0317
r	0.81624
ρ	0.69091
Sxx	110.00
VarianssiY	41.276
Sxy	55.000

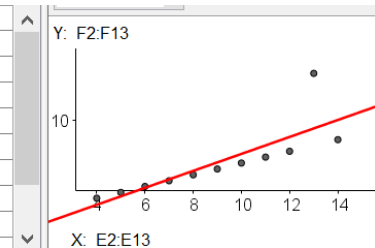


Regressiomalli

Lineaarinen $y = 0.50000x + 3.0009$

Aineisto III

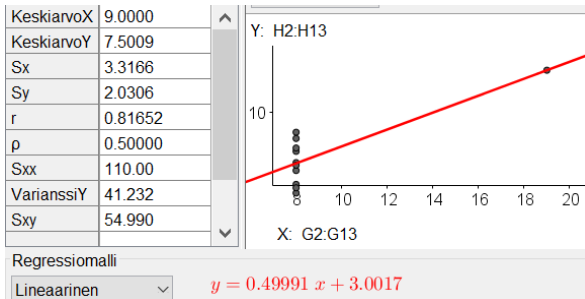
KeskiarvoX	9.0000
KeskiarvoY	7.5000
Sx	3.3166
Sy	2.0304
r	0.81629
ρ	0.99091
Sxx	110.00
VarianssiY	41.226
Sxy	54.970



Regressiomalli

Lineaarinen $y = 0.49973x + 3.0025$

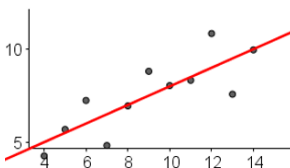
Aineisto IV



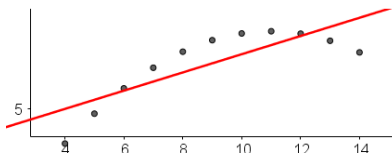
Huomataan, että vaikka kaikkien aineistojen hajontakuviot ovat erilaiset ja vain aineistojen I ja III muuttujat ovat selkeästi lineaarisesti riippuvia, on kaikissa korrelaatiokertoimen kahden desimaalin likiarvo yhtä suuri $r = 0,82$. Kaikkien aineistojen regressiosuoraksi saadaan kahden desimaalin tarkkuudella sama suora $y = 0,50x + 3,00$.

Vastaus: Hajontakuviot ovat erilaiset, mutta kaikkien aineistojen korrelaatiokerroin on $r = 0,82$ ja regressiosuoran yhtälö $y = 0,50x + 3,00$.

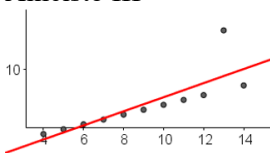
Aineisto I



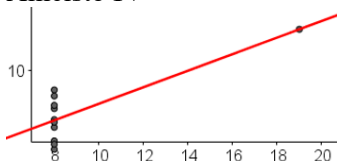
Aineisto II



Aineisto III



Aineisto IV



- c) Pelkkä korrelaatiokerroin voi antaa harhaanjohtavan kuvan muuttujien riippuvuudesta ja ilmiön luonteesta. Aineistoissa II ja IV hajontakuvio paljastaa selkeästi, että muuttujien välinen riippuvuus ei ole lineaarista.

Vastaus: Pelkkä korrelaatiokerroin voi antaa harhaanjohtavan kuvan muuttujien riippuvuudesta ja ilmiön luonteesta. Aineistoissa II ja IV hajontakuvio paljastaa selkeästi, että muuttujien välinen riippuvuus ei ole lineaarista.

368B. Valtion A tulojen keskiarvo on $\bar{x} = 47\,770$ ja keskihajonta $s = 40\,228,251\dots$

n	49
Keskiarvo	47770
σ	39815.6433
s	40228.2517
Σx	2340730
Σx^2	1894956594...
Min	0
Q1	25713
Mediaani	43649
Q3	60736.5
Max	261003

Valtion A variaatiokerroin on $\frac{40\,228,251\dots}{47\,770} = 0,842\dots$

Valtion B tulojen keskiarvo on $\bar{x} = 34\,996,612\dots$ ja keskihajonta $s = 36\,299,310\dots$

n	49
Keskiarvo	34996.6122
σ	35926.9999
s	36299.3103
Σx	1714834
Σx^2	1232600972...
Min	0
Q1	14083.5
Mediaani	29132
Q3	45012.5
Max	231921

Valtion B variaatiokerroin on $\frac{36\,299,310\dots}{34\,996,612\dots} = 1,037\dots$

Valtion A variaatiokerroin on pienempi kuin valtion B, joten valtiossa A tulot jakautuvat otoksen perusteella suhteellisesti tasaisemmin.

Vastaus: valtiossa A

- 369B. a)** Koska otoskeskihajonnan jakajassa on pienempi luku, on otoskeskihajonta suurempi kuin keskihajonta.

Vastaus: Otoskeskihajonta on suurempi kuin keskihajonta.

- b)** Kymmenen päivän ajanjaksolla keskihajonta on $57,099\dots \approx 57,1$ ja otoskeskihajonta $60,188\dots \approx 60,2$.

n	10
Keskisarvo	620.8
σ	57.0996
s	60.1882
Σx	6208
Σx^2	3886530
Min	530
Q1	563
Mediaani	651.5
Q3	671
Max	673

Kolmenkymmenen päivän ajanjaksolla keskihajonta on $104,813\dots \approx 104,8$ ja otoskeskihajonta $106,605\dots \approx 106,6$.

n	30
Keskisarvo	633.9333
σ	104.8138
s	106.6056
Σx	19018
Σx^2	12385722
Min	399
Q1	599
Mediaani	671
Q3	719
Max	737

Yhdeksänkymmenen päivän ajanjaksolla keskihajonta on $106,494\dots \approx 106,5$ ja otoskeskihajonta $107,090\dots \approx 107,1$.

n	90
Keskisarvo	597.2444
σ	106.4944
s	107.091
Σx	53752
Σx^2	33123778
Min	360
Q1	529
Mediaani	621
Q3	680
Max	737

Huomataan, että hajontaluvut lähestyvät toisiaan otoskoon kasvaessa.

Vastaus: Hajontaluvut lähestyvät toisiaan otoskoon kasvaessa.

- c) Koska otoskeskihajonta on suurempi kuin keskihajonta, on otoskeskihajonnan avulla laskettu väli pidempi kuin keskihajonnalla laskettu väli.

Pienten otosten perusteella tehdyt johtopäätökset ovat epävarmempia kuin suuremman otoksen perusteella tehdyt johtopäätökset. Tällöin pienillä otoksilla on järkevää käyttää jakajana lukua $n - 1$, koska silloin saadaan suurempi varmuusväli.

Vastaus: Otoskeskihajonnan avulla laskettu väli pidempi kuin keskihajonnalla laskettu väli. Saadaan suurempi varmuusväli.

- d) Jos n on suuri, on $n - 1$ melkein yhtä suuri, jolloin keskihajonta ja otoskeskihajonta ovat melkein yhtä suuret.

Vastaus: Jos n on suuri, on $n - 1$ melkein yhtä suuri.