

5 TILASTOJEN ANALYSOINTIA

POHDITTAVAA

1. Lasketaan jogurttipurkkien massojen keskiarvo taulukkolaskentaohjelmalla.

A102						
	A	B	C	D	E	F
82	143					
83	148					
84	143					
85	153					
86	152					
87	148					
88	145					
89	149					
90	157					
91	155					
92	150					
93	147					
94	155					
95	153					
96	154					
97	148					
98	145					
99	146					
100	142					
101	147					
102	149,6					

Ohjelmalla saadaan 100 jogurttipurkin sisällön keskiarvopainoksi 149,6 g.

Vastaus: 149,6 g

2. Lasketaan jogurttipurkkien massojen keskiarvo taulukkolaskentaohjelmalla.

A13						
	A	B	C	D	E	F
1	Massa (g)					
2	148					
3	148					
4	154					
5	145					
6	153					
7	156					
8	148					
9	152					
10	151					
11	155					
12	147					
13	150,6					
14						

Otoksen keskiarvopainoksi saadaan 150,6 g, joka on melko hyvä arvio kaikkien purkkien keskiarvopainoksi.

Vastaus: 150,6 g, melko hyvän arvion

5.1 Hajontalukuja

ALOITA PERUSTEISTA

- 501.** a) Suurin aallonkorkeus on noin 8,2 m.

Vastaus: 8,2 m

- b) Pienin aallonkorkeus on noin 1,4 m.

Vastaus: 1,4 m

- c) Suurin arvo on noin 8,2 m ja pienin arvo noin 1,4 m, joten vaihteluväli on metreinä [1,4; 8,2].

Vastaus: [1,4; 8,2]

- d) Vaihteluvälin pituus on $8,2 \text{ m} - 1,4 \text{ m} = 6,8 \text{ m}$

Vastaus: 6,8 m

- 502.** Jakaumassa B havaintoarvot ovat keskittyneet voimakkaasti keskiarvon ympärille, joten keskihajonta on siinä pienin.

Jakaumassa A suurin osa havaintoarvoista on hyvin kaukana keskiarvosta, joten keskihajonta on siinä suurin.

Vastaus: B, C ja A

- 503.** Jakaumassa C kaikki arvosanat ovat vähintään 8. Kymppejä on enemmän kuin kaseja, joten keskiarvo on yli 9. Jakaumissa A ja B on myös arvosanoja 4–8, joten keskiarvo on selvästi alempi kuin jakaumassa C. Jakaumassa A on eniten äärimmäisiä arvosanoja (kymppejä ja nelosia), joten keskihajonta on suurin.

Vastaus: suurin keskiarvo C, suurin keskihajonta A

504. Lasketaan ensin kahvintuotanto kilogrammoina.

Vuosi	2013	2014	2015	2016
Tuotanto (60 kg:n säkkeinä)	56 925	55 050	51 116	48 423
Tuotanto (kg)	3415500	3303000	3066960	2905380

Tuotannon keskiarvo kilogrammoina

$$\bar{x} = \frac{3\,415\,500 + 3\,303\,000 + 3\,066\,960 + 2\,905\,380}{4} = 3\,172\,710 \approx 3\,170\,000.$$

Keskiarvo kilogrammoina on noin 3 170 000 kg.

Keskihajonta kilogrammoina

$$s = \sqrt{\frac{(3\,415\,500 - \bar{x})^2 + (3\,303\,000 - \bar{x})^2 + (3\,066\,960 - \bar{x})^2 + (2\,905\,380 - \bar{x})^2}{4 - 1}} \\ = 229906,400\dots$$

Keskihajonta kilogrammoina on 229 906,400... kg $\approx$ 230 000 kg.

Vastaus: $\bar{x} = 3\,170\,000$ kg, $s = 230\,000$ kg

- 505.** Kopioidaan taulukko sopivaan ohjelmaan. Määritetään ohjelman avulla patenttien vuosittaisen määrän keskiarvo, vaihteluväli ja keskihajonta.

	A	B
1	Vuosi	Patenttien määrä
2	2000	1937
3	2001	2046
4	2002	2052
5	2003	2400
6	2004	2075
7	2005	1757
8	2006	1059
9	2007	920
10	2008	998
11	2009	1055
12	2010	923
13	2011	846
14	2012	835
15	2013	711
16	2014	786
17	2015	931
18	keskiarvo	=KESKIARVO(B2:B17)
19	keskihajonta	=KESKIHAJONTA(B2:B17)
20	suurin arvo	=MAKS(B2:B17)
21	pienin arvo	=MIN(B2:B17)

Ohjelman tilastotoiminnoilla saadaan patenttien määrän vuosittaiseksi keskiarvoksi noin 1333 ja keskihajonnaksi noin 589. Ohjelman toiminnolla tai taulukosta saadaan suurimmaksi arvoksi 2400 ja pienimmäksi 711, joten vaihteluväli on [711, 2400].

Vastaus: $\bar{x} \approx 1333$, vaihteluväli [711, 2400], $s \approx 589$

- 506. a)** Ensimmäisen luokan todelliset rajat ovat 1 ja 10, koska teoriakokeessa saa vain kokonaislukupisteitä. Ensimmäisen luokan luokkakeskus on $\frac{1+10}{2} = 5,5$.

Muut luokkakeskukset ovat vastaavasti 15,5; 25,5; 35,5; 45,5 ja 55,5.

Kokeen suorittaneita on $3 + 5 + 8 + 10 + 34 + 240 = 300$.

Pisteiden keskiarvo on

$$\begin{aligned}\bar{x} &= \frac{3 \cdot 5,5 + 5 \cdot 15,5 + 8 \cdot 25,5 + 10 \cdot 31,5 + 34 \cdot 45,5 + 240 \cdot 55,5}{300} \\ &= 51,733... \\ &\approx 51,7\end{aligned}$$

Vastaus: $\bar{x} \approx 51,7$ pistettä

- b)** Keskihajonta on

$$\begin{aligned}s &= \sqrt{\frac{3 \cdot (5,5 - \bar{x})^2 + 5 \cdot (15,5 - \bar{x})^2 + \dots + 240 \cdot (55,5 - \bar{x})^2}{300 - 1}} \\ &= 9,296... \\ &\approx 9,3\end{aligned}$$

Vastaus: $s \approx 9,3$ pistettä

VAHVISTA OSAAMISTA

- 507.** Kopioidaan tiedot taulukkolaskentaohjelmaan ja määritetään keskihajonta ohjelman avulla.

Ajat ovat sekunteina 65, 60, 57, 62, 67, 64, 66, 63, 55, 68, 66, 66 ja 65.

	A	B
1	Aika (s)	
2	65	
3	60	
4	57	
5	62	
6	67	
7	64	
8	66	
9	63	
10	55	
11	68	
12	66	
13	66	
14	65	
15		
16	keskihajonta	=KESKIHAJONTA(A2:A14)
17	keskiarvo	=KESKIARVO(A2:A14)

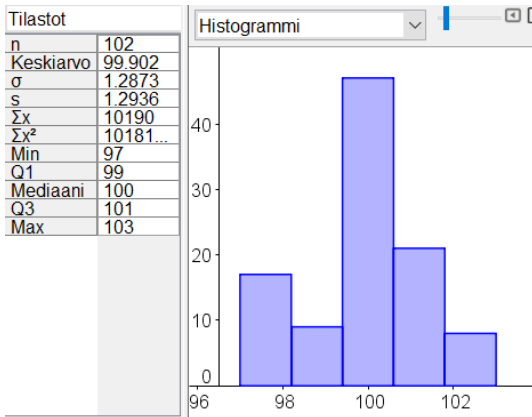
Keskihajonnaksi saadaan sopivalla ohjelmalla $3,927\dots s \approx 3,9 s$.

Kauden paras tulos on 55 s. Jotta se poikkeaisi merkittävästi muista tuloksista, on sen oltava yli kahden keskihajonnan päässä keskiarvosta.

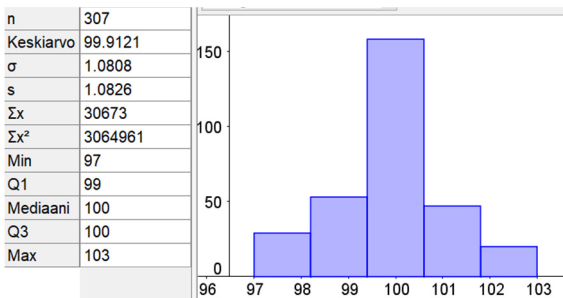
Lasketaan aikojen keskiarvo. Keskiarvoksi saadaan sopivalla ohjelmalla $63,384\dots s$. Merkittävän poikkeaman alaraja on $63,384\dots s - 2 \cdot 3,927\dots s = 55,530\dots s$. Kauden paras tulos 55 s alittaa alarajan, joten se poikkeaa merkittävästi muista tuloksista.

Vastaus: $s \approx 3,9 s$, poikkeaa

- 508.** Kopioidaan taulukko sopivaan ohjelmaan. Määritetään ohjelman avulla koneen A pussien massojen tunnusluvut.



Määritetään koneen B tunnusluvut.



- a) Kone pakkaa sitä tarkemmin, mitä lähempänä pakkausten massojen keskiarvo on tavoiteltavaa 100 g:n massaa ja mitä pienempi keskihajonta pakkausten massoilla on.

kone A: keskiarvo $\bar{x} \approx 99,9$ g ja keskihajonta $s \approx 1,29$ g

kone B: keskiarvo $\bar{x} \approx 99,9$ g ja keskihajonta $s \approx 1,08$ g

Kummankin koneen pakkausten massojen keskiarvo on 99,9 g. Koska koneen B pakkausten massojen keskihajonta on pienempi kuin koneen A, voidaan sanoa, että kone B pakkaa tarkemmin.

Vastaus: kone B

- b) Lasketaan merkittävän poikkeaman rajat lisäämällä ja vähentämällä keskiarvoista kaksinkertainen keskihajonta.

kone A: yläraja $99,9 + 2 \cdot 1,29 = 102,48$

 alaraja $99,9 - 2 \cdot 1,29 = 97,32$

kone B: yläraja $99,9 + 2 \cdot 1,08 = 102,6$

 alaraja $249,9 - 2 \cdot 1,06 = 97,78$

Koska molemmilla koneilla pakkauksen massa 103 g on rajojen ulkopuolella, voidaan sanoa, että molempien koneiden pakkaamat 103 g:n pakkaukset poikkeavat merkittävästi tavoitepainosta.

Vastaavasti myös koneiden pakkaamat kevyimmät 97 g:n pakkaukset poikkeavat merkittävästi koneen tavoitepainosta.

Vastaus: poikkeavat

- 509. a)** Kopioidaan taulukko sopivaan ohjelmaan. Määritetään tilavuuden keskiarvo ja keskihajonta taulukkolaskentaohjelman avulla.

Keskiarvo on

$$\bar{x} = 87,975 \text{ cl} \approx 88 \text{ cl.}$$

Keskihajonta on

$$s = 4,181... \text{ cl} \approx 4,2 \text{ cl.}$$

Vastaus: $\bar{x} \approx 88 \text{ cl}$, $s \approx 4,2 \text{ cl}$

	A	B
1	Tilavuus (cl)	
2	87,1	
3	88	
4	87,7	
5	90,7	
6	84,7	
7	85,8	
8	79,2	
9	80,3	
10	86,2	
11	88,2	
12	92,8	
13	89,1	
14	93,2	
15	94,7	
16	95,7	
17	88,3	
18	87,3	
19	86,5	
20	85	
21	89	
22		
23	keskiarvo	=KESKIARVO(A2:A21)
24	keskihajonta	=KESKIAJONTA(A2:A21)

- b)** Suurin hyväksyttävä pullon tilavuus on $87,975 \text{ cl} + 4,181... \text{ cl} = 92,156... \text{ cl}$ ja pienin hyväksyttävä tilavuus on $87,975 \text{ cl} - 4,181 \text{ cl} = 83,793... \text{ cl}$.

Merkitään taulukkoon liian pienet pullot sinisellä ja liian suuret pullot keltaisella.

Hylättäviä pulloja on 6 kaikista otoksen

20 pullosta. Tämä on $\frac{6}{20} = 30\%$ pulloista.

Vastaus: 30 %

	A
1	Tilavuus (cl)
2	87,1
3	88
4	87,7
5	90,7
6	84,7
7	85,8
8	79,2
9	80,3
10	86,2
11	88,2
12	92,8
13	89,1
14	93,2
15	94,7
16	95,7
17	88,3
18	87,3
19	86,5
20	85
21	89

510. a) Kopioidaan luvut sopivaan ohjelmaan. Määritetään keskiarvo ja keskihajonta ohjelman avulla.

	A	B
1	Jokerit	9173
2	HIFK	8266
3	Kärpät	5821
4	TPS	5534
5	Tappara	5359
6	Ilves	5177
7		
8	keskiarvo	=KESKIARVO(B1:B6)
9	keskihajonta	=KESKIHAJONTA(B1:B6)

Keskiarvo on 6555.

Keskihajonta on $1714,172... \approx 1714$.

Vastaus: $\bar{x} = 6555$, $s \approx 1714$

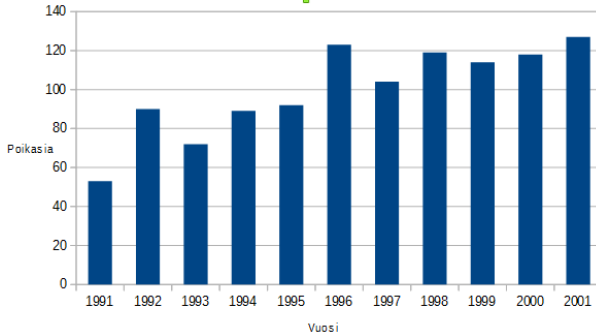
- b) Poikkeama keskiarvosta ylöspäin saadaan keskiarvon ja keskihajonnan summan avulla
 $6555 + 1714,172... = 8269,172...$ Vain Jokerien katsojaluku on tätä suurempi.

Poikkeama keskiarvosta alaspäin saadaan keskiarvon ja keskihajonnan erotuksen avulla
 $6555 - 1714,172... = 4840,828...$ Yhdenkään joukkueen katsojaluvut eivät ole tätä pienempiä.

Vastaus: Jokerit

511. Kopioidaan taulukko sopivaan ohjelmaan.

Piirretään pylväsdiagrammi.



Määritetään poikasten määrän keskiarvo ja keskihajonta ohjelman avulla.

Keskiarvo on $\bar{x} = 99$ poikasta.

Keskihajonta on $s = 22,542... \approx 23$.

Poikasmäärä poikkeaa merkitsevästi ylöspäin keskiarvosta, jos se on vähintään $99 + 2 \cdot 22,542... = 144,085...$
Vuoden 2001 poikasmäärä ei ole näin suuri, joten se ei poikkeaa merkitsevästi keskiarvosta.

Poikasmäärä poikkeaa merkitsevästi alaspäin keskiarvosta, jos se on korkeintaan $99 - 2 \cdot 22,542... = 53,914...$
Vuoden 1991 poikasmäärä alittaa tämän, joten poikkeama on merkittävä.

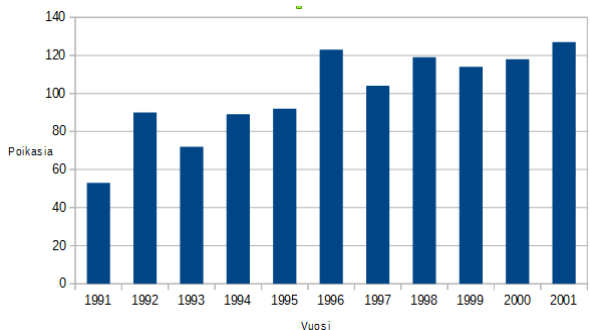
	A	B
1	Vuosi	
2	1990	87
3	1991	53
4	1992	90
5	1993	72
6	1994	89
7	1995	92
8	1996	123
9	1997	104
10	1998	119
11	1999	114
12	2000	118
13	2001	127
14		
15	keskiarvo	=KESKIARVO(B2:B13)
16	keskihajonta	=KESKHAJONTA(B2:B13)

Vastaus:

$\bar{x} = 99$, $s \approx 23$,

vuoden 1991 poikkeaa,

vuoden 2001 ei poikkeaa



- 512. a)** Kopioidaan taulukko sopivaan ohjelmaan. Lasketaan hinta yhtä matkaa kohti jakamalla kokonaiskustannukset matkustajien määrällä.
 Lasketaan matkakohtainen hinta ensin linjan 70 bussille.

	A	B	C	D
1	linja	matkustajia	Kokonaiskust.	Kust./matka (€)
2	70	13135	102647	=C2/B2
3	91	19378	275762	
4	96H	8372	39438	
5	243	92611	528542	
6	432	10743	69694	
7	433	49382	244122	
8	443	26304	234583	
9	444	10012	57254	
10	446	9104	57572	
11	625	5544	43854	

Kopioidaan sitten hinnan laskulauseke alemmille riveille.

Pyöristetään luvut kahden desimaalin tarkkuuteen.

	A	B	C	D
1	linja	matkustajia	Kokonaiskust.	Kust./matka (€)
2	70	13135	102647	7,81
3	91	19378	275762	14,23
4	96H	8372	39438	4,71
5	243	92611	528542	5,71
6	432	10743	69694	6,49
7	433	49382	244122	4,94
8	443	26304	234583	8,92
9	444	10012	57254	5,72
10	446	9104	57572	6,32
11	625	5544	43854	7,91

Vastaus:

Linja	Kustannukset/matka (€)
70	7,81
91	14,23
96H	4,71
243	5,71
432	6,49
433	4,94
443	8,92
444	5,72
446	6,32
625	7,91

- b) Lisätään a-kohdan taulukkoon matkustajamäärät ja määritetään vaihteluväli, keskiarvo ja keskihajonta. Matkustajamäärät ovat frekvenssejä ja keskimääräinen kustannus matkustajaa kohden muuttujan arvo.

Yhden matkan hinta	Matkustajia
7.81	13135
14.23	19378
4.71	8372
5.71	92611
6.49	10743
4.94	49382
8.92	26304
5.72	10012
6.32	9104
7.91	5544

n	244585
Keskiarvo	6.7606
σ	2.5093
s	2.5093
Σx	1653534.01
Σx^2	12718913.9519
Min	4.71
Q1	5.71
Mediaani	5.71
Q3	7.81
Max	14.23

Vaihteluväli on euroina $[4,71; 14,23]$.

Keskiarvo on $\bar{x} = 6,760\dots \text{€} \approx 6,76 \text{ €}$.

Keskihajonta on $s = 2,509\dots \text{€} \approx 2,51 \text{ €}$.

Vastaus: $[4,71; 14,23]$, $\bar{x} \approx 6,76 \text{ €}$, $s \approx 2,51 \text{ €}$

- 513.** a) Jamin keskiarvotulos on 64,086 m ja Jimin 62,803 m, joten Jamin keskiarvotulos on parempi.

Vastaus: Jamin

- b) Jamin paras tulos on 66,42 m ja Jimin 70,66 m, joten Jimin tulos on parempi.

Vastaus: Jimin

- c) Jamin tulosten keskihajonta on 0,937 m ja Jimin 4,851 m, joten Jami heittää tasaisemmin.

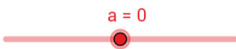
Vastaus: Jami

- d) Jamin valitsemista puoltaa hänen parempi keskiarvonsa. Jimin valintaa puoltaa hänen Jamia parempi tuloksensa edellisessä kilpailussa.

Vastaus: Jamin valitsemista puoltaa hänen parempi keskiarvonsa. Jimin valintaa puoltaa hänen Jamia parempi tuloksensa edellisessä kilpailussa.

- 514.** a) Appletin avulla havaitaan, että keskihajonta on aluksi 2,58. Kun luvut kerrotaan luvulla 10, havaitaan, että keskihajonta on 25,82. Keskihajonta tulee siis kymmenkertaiseksi.



Lisää vakio: 

Kerro vakiolla: 

Lukujen $\{-40, -30, -20, -10, 0, 10, 20, 30, 40\}$
keskihajonta on 25,82.

Vastaus: tulee kymmenkertaiseksi

- b) Huomataan, että keskihajonta ei muutu, jos kaikkiin havaintoarvoihin lisätään sama luku 10.



Lisää vakio:  $a = 10$

Kerro vakiolla:  $b = 1$

Lukujen $\{6, 7, 8, 9, 10, 11, 12, 13, 14\}$
keskihajonta on 2.58.

Vastaus: ei muutu

- c) Kun kaikki luvut kerrotaan luvulla 10, myös lukujen keskinäiset etäisyydet kymmenkertaistuvat, joten keskihajonta kymmenkertaistuu. Kun kaikkiin lukuihin lisätään luku 10, luvut siirtyvät lukusuoralla 10 askelta oikealle, mutta niiden keskinäiset etäisyydet eivät muutu, joten keskihajontakaan ei muutu.

515. Kolme jakaumaa A, B ja C, joilla on sama keskiarvo, mutta eri keskihajonta, voidaan muodostaa esimerkiksi seuraavasti. Valitaan yhteiseksi keskiarvoksi jokin luku, esimerkiksi luku 100. Valitaan jakaumaan A kaksi lukua luvun 100 molemmilta puolilta yhtä kaukaa, jotta keskiarvo säilyy, vaikkapa 99 ja 101. Jakauma A on siis 99, 100, 101.

Valitaan jakaumaan B kaksi lukua luvun 100 molemmilta puolilta yhtä kaukaa, mutta kauempaa kuin ensimmäisessä jakaumassa, esimerkiksi 90 ja 110. Jakauma B on siis 90, 100, 110.

Muodostetaan jakauma C vastaavalla tavalla, esimerkiksi 50, 100 ja 150.

Jokaisen jakauman keskiarvo on nyt 100, ja keskihajonnat ovat keskenään erisuuruksia.

Vastaus: esim. jakauma 1: 99, 100, 101, jakauma 2: 90, 100, 110, jakauma 3: 50, 100, 150

- 516.** Kopioidaan taulukko sopivaan ohjelmaan ja lasketaan jakauman keskiarvo ja keskihajonta käyttämällä suhteellisia osuuksia frekvensseinä.

	A	B
1	puoltoäänet	f
2		7 0.035
3		6 0.132
4		5 0.181
5		4 0.265
6		3 0.22
7		2 0.122
8		0 0.044
9		

Keskiarvo on $\bar{x} = 3,909... \approx 3,9$.

Keskihajonta on $s = 1,558... \approx 1,6$.

Vastaus: $\bar{x} \approx 3,9$ puoltoääntä, $s \approx 1,6$ puoltoääntä

- 517.** a) Kopioidaan taulukko sopivaan ohjelmaan. Lasketaan kaikkina kuukausina syntyneiden lasten määrän keskiarvo ohjelman avulla.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Koko vuosi	T	H	M	H	T	K	H	E	S	L	M	J
2	2010	60980	5007	4733	5297	4956	5095	5111	5441	5272	5280	5251	4716	4821
3	2011	59961	5070	4592	5072	4790	5091	5057	5309	5434	5274	4888	4803	4581
4	2012	59493	4991	4722	5061	4822	5003	5047	5384	5388	4990	4832	4769	4484
5	2013	58134	4837	4463	4970	4840	5014	4866	5247	5134	5025	4987	4400	4351
6	2014	57232	4800	4333	4806	4683	4997	4880	5227	5052	5065	4781	4232	4376
7	2015	55472	4659	4209	4603	4629	4633	4696	5010	4882	4756	4615	4500	4280
8														
9														
10	Kaikkien kuukausien keskiarvo													
11	=KESKIARVO(C2:N7)													

Kaikkien kuukausien keskiarvo
4878,778

Keskiarvoksi saadaan noin 4879 syntymää kuukaudessa.

Vastaus: 4879 lasta

- b) Lasketaan kalenterikuukausien syntymien keskiarvot. Aloitetaan laskemalla tammikuussa syntyneiden lasten määrien keskiarvo ohjelman avulla.

	A	B	C	D	E	F	G
1		Koko vuosi	T	H	M	H	
2	2010	60980	5007	4733	5297	4956	5
3	2011	59961	5070	4592	5072	4790	5
4	2012	59493	4991	4722	5061	4822	5
5	2013	58134	4837	4463	4970	4840	5
6	2014	57232	4800	4333	4806	4683	4
7	2015	55472	4659	6 R x 1 S	3	4629	4
8	Keskiarvo		=KESKIARVO(C2:C7)				

Kopioidaan keskiarvon laskukaava muiden kuukausien kohdalle.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Koko vuosi	T	H	M	H	T	K	H	E	S	L	M	J
2	2010	60980	5007	4733	5297	4956	5095	5111	5441	5272	5280	5251	4716	4821
3	2011	59961	5070	4592	5072	4790	5091	5057	5309	5434	5274	4888	4803	4581
4	2012	59493	4991	4722	5061	4822	5003	5047	5384	5388	4990	4832	4769	4484
5	2013	58134	4837	4463	4970	4840	5014	4866	5247	5134	5025	4987	4400	4351
6	2014	57232	4800	4333	4806	4683	4997	4880	5227	5052	5065	4781	4232	4376
7	2015	55472	4659	4209	4603	4629	4633	4696	5010	4882	4756	4615	4500	4280
8	Keskiarvo		4894	4509	4968	4787	4972	4943	5270	5194	5065	4892	4570	4482

Vastaus:

Kuukausi	Syntyneitä keskimäärin
tammikuu	4894
helmikuu	4509
maaliskuu	4968
huhtikuu	4787
toukokuu	4972
kesäkuu	4943
heinäkuu	5270
elokuu	5194
syyskuu	5065
lokakuu	4892
marraskuu	4570
joulukuu	4482

- c) Suurin syntyneiden määrän keskiarvo on heinäkuussa, 5270. Sen poikkeama keskiarvosta on $5270 - 4879 = 391$.

Pienin syntyneiden määrän keskiarvo on joulukuussa, 4482. Sen poikkeama keskiarvosta on $4482 - 4879 = -397$.

Lasketaan kaikkien tutkittujen kuukausien syntyneiden määrän keskihajonta ohjelman avulla.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Koko vuosi	T	H	M	H	T	K	H	E	S	L	M	J
2	2010	60980	5007	4733	5297	4956	5095	5111	5441	5272	5280	5251	4716	4821
3	2011	59961	5070	4592	5072	4790	5091	5057	5309	5434	5274	4888	4803	4581
4	2012	59493	4991	4722	5061	4822	5003	5047	5384	5388	4990	4832	4769	4484
5	2013	58134	4837	4463	4970	4840	5014	4866	5247	5134	5025	4987	4400	4351
6	2014	57232	4800	4333	4806	4683	4997	4880	5227	5052	5065	4781	4232	4376
7	2015	55472	4659	4209	4603	4629	4633	4696	5010	4882	4756	4615	4500	4280
8	Keskiarvo		4894	4509	4968	4787	4972	4943	5270	5194	5065	4892	4570	4482
9														
10	Kaikkien kuukausien keskiarvo													
11	4878,778													
12	Kaikkien kuukausien keskihajonta													
13	=KESKIHAJONTA(C2:N7)													

Kaikkien kuukausien keskihajonta														
300,8063														

Keskihajonnaksi saadaan 300,806...

Syntymien määrä, joka on kaksi keskihajontaa keskiarvoa suurempi, on $4878,777... + 2 \cdot 300,806... = 5480,390...$

Syntymien määrä, joka on kaksi keskihajontaa keskiarvoa pienempi, on $4878,777... - 2 \cdot 300,806... = 4277,165...$

Yksikään keskiarvoista ei ylitä arvoa 5480,390... eikä alita arvoa 4277,165..., joten poikkeamat eivät ole merkittäviä.

Vastaus: ei poikkea

- 518.** Puhelinten käyttöikien moodi on 61–70 kk, koska sen frekvenssi on suurin. Muiden sijaintilukujen määrittämiseksi lasketaan luokkakeskukset ja kirjoitetaan frekvenssit sopivaan ohjelmaan. Ensimmäisen luokan 21–30 todellinen alaraja on 20,5 ja todellinen yläraja 30,5, joten luokkakeskus on $\frac{20,5 + 30,5}{2} = 25,5$.

Lasketaan kysytyt tunnusluvut ohjelman avulla.

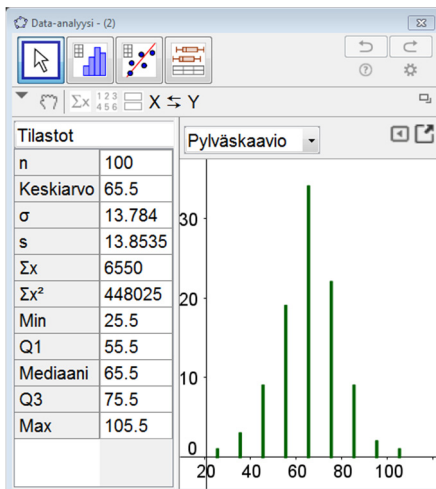
GeoGebra

Tilastot Kirjautu sisään...

f_x L K $\{1,2\}$ Σ

	A	B	C
1	25.5	1	
2	35.5	3	
3	45.5	9	
4	55.5	19	
5	65.5	34	
6	75.5	22	
7	85.5	9	
8	95.5	2	
9	105.5	1	
10			
11			
12			
13			

Syöttökenttä:



Ohjelma antaa alakvartiiliksi $Q_1 = 55,5$ kk, mediaaniksi $Q_2 = 65,5$ kk ja yläkvartiiliksi $Q_3 = 75,5$ kk sekä keskiarvoksi 65,5 kk ja keskihajonnaksi $s = 13,853\dots$ kk $\approx 13,9$ kk

Vastaus: $Mo = 61-70$ kk, $Q_1 = 55,5$ kk, $Q_2 = 65,5$ kk, $Q_3 = 75,5$ kk, $\bar{x} = 65,5$ kk, $s \approx 13,9$ kk

- 519. a)** Naisia on ikäluokassa 65–69 eniten, joten naisten moodi on 65–69 vuotta. Vastaavasti miesten moodi on 50–54 vuotta. Muiden tunnuslukujen määrittämiseksi lasketaan luokkakeskukset ja kirjoitetaan frekvenssit sopivaan ohjelmaan. Ensimmäisen luokan 0–4 todellinen alaraja on 0 ja todellinen yläraja 5, koska ikä pyöristetään aina alaspäin. Ensimmäisen luokan luokkakeskus on $\frac{0+5}{2} = 2,5$.

	A	B	C
1	Luokkakeskus	Miehet	Naiset
2	2.5	150579	143799
3	7.5	157405	150313
4	12.5	150064	143863
5	17.5	154051	147120
6	22.5	173332	165884
7	27.5	175568	166960
8	32.5	183915	173017
9	37.5	177811	168005
10	42.5	162832	155047
11	47.5	177270	173099
12	52.5	188139	186362
13	57.5	182452	185658
14	62.5	181417	190243
15	67.5	183252	198011
16	72.5	121120	140045
17	77.5	84974	112220
18	82.5	56388	87847
19	87.5	30175	63529
20	92.5	9248	28158
21	97.5	1386	5991
22	102.5	112	647

Määritetään naisten ikäjakauman tunnusluvut.

Naisten iän jakaumassa

$Md = 42,5$ vuotta

$\bar{x} \approx 43,6$ vuotta

$Q_1 = 22,5$ vuotta

$Q_3 = 62,5$ vuotta

$s \approx 24,3$ vuotta

n	2785818
Keskiarvo	43.63
σ	24.3048
s	24.3048
Σx	121545265
Σx^2	6948672912.5
Min	2.5
Q1	22.5
Mediaani	42.5
Q3	62.5
Max	102.5

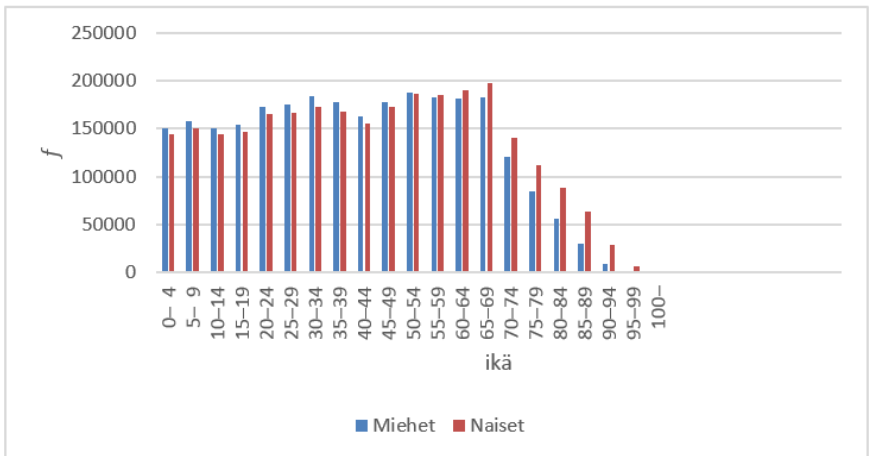
Vastaavasti saadaan miesten jakauman tunnusluvut.

Miesten iän jakaumassa
 $Md = 42,5$ vuotta
 $\bar{x} \approx 40,9$ vuotta
 $Q_1 = 22,5$ vuotta
 $Q_3 = 57,5$ vuotta
 $s \approx 23,0$ vuotta

n	2701490
Keskiarvo	40.933
σ	23.0368
s	23.0368
Σx	110580025
Σx^2	5960034012.5
Min	2.5
Q1	22.5
Mediaani	42.5
Q3	57.5
Max	102.5

Vastaus: Naiset: $Mo = 65-69$ vuotta $Md = 42,5$ vuotta,
 $Q_1 = 22,5$ vuotta, $Q_3 = 62,5$ vuotta, $\bar{x} \approx 43,6$ vuotta, $s \approx 24,3$ vuotta,
miehet: $Mo = 50-54$ vuotta, $Md = 42,5$ vuotta, $Q_1 = 22,5$ vuotta,
 $Q_3 = 57,5$ vuotta, $\bar{x} \approx 40,9$ vuotta, $s \approx 23,0$ vuotta

b) Piirretään jakaumien pylväskaavio sopivalla ohjelmalla.



520. a) Kopioidaan luvut taulukkolaskentaohjelmaan. Määritetään Anteron tulosten keskiarvo ja keskihajonta ohjelman avulla.

	A	B	C
1		Antero	Lauri
2		620	590
3		630	670
4		628	610
5		645	678
6		602	560
7		635	650
8		630	634
9		644	682
10		657	604
11		630	625
12		634	640
13		642	592
14		633	671
15			
16	keskiarvo	=KESKIARVO(B2:B14)	=KESKIARVO(C2:C14)
17	keskihajonta	=KESKIAJONTA(B2:B14)	=KESKIAJONTA(C2:C14)

Anteron tuloskeskiarvo on noin 633 cm ja Laurin 631 cm.

Määritetään Anteron tulosten keskihajonta ohjelman avulla.

Anteron tulosten keskihajonta on noin 13 cm ja Laurin noin 39 cm.

Vastaus: Antero: $\bar{x} \approx 633$ cm, $s \approx 13$ cm

Lauri: $\bar{x} \approx 631$ cm, $s \approx 39$ cm

- b) Jos kilpailussa hypätään vain yksi hyppy, niin Antero luultavasti voittaa kilpailun, koska hänen tulostensa keskiarvo on parempi kuin Laurin.

Jos kilpailussa saa hypätä niin monta hyppyä kuin haluaa, niin Lauri luultavasti voittaa kilpailun, koska hänen tulostensa keskihajonta on paljon suurempi kuin Laurin. Vaikka Lauri hyppää keskimäärin huonommin, niin hänen hyppyjensä sekaan mahtuu hyvin pitkiä hyppyjä, joita Antero ei juuri koskaan hyppää.

Vastaus: yhden hypyn kilpailussa Antero, rajoittamattoman hyppymäärän kilpailussa Lauri

SYVENNÄ YMMÄRRYSTÄ

521. Esimerkiksi jakauma

x	f
1	5
2	10
3	5

toteuttaa vaaditun ehdon sillä jakauman keskiarvo on 2 ja keskihajonta noin 0,73, jolloin havaintoarvot 1 ja 3 eivät ole yli kahden keskihajonnan päässä keskiarvosta 2.

n	20
Keskiarvo	2
σ	0.7071
s	0.7255
Σx	40
Σx^2	90
Min	1
Q1	1.5
Mediaani	2
Q3	2.5
Max	3

Vastaus: esim.

x	1	2	3
f	5	10	5

522. Merkitään luokkien ylärajat taulukkoon ja luokitellaan aineisto niiden mukaan.

Sikiäimen pituus	luokan yläraja	Frekvenssi	Kypsan kurkun pituus	luokan yläraja	Frekvenssi
1,0-1,5	1,5	2	12-16	16	10
1,5-2,0	2	8	16-20	20	19
2,0-2,5	2,5	32	20-24	24	36
2,5-3,0	3	29	24-28	28	31
3,0-3,5	3,5	27	28-32	32	23
3,5-4,0	4	23	32-36	36	10
4,0-4,5	4,5	8	36-40	40	5
4,5-5,0	5	4	40-44	44	1
5,0-5,5	5,5	1			0
5,5-6,0	6	1			
		0			

Määritetään luokittelemattoman aineiston tunnusluvut.

	5,90	29,50
keskiarvo	2,998	24,812
keskihajonta	0,814	5,979

Sikiäin: keskiarvo $\bar{x} \approx 3,00$ cm ja keskihajonta $s \approx 0,81$ cm

Kypsa kurkku: keskiarvo $\bar{x} \approx 24,8$ cm ja keskihajonta $s \approx 5,98$ cm

Määritetään luokitellun aineiston tunnusluvut sikiäimelle. Ensimmäisen luokan 1,0–1,5 todellinen alaraja on 0,95 ja todellinen yläraja 1,45 ja

luokkakeskus on $\frac{0,95 + 1,45}{2} = 1,2$.

Sikiäimen pituus (cm)	f
1.2	2
1.7	8
2.2	32
2.7	29
3.2	27
3.7	23
4.2	8
4.7	4
5.2	1
5.7	1

n	135
Keskiarvo	2.9593
σ	0.8184
s	0.8215
Σx	399.5
Σx^2	1272.65
Min	1.2
Q1	2.2
Mediaani	2.7
Q3	3.7
Max	5.7

Sikiäimen keskiarvo
 $\bar{x} \approx 2,96$ cm ja
 keskihajonta
 $s \approx 0,82$ cm

Määritetään luokitellun aineiston tunnusluvut kypsälle kurkulle.
Ensimmäisen luokan 12–16 todellinen alaraja on 11,5 ja todellinen yläraja 16,5 ja luokkakeskus on $\frac{11,5 + 16,5}{2} = 14$.

Kypsän kurkun pituus (cm)	f
14	10
18	19
22	36
26	31
30	23
34	10
38	5
42	1

n	135
Keskiarvo	24,7556
σ	6,09
s	6,1127
Σx	3342
Σx^2	87740
Min	14
Q1	22
Mediaani	26
Q3	30
Max	42

Kypsän kurkun keskiarvo $\bar{x} \approx 24,8$ cm ja keskihajonta $s \approx 6,11$ cm.

Havaitaan, että luokitellun aineiston keskiarvot ovat kummallakin hieman suuremmat kuin luokittelemattomassa aineistossa.

Vastaus: Luokittelematon aineisto
sikiäin: $\bar{x} \approx 3,00$ cm, $s \approx 0,81$ cm
kypsä kurkku: $\bar{x} \approx 24,8$ cm, $s \approx 5,98$ cm

Luokiteltu aineisto
sikiäin: $\bar{x} \approx 2,96$ cm, $s \approx 0,82$ cm
kypsä kurkku: $\bar{x} \approx 24,8$ cm, $s \approx 6,11$ cm,

Luokitellun ja luokittelemattoman aineiston tunnusluvut ovat hyvin lähellä toisiaan. Kypsän kurkun luokitellun aineiston keskihajonta on hieman suurempi kuin luokittelemattomassa aineistossa.

523. Lukujen x , $4x + 3$, $5x + 1$, $4 - 2x$ ja $-6x$ keskiarvo on 1. Muodostetaan tämän tiedon perusteella yhtälö ja ratkaistaan siitä x .

$$\begin{aligned} \frac{x + 4x + 3 + 5x + 1 + 4 - 2x - 6x}{5} &= 1 \\ \frac{2x + 8}{5} &= 1 && \parallel \cdot 5 \\ 2x + 8 &= 5 \\ 2x &= 5 - 8 \\ 2x &= -3 && \parallel : 2 \\ x &= -1,5 \end{aligned}$$

Luvut ovat siis

$$x = -1,5$$

$$4x + 3 = 4 \cdot (-1,5) + 3 = -3$$

$$5x + 1 = 5 \cdot (-1,5) + 1 = -6,5$$

$$4 - 2x = 4 - 2 \cdot (-1,5) = 7$$

$$-6x = -6 \cdot (-1,5) = 9$$

Määritetään lukujen keskihajonta sopivalla ohjelmalla.

	A	B	C
1		-1,5	
2		-3	
3		-6,5	
4		7	
5		9	
6	Keskihajonta	=KESKIHAJONTA(B1:B5)	

	A	B
1		-1,5
2		-3
3		-6,5
4		7
5		9
6	Keskihajonta	6,680194608

Keskihajonta on $s = 6,680 \dots \approx 6,7$.

Vastaus: $s \approx 6,7$

524. Määritetään lukumääräarvioille keskiarvot ja keskihajonnat.

Aineisto luokiteltuna

Lukumäärän 5 arviot	<i>f</i>	Keskiarvo Keskihajonta			
5	102	5,1	0,27		
6	4				
7	1				
Lukumäärän 6 arviot	<i>f</i>	6,09	0,42		
5	7				
6	120				
7	20	7,04	0,52		
				8,00	0,68
Lukumäärän 7 arviot	<i>f</i>				
6	18				
7	113	8,00	0,68		
8	25				
				8,00	0,68
Lukumäärän 8 arviot	<i>f</i>				
7	30				
8	76	8,00	0,68		
9	28				
10	1				

Lukumäärän 9 arviot	f	Keskiarvo Keskihajonta	
7	2		
8	24		
9	76		
10	18		
11	2		

8,95 0,69

Lukumäärän 10 arviot	f
8	6
9	37
10	46
11	16
12	2

9,73 0,85

Lukumäärän 11 arviot	f
8	1
9	11
10	19
11	26
12	12

10,54 1,01

Lukumäärän 12 arviot	f	Keskiarvo Keskihajonta	
9	1		
10	4		
11	17		
12	19		
13	3		
14	1		

11,49 0,92

Lukumäärän 13 arviot	f
11	7
12	11
13	6
14	1
15	1

12,15 1,01

Lukumäärän 14 arviot	f
11	2
12	3
13	3
14	4
15	2

13,07 1,33

Lukumäärän 15 arviot	f	Keskiarvo Keskihajonta
12	2	
13	1	
14	6	
15	2	

13,73 1,01

Vastaus:

	Arvioiden keskiarvo	Arvioiden keskihajonta
Lukumäärän 3 arvio	3,0	0,0
Lukumäärän 4 arvio	4,0	0,0
Lukumäärän 5 arviot	5,1	0,27
Lukumäärän 6 arviot	6,09	0,42
Lukumäärän 7 arviot	7,04	0,52
Lukumäärän 8 arviot	8,00	0,68
Lukumäärän 9 arviot	8,95	0,69
Lukumäärän 10 arviot	9,73	0,85
Lukumäärän 11 arviot	10,54	1,01
Lukumäärän 12 arviot	11,49	0,92
Lukumäärän 13 arviot	12,2	1,01
Lukumäärän 14 arviot	13,1	1,33
Lukumäärän 15 arviot	13,7	1,01

- 525.** Kopioidaan taulukko sopivaan ohjelmaan. Määritetään ohjelman avulla koneiden A ja B pussien keskiarvot ja keskihajonnat.

	A	B	C
1		Kone A	Kone B
2		iso pussi (g)	pieni pussi (g)
3		1010	205
4		980	198
5		975	189
6		1005	215
7		1012	210
8		1020	190
9		990	189
10		985	211
11		1011	209
12	Keskiarvo	998,6666667	201,7777778
13	Keskihajonta	16,29417074	10,42566278
14	Variaatiokerroin	=B13/B12	

Variaatiokerroin on koneella A 0,0163... ja koneella B 0,0516..., joten kone A pakkaa suhteellisesti tarkemmin.

Vastaus: kone A

- 526.** Normitetaan Rasmuksen koetulokset.

Biologia: $\frac{15 - 20,4}{4,5} = -1,2$ eli 1,2 keskihajontaa alle keskiarvon.

Kemia: $\frac{21 - 25,3}{3,7} = -1,162...$ eli 1,162... keskihajontaa alle keskiarvon.

Rasmuksen kemian koetulos poikkesi keskiarvosta suhteellisesti vähemmän alaspäin kuin kemian koetulos, joten hän menestyi kemiassa paremmin kuin biologiassa.

Vastaus: kemiassa

5.2 Regressio ja korrelaatio

ALOITA PERUSTEISTA

- 527.** Kun korrelaatio on negatiivinen, hajontakuvio asettuu laskevan suoran muotoon, joten ilmaisu A vastaa hajontakuvio II.

Kun korrelaatiota ei ole, hajontakuvio on sellainen, johon ei voi selkeästi asettaa sitä edustavaa suoraa, joten ilmaisuun B sopii hajontakuvio III.

Kun korrelaatio on positiivinen, hajontakuvio asettuu nousevan suoran muotoon, joten ilmaisu C vastaa hajontakuvio I.

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

- 528.** Kuviosta I havaitaan, että kyseessä on voimakas negatiivinen korrelaatio, joten siihen sopii korrelaatiokerroin C: $-0,995$.

Kuviosta II havaitaan, että kyseessä on merkityksetön positiivinen korrelaatio, joten siihen sopii korrelaatiokerroin A: $0,123$.

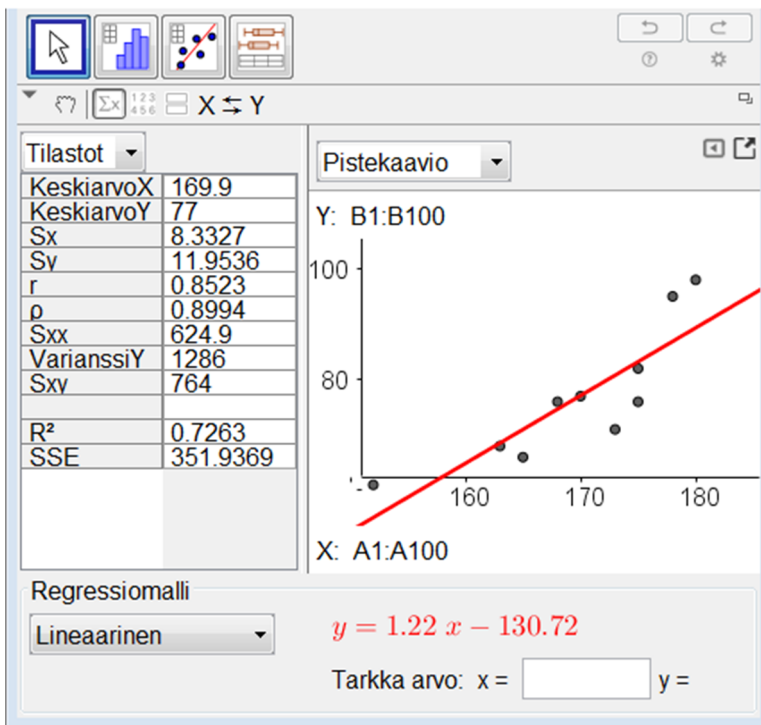
Kuviosta III havaitaan, että kyseessä on voimakas positiivinen korrelaatio, joten siihen sopii korrelaatiokerroin B: $0,995$.

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

- 529.** Korrelaatiokertoimen ohjeellisen voimakkuustaulukon perusteella
- a) $r = 0,2$ korrelaatio on merkityksetön
 - b) $r = -0,81$ korrelaatio on voimakas
 - c) $r = 0,99$ korrelaatio on voimakas
 - d) $r = 0,32$ korrelaatio on kohtalainen
 - e) $r = 0,72$ korrelaatio on huomattava
 - f) $r = -0,01$ korrelaatio on merkityksetön

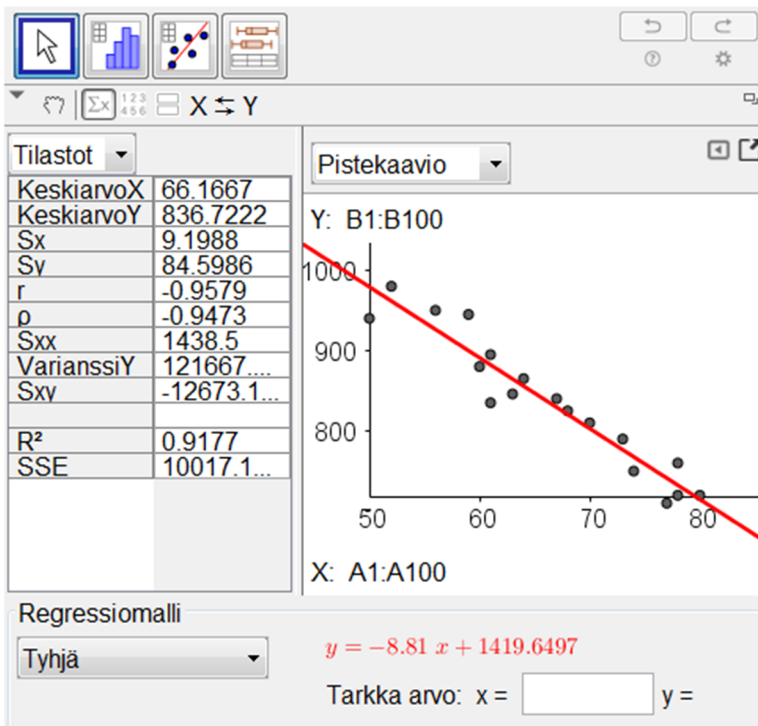
Vastaus: a) merkityksetön b) voimakas c) voimakas
d) kohtalainen e) huomattava f) merkityksetön

- 530.** Määritetään korrelaatiokerroin ohjelman avulla.
Ohjelma antaa korrelaatiokertoimeksi $r \approx 0,85$ pituuden ja massan välille.



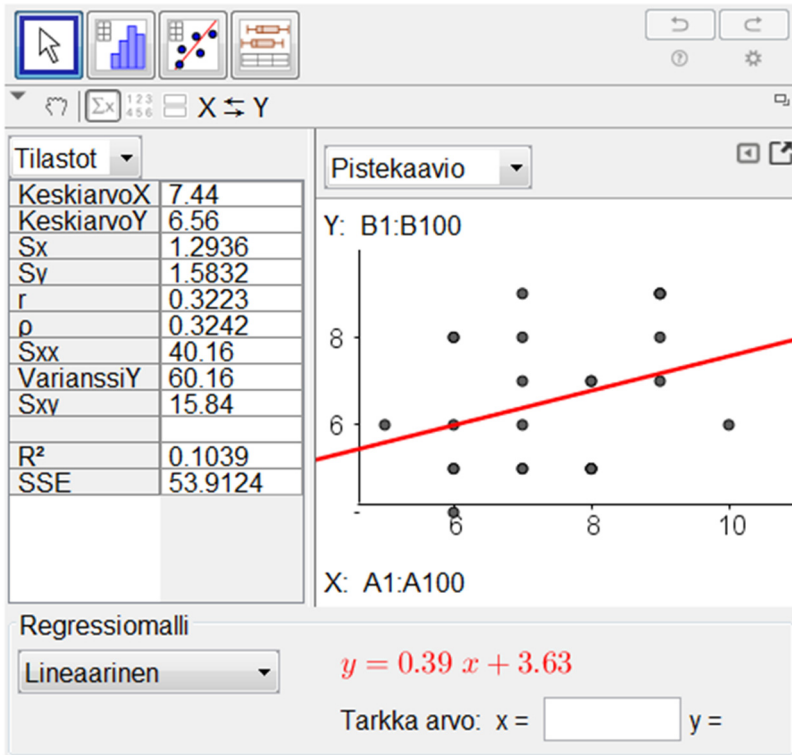
Vastaus: $r \approx 0,85$

531. Ohjelma antaa selitystasteeksi $r^2 = 0,9177 \approx 0,92 = 92 \%$, joten ihmisen ikä selittää luiden tiheydestä noin 92 %.



Vastaus: n. 92 %

532. Määritetään taulukon ja ohjelman avulla tilastolliset tunnusluvut.



- a) Matematiikan ja ruotsin kielen arvosanojen välinen korrelaatiokerroin on $r = 0,3223 \approx 0,32$. Riippuvuuden voimakkuus on kohtalainen.

Vastaus: 0,32, kohtalainen positiivinen

- b) Matematiikan ja ruotsin kielen arvosanojen riippuvuutta kuvaavan regressiosuoran yhtälö on $y = 0,39x + 3,63$, jossa x on matematiikan arvosana ja y on ruotsinkielen arvosana.

Vastaus: $y = 0,39x + 3,63$

- c) Lasketaan ohjelman avulla arvio ruotsin kielen arvosanalle, kun matematiikan arvosana on 8.

Regressiomalli

Tyhjä

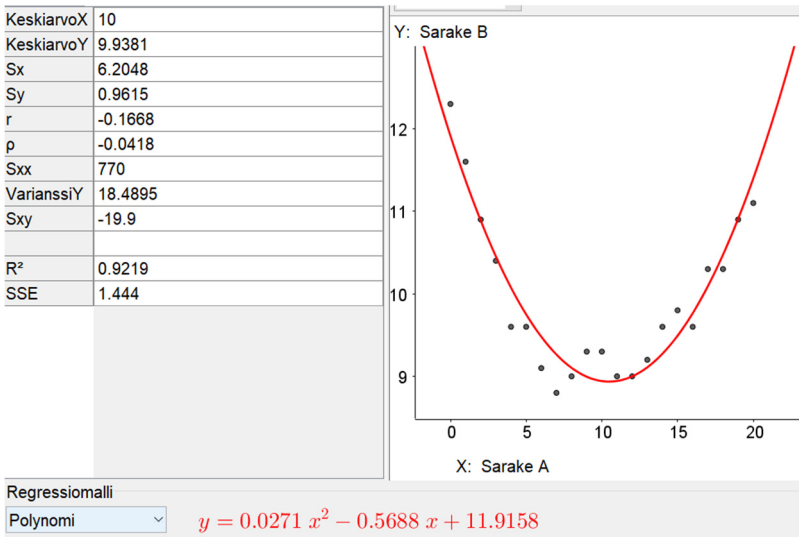
$y = 0.39x + 3.63$

Tarkka arvo: x = 8 y = 6.7809

Mallin mukaan ruotsin arvosana on $6,780... \approx 7$, jos matematiikan arvosana on 8.

Vastaus: arvosana 7

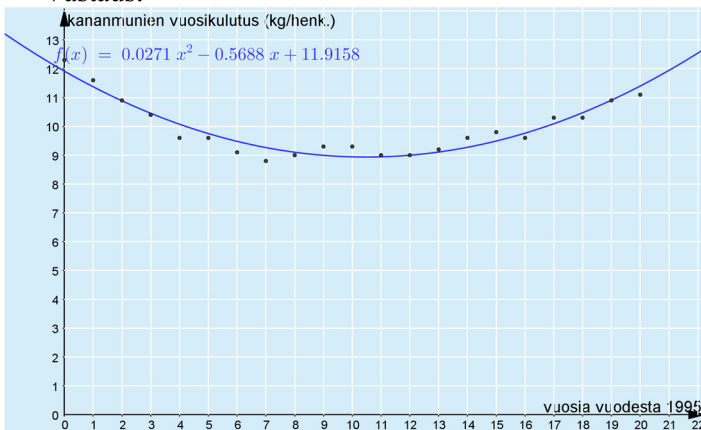
533. a) Kopioidaan taulukko sopivaan ohjelmaan ja sovitetaan havaintoaineistoon 2. asteen polynominen malli.



Ohjelma antaa malliksi $y = 0,0271x^2 - 0,5688x + 11,9158$, jossa y on kananmunien kulutus kilogrammoina henkeä kohti ja x on aika vuosina vuodesta 1995.

Mallin selityssaste on $R^2 = 0,9219 \approx 0,92$ eli aika selittää noin 92 % kananmunien kulutuksesta.

Vastaus:



$$y = 0,0271x^2 - 0,5688x + 11,9158, r^2 \approx 0,92.$$

- b) Määritetään ohjelmalla arviot kananmunien kulutukselle vuosina 2009 ja 2021. Syötetään muuttujan arvoiksi $2009 - 1995 = 14$ ja $2021 - 1995 = 26$.

X: Sarake A

Regressiomalli

Polynomi $y = 0.0271 x^2 - 0.5688 x + 11.9158$

2 Tarkka arvo: x = y = 9.2737

Regressiomalli

Polynomi $y = 0.0271 x^2 - 0.5688 x + 11.9158$

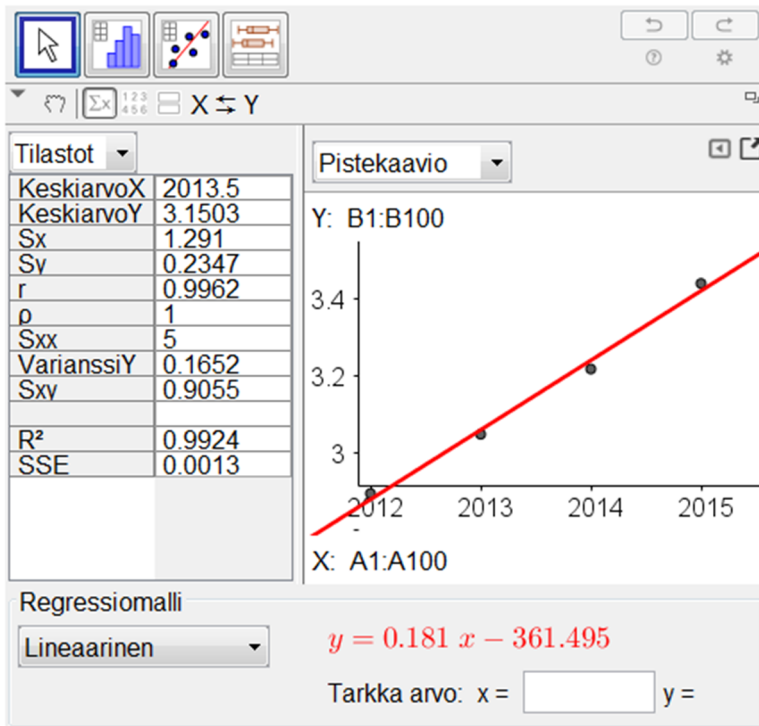
2 Tarkka arvo: x = y = 15.4787

Arvioitu vuosikulutus vuonna 2009 on noin 9,3 kg/hlö ja vuonna 2021 noin 15,5 kg/hlö.

Vastaus: 2009: 9,3 kg/henkilö, 2021: 15,5 kg/henkilö

VAHVISTA OSAAMISTA

534. Kopioidaan taulukko sopivaan ohjelmaan ja analysoidaan aineistoa.



- a) Vuosiluvun ja matkustajamäärän välinen korrelaatiokerroin $r = 0,9962 \approx 0,996$. Korrelaatio on voimakas ja positiivinen.

Vastaus: $r \approx 0,996$, voimakas positiivinen

- b) Vuosiluvun ja matkustajamäärän riippuvuutta kuvaavan regressiosuoran yhtälö on $y = 0,181x - 361,495$, jossa x on vuosiluku ja y on lentomatkustajien määrä miljardeina.

Vastaus: $y = 0,181x - 361,495$, jossa x on vuosiluku ja y on lentomatkustajien määrä miljardeina

- c) Lasketaan ohjelman avulla arvio lentomatkustajien määrälle vuonna 2025.

Regressiomalli

Lineaarinen

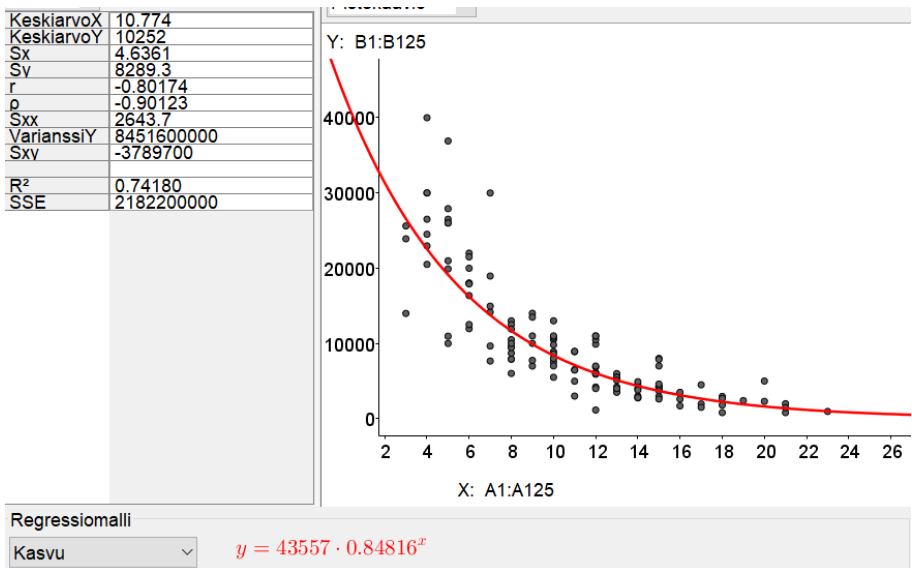
$y = 0.1811x - 361.4946$

Tarkka arvo: x = 2025 y = 5.2329

Lentomatkustajien määrä vuonna 2025 on mallin mukaan 5,23 miljardia.

Vastaus: 5,23 mrd

535. a) Sovitetaan havaintoaineistoon eksponentiaalinen malli.



Eksponentiaalinen malli on $y = 43\,557 \cdot 0,848^x$, jossa x on auton ikä ja y on auton hinta.

Vastaus: $y = 43\,557 \cdot 0,848^x$, jossa x on auton ikä ja y on auton hinta.

b) Selitysaste on $0,7418 \approx 74 \%$.

Vastaus: 74 %

c) Eksponentiaalisen mallin kantaluvusta 0,848 nähdään, että auton arvo on joka vuosi 0,848-kertainen edellisen vuoden arvoon verrattuna, joten auton arvo vähenee joka vuosi $100 \% - 84,8 \% = 15,2 \% \approx 15 \%$.

d) Ohjelma antaa 4 vuotta vanhan auton keskihinnaksi noin 22 540 €.

$$y = 43557 \cdot 0.84816^x$$

Tarkka arvo: x = 4 y = 22540

Vastaus: 22 540 €

536.

	Pojan pituus	Tytön pituus	Äidin pituus	Isän pituus
Pojan pituus	1,000		0,508	0,435
Tytön pituus		1,000	0,506	0,551
Äidin pituus			1,000	0,248
Isän pituus				1,000

Lähde: Tuomas Reiterä, HY, kandidaatin tutkielma (2015)

Lasten ja vanhempien pituuksien välistä korrelaatiokerrointaulukkoa tulkitaan seuraavasti:

1. rivin 1. ruutu: pojan pituuden korrelaatiokerroin itsensä kanssa on 1,000 eli täydellinen riippuvuus.

1. rivin 3. ruutu: pojan ja äidin pituuden korrelaatiokerroin on 0,508 eli kohtalainen positiivinen riippuvuus.

1. rivin 4. ruutu: pojan ja isän pituuden korrelaatiokerroin on 0,435 eli kohtalainen positiivinen riippuvuus.

2. rivin 2. ruutu: tytön pituuden korrelaatiokerroin itsensä kanssa on 1,000 eli täydellinen riippuvuus.

2. rivin 3. ruutu: tytön ja äidin pituuden korrelaatiokerroin on 0,506 eli kohtalainen positiivinen riippuvuus.

2. rivin 4. ruutu: tytön ja isän pituuden korrelaatiokerroin on 0,551 eli kohtalainen positiivinen riippuvuus.

3. rivin 3. ruutu: äidin pituuden korrelaatiokerroin itsensä kanssa on 1,000 eli täydellinen riippuvuus.

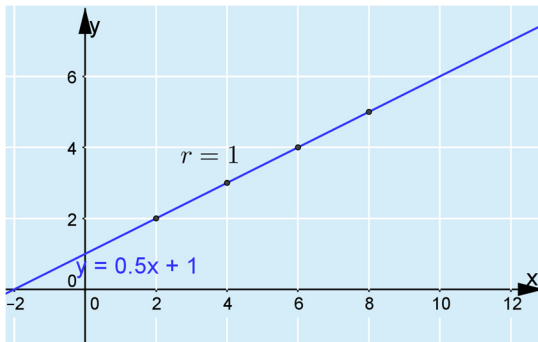
3. rivin 4. ruutu: äidin ja isän pituuden korrelaatiokerroin on 0,248 eli merkityksetön positiivinen riippuvuus.

4. rivin 4. ruutu: isän pituuden korrelaatiokerroin itsensä kanssa on 1,000 eli täydellinen riippuvuus.

Vastaus: Äidin ja lasten pituuksien välinen riippuvuus on kohtalainen positiivinen, samoin isän ja lasten. Isän ja tytön pituuksien välinen riippuvuus on voimakkain. Aviopuolisoiden välinen riippuvuus on merkityksetön positiivinen

537. a) Täydellinen positiivinen ($r = 1$) korrelaatio on kyseessä silloin, kun kaikki pisteet sijaitsevat samalla suoralla. Esimerkiksi aineisto

x	y
2	2
4	3
6	4
8	5

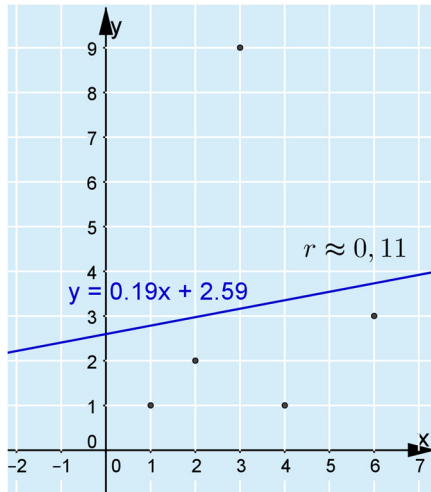


Vastaus: esim.

x	y
2	2
4	3
6	4
8	5

- b) kahdella muuttujalla ei ole korrelaatiota, jos pisteet sijaitsevat niin hajallaan, ettei pistejoukkoon voi asettaa sopivaa suoraa. Esimerkiksi aineisto

x	y
1	1
2	2
3	9
4	1
6	3

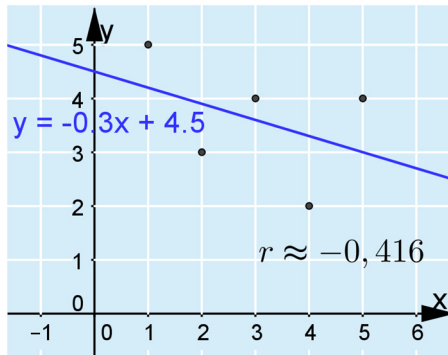


Vastaus: esim.

x	y
1	1
2	2
3	9
4	1
6	3

- c) Muuttujien välillä on kohtalainen negatiivinen korrelaatio, kun pistejoukkoon voi sovittaa laskevan suoran. Esimerkiksi aineisto

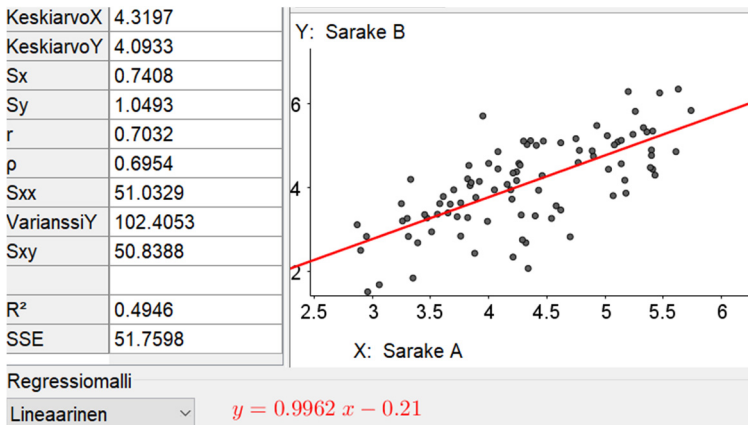
x	y
1	5
2	3
3	4
4	2
5	4



Vastaus: esim.

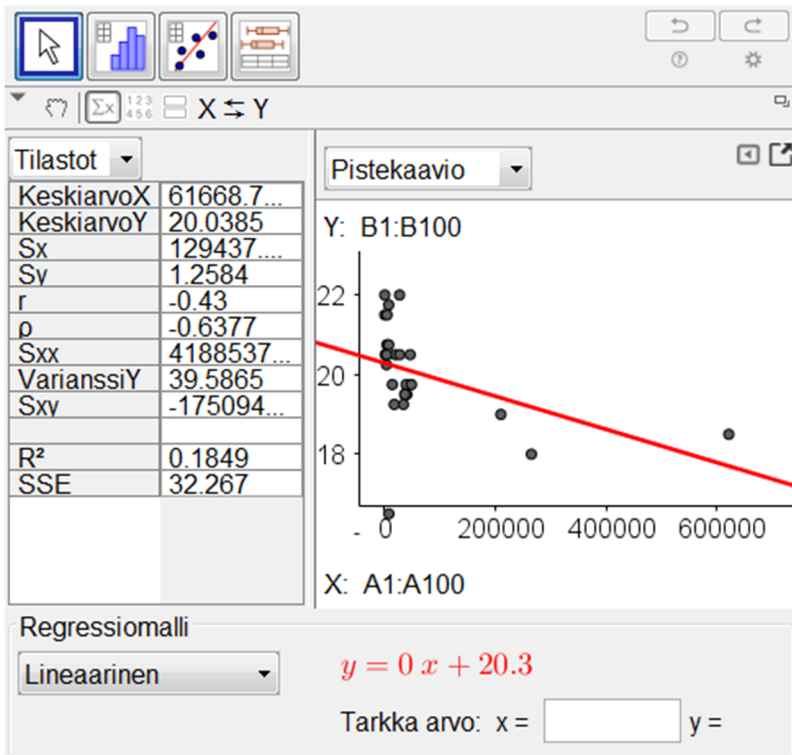
x	y
1	5
2	3
3	4
4	2
5	4

- 538.** Kopioidaan taulukko sopivaan ohjelmaan. Korrelaatiokertoimeksi saadaan $r = 0,70$, joka on huomattava positiivinen korrelaatio.



Vastaus: $r = 0,70$, joka on huomattava positiivinen korrelaatio.

- 539.** Kopioidaan taulukko sopivaan ohjelmaan. Määritetään ohjelman avulla kunnan väkiluku ja kunnallisveroprosentin välinen korrelaatiokerroin ja selitysaste.



Ohjelma antaa korrelaatiokertoimeksi $r = -0,43$ ja selitysasteeksi $r^2 = 0,1849 \approx 18 \%$.

Korrelaatio on kohtalaisen negatiivinen eli väkiluvun kasvaessa veroprosentti pienenee. Väkiluku selittää noin 18 % veroprosentin vaihtelusta.

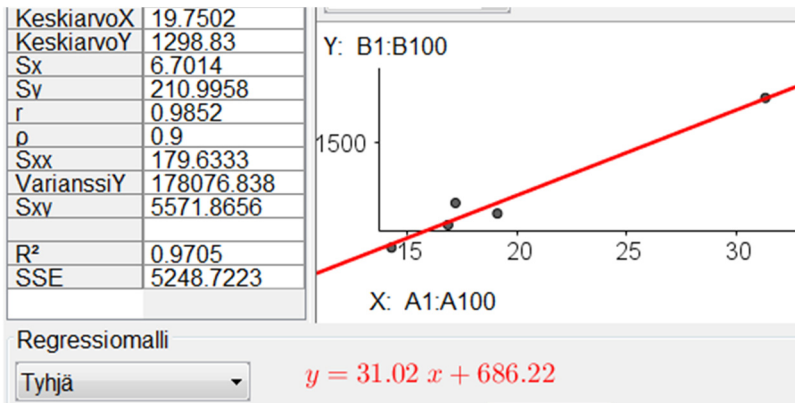
Vastaus: $r = -0,43$, 18 %

540. Kopioidaan taulukko sopivaan ohjelmaan ja määritetään kaikki kolme riippuvuussuhdetta: mobiili ja kiinteä, kiinteä ja SMS sekä mobiili ja SMS.

Viestintävirasto	Mobiililaajakaistaliittymät (kpl)	Lankapuhelinliittymät (kpl)	SMS- ja MMS-viestit yhteensä (milj. kpl)
6/2007	0	1830000	1556
12/2007	143100	1760000	1662
6/2008	307100	1690000	1686
12/2008	479700	1650000	1836
6/2009	664300	1530000	1865
12/2009	908000	1430000	1975
6/2010	1152200	1350000	1950
12/2010	1636300	1250000	2094
6/2011	1825400	1180000	2068
12/2011	1989500	1070000	2219
6/2012	2153500	980000	2181
12/2012	2347200	890000	2201
6/2013	2900000	810000	2016
12/2013	3380000	750000	1939
6/2014	3730000	690000	1722
	korrelaatiokerroin	selitysaste	
mobiili ja lankapuhelin	$-0.985 \dots \approx 0,99$	$0,971 \dots \approx 0,97$	
lankapuhelin ja SMS	$-0.590 \dots \approx -0,59$	$0,348 \dots \approx 0,35$	
mobiili ja SMS	$0.467 \dots \approx 0,47$	$0,218 \dots \approx 0,22$	

Vastaus: Mobiililaajakaistaliittymien ja lankapuhelinliittymien välillä $r \approx -0,99$ ja selitysaste 97 %, mobiililaajakaistaliittymien ja viestien välillä $r \approx 0,47$ ja selitysaste 22 %, lankapuhelinliittymien ja viestien välillä $r \approx -0,59$ ja selitysaste 35 %

- 541.** Kopioidaan taulukko sopivaan ohjelmaan ja analysoidaan aineistoa ohjelman avulla.



Tuloksiksi saadaan seuraavat:

$$r = 0,9852$$

$$r^2 = 0,9705 \approx 97 \%$$

$$\text{regressiosuoran yhtälö } y = 31,02x + 686,22$$

Saatujen tulosten perusteella:

- a)** Väite on epätosi, koska hopean ja kullan hinnoilla on voimakas positiivinen korrelaatio.

Vastaus: epätosi, voimakas

- b)** Väite on tosi, koska hopean hinta selittää 97 % kullan hinnan vaihtelusta.

Vastaus: tosi

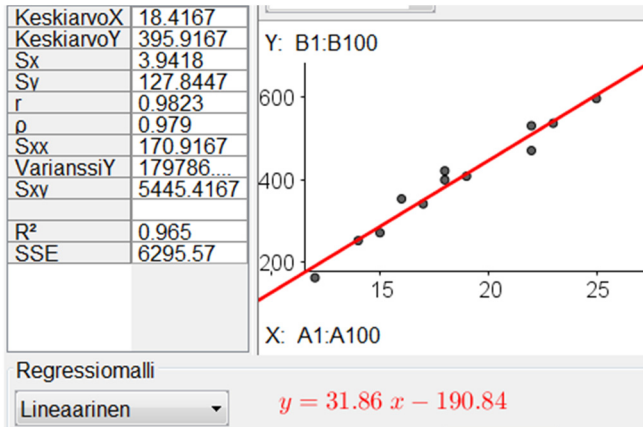
- c)** Väite on tosi, koska kullan hinta = $31,02 \cdot \text{hopean hinta} + 686,22$.

Vastaus: tosi

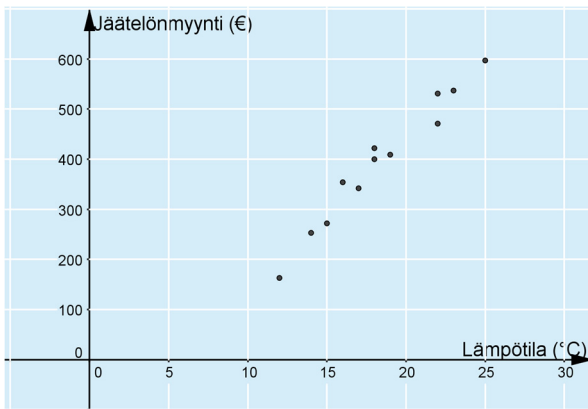
- d)** Väite on epätosi, koska kullan hinnan noustessa hopean hinta nousee, koska regressiosuora on nouseva.

Vastaus: epätosi, nousee

542. Kopioidaan taulukko sopivaan ohjelmaan ja analysoidaan aineistoa ohjelman avulla.



- a) Piirretään hajontakuviio.



- b) Selitysaste on $r^2 = 0,965 \approx 97\%$, joten lämpötila selittää 97 % jäätelönmyynnin vaihtelusta.

Vastaus: 97 %

- c) Lämpötilan ja jäätelömyynnin riippuvuutta kuvaavan regressiosuoran yhtälö on $y = 31,86x - 190,84$.

Vastaus: $y = 31,86x - 190,84$

- d) Ohjelma antaa mallin mukaiseksi myynnin arvoksi noin 765 €, kun lämpötila x on 30 astetta.

Regressiomalli	
Lineaarinen	$y = 31.86007x - 190.83959$
Tarkka arvo: $x =$ 30	$y = 764.96246$

Vastaus: 765 €

543. Kopioidaan taulukko sopivaan ohjelmaan ja analysoidaan aineistoa ohjelman avulla.

- a) Määritetään korrelaatiokertoimet muuttujien välille.
vuosiluku ja hedelmällisyysluku: $r = -0,996$
vuosiluku ja elinajanodote: $r = 0,994$
hedelmällisyysluku ja elinajanodote: $r = -0,987$

Vastaus: vuosiluku ja hedelmällisyysluku $r \approx -0,996$, vuosiluku ja elinajanodote $r \approx 0,994$, hedelmällisyysluku ja elinajanodote $r \approx -0,987$

- b) Koska vuosiluvun ja hedelmällisyysluvun välinen korrelaatio on voimakas negatiivinen, voidaan sanoa, että nykyään maailmassa naiset synnyttävät vähemmän lapsia elinaikanaan kuin ennen.
Koska vuosiluvun ja elinajanodotteen välinen korrelaatio on voimakas positiivinen, voidaan sanoa, että nykyään eletään pidempään kuin ennen.

Vastaus: Väite pitää paikkansa

- c) Hedelmällisyyslukujen ja elinajanodotteen välillä on voimakas negatiivinen korrelaatio, eli hedelmällisyysluvun pienetessä elinajanodote kasvaa.

Vastaus: voimakas negatiivinen korrelaatio

SYVENNÄ YMMÄRRYSTÄ

- 544.** Kopioidaan taulukko sopivaan ohjelmaan ja määritetään muuttujien väliset korrelaatiokertoimet.

Kuolemien ja avioerojen välillä $r \approx 0,91$

Avioerojen ja väestömäärän välillä $r \approx 0,96$

Kuolemien ja väestömäärän välillä $r \approx 0,96$

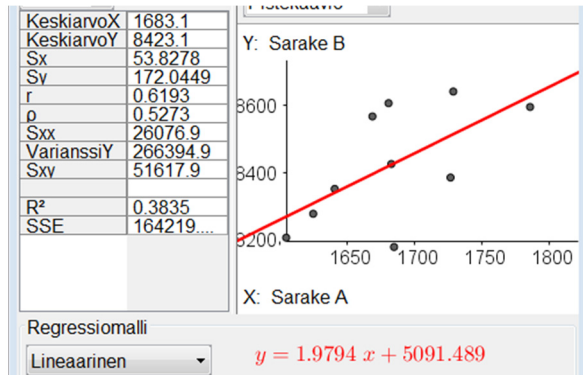
Kaikkien muuttujien väliset korrelaatiot ovat voimakkaan positiivisia, joka johtuu siitä, että väestömäärän ollessa suuri on kuolemien ja avioerojenkin määrä suuri. Avioerojen ja kuolemien määrien välillä on voimakas positiivinen korrelaatio, mutta niiden välillä tuskin on syy-seuraussuhdetta, vaan kolmas muuttuja väestömäärä selittää avioerojen ja kuolemien määrien korrelaation.

Kuolemien ja avioerojen välinen korrelaatio on voimakas, muttei niin voimakas kuin kuolemien ja väestömäärän välinen korrelaatio. Tämä voi johtua esimerkiksi siitä, että joissain maissa on kulttuurisyyistä vähemmän avioeroja kuin toisissa.

Vastaus: Kuolemat ja avioerot $r \approx 0,91$, avioerot ja väestömäärä $r \approx 0,96$, kuolemat ja väestömäärä $r \approx 0,96$.

545. Kopioidaan taulukko sopivaan ohjelmaan ja analysoidaan aineistoa.

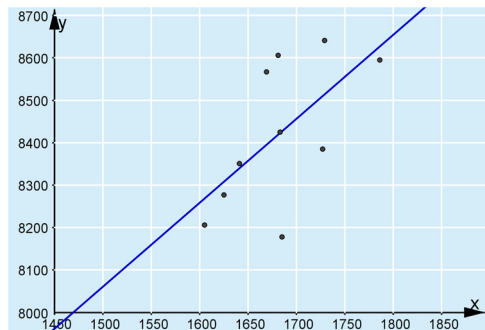
Määritetään
100 m juoksun sekä
kuulantäynnön
yhteispistemäärän ja
ottelun
kokonaispistemäärän
korrelaatiokerroin ja
regressiosuoran
yhtälö.



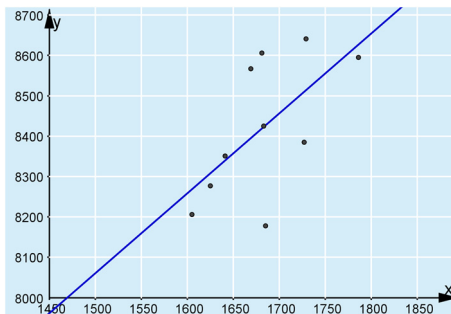
Korrelaatiokerroin on $r = 0,6193 \approx 0,619$ ja regressiosuoran yhtälö on $y = 1,9794x + 5091,489$.

Eduard Hämäläisen juoksun ja kuulantäynnön pistesumma oli $858 + 732 = 1590$. Hämäläisen odotettavissa olevaksi lopulliseksi kokonaispistemääräksi saadaan ohjelman avulla $8238,813... \approx 8239$.

Piirretään
korrelaatiodiagrammi.

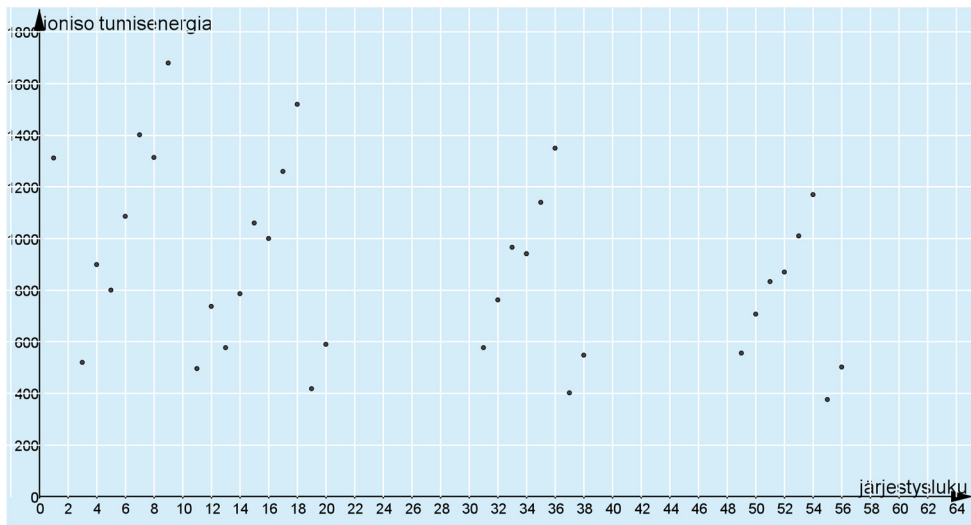


Vastaus: $r \approx 0,619$, $y = 1,9794x + 5091,489$, pistemäärä 8239,



546. Kopioidaan taulukko sopivaan ohjelmaan ja analysoidaan aineistoa.

- a) Piirretään hajontakuvio, jossa vaaka-akselilla on alkuaineen järjestysluku ja pystyakselilla ionisoitumisenergia.



Huomataan, että ionisoitumisenergian ja järjestysluvun riippuvuutta osoittavat pisteet ovat kahdeksassa pistejoukossa, joita voidaan kuvata kahdeksalla suoralla.

Vastaus: Ionisoitumisenergian ja järjestysluvun riippuvuutta voidaan kuvata kahdeksalla suoralla.

- b) Ohjelman avulla saadaan alkuaineiden 3–10 järjestysluvun ja ionisoitumisenergian väliseksi korrelaatiokertoimeksi $r = 0,96$ ja regressiosuoran yhtälöksi $y = 198,61x - 68,32$.

Vastaus: $r \approx 0,96$ ja $y = 198,61x - 68,32$

547. Kokeilemalla havaitaan, että esimerkiksi lihantuotannon y bruttokansantuotteen x riippuvuutta voi kuvata 6. asteen polynomisella mallilla

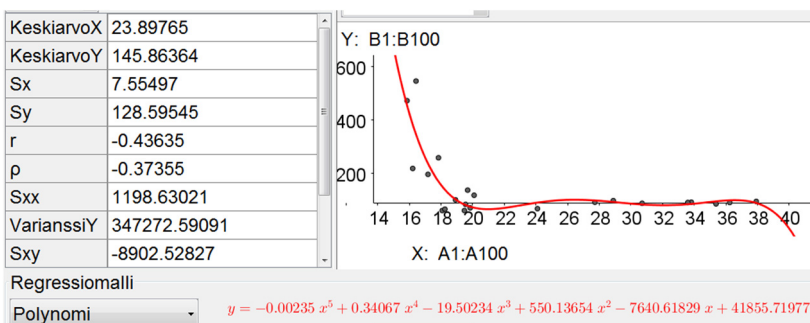
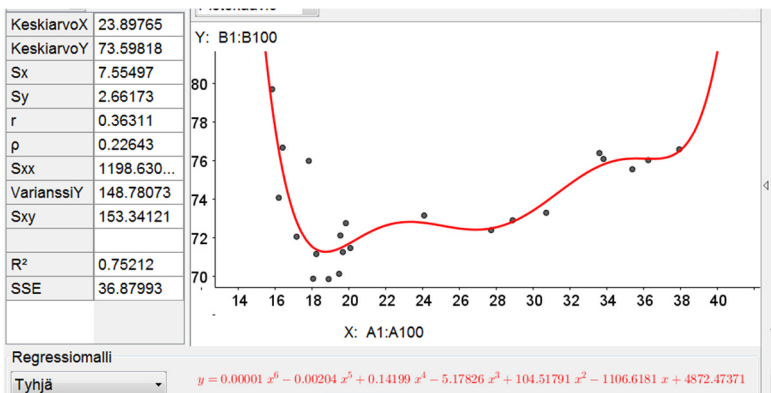
$$y = 0,00001x^6 - 0,00204x^5 + 0,142x^4 - 5,178x^3 + 104,5x^2 - 1106,6x + 4872.$$

Selitysaste on 75 %.

Lihajätteen y ja bruttokansantuotteen x riippuvuutta voi kuvata 5. asteen polynomisella mallilla

$$y = -0,00235x^5 + 0,341x^4 - 19,502x^3 + 550,14x^2 - 7640,6x + 41856$$

Selitysaste on 72 %.



Vastaus: esim. lihantuotanto y bruttokansantuotteen x funktiona

$$y = 0,00001x^6 - 0,00204x^5 + 0,142x^4 - 5,178x^3 + 104,5x^2 - 1106,6x + 4872,$$

selitysaste 75 %, lihajäte bruttokansantuotteen funktiona

$$y = -0,00235x^5 + 0,341x^4 - 19,502x^3 + 550,14x^2 - 7640,6x + 41856,$$

selitysaste 72 %

548. Kopioidaan taulukko sopivaan ohjelmaan ja analysoidaan aineistoa.

- a) Täydennetään taulukko. Lasketaan ensin muiden vuosikulujen kertymät rivillä 5.

	A	B	C	D	E	F
1						
2	Laitteen ikä (vuotta)	1	2	3	4	5
3	Arvon vähenemisen kertymä (€)	50000	70000	85000	90000	95000
4	Muut kulut vuodessa (€)	4000	8000	14000	40000	60000
5	Vuosikulujen kertymä (€)	4000	12000	=C5+D4		
6	Kokonaiskulujen kertymä (€)	54000	82000			
7	Kokonaiskulut ikävuotta kohti (€)	54000				

Seuraavaksi täydennetään kokonaiskulujen laskemalla yhteen arvon väheneminen ja muiden vuosikulujen kertymä.

	A	B	C	D	E	F
1						
2	Laitteen ikä (vuotta)	1	2	3	4	5
3	Arvon vähenemisen kertymä (€)	50000	70000	85000	90000	95000
4	Muut kulut vuodessa (€)	4000	8000	14000	40000	60000
5	Vuosikulujen kertymä (€)	4000	12000	26000	66000	126000
6	Kokonaiskulujen kertymä (€)	54000	82000	=D3+D5		
7	Kokonaiskulut ikävuotta kohti (€)	54000				

Seuraavaksi lasketaan kokonaiskulut ikävuotta kohti jakamalla kokonaiskulut ikävuosilla.

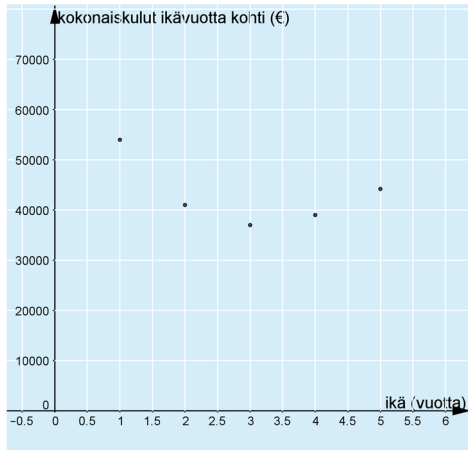
	A	B	C	D	E	F
1						
2	Laitteen ikä (vuotta)	1	2	3	4	5
3	Arvon vähenemisen kertymä (€)	50000	70000	85000	90000	95000
4	Muut kulut vuodessa (€)	4000	8000	14000	40000	60000
5	Vuosikulujen kertymä (€)	4000	12000	26000	66000	126000
6	Kokonaiskulujen kertymä (€)	54000	82000	111000	156000	221000
7	Kokonaiskulut ikävuotta kohti (€)	54000	=C6/C2			

Sovitetaan polynominen malli, joka ilmaisee laitteen kokonaiskulut ikävuotta kohti laitteen iän x funktiona.

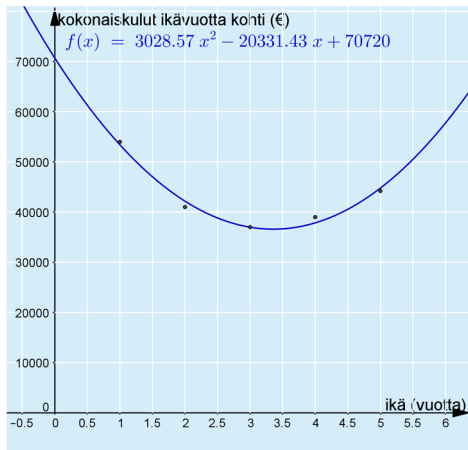
	A	B	C	D	E	F
1						
2	Laitteen ikä (vuotta)	1	2	3	4	5
3	Arvon vähenemisen kertymä (€)	50000	70000	85000	90000	95000
4	Muut kulut vuodessa (€)	4000	8000	14000	40000	60000
5	Vuosikulujen kertymä (€)	4000	12000	26000	66000	126000
6	Kokonaiskulujen kertymä (€)	54000	82000	111000	156000	221000
7	Kokonaiskulut ikävuotta kohti (€)	54000	41000	37000	39000	44200

Piirretään ensin
hajontakuvio.

Huomataan, että se
muistuttaa paraabelia,
joten sovitetaan
aineistoon 2. asteen
polynomimalli.



Malliksi saadaan $f(x) = 3028,57x^2 - 20\,331,43x + 70\,720$, jossa
funktion arvona ovat
laitteen kokonaiskulut
ikävuotta kohti iän x
funktiona.

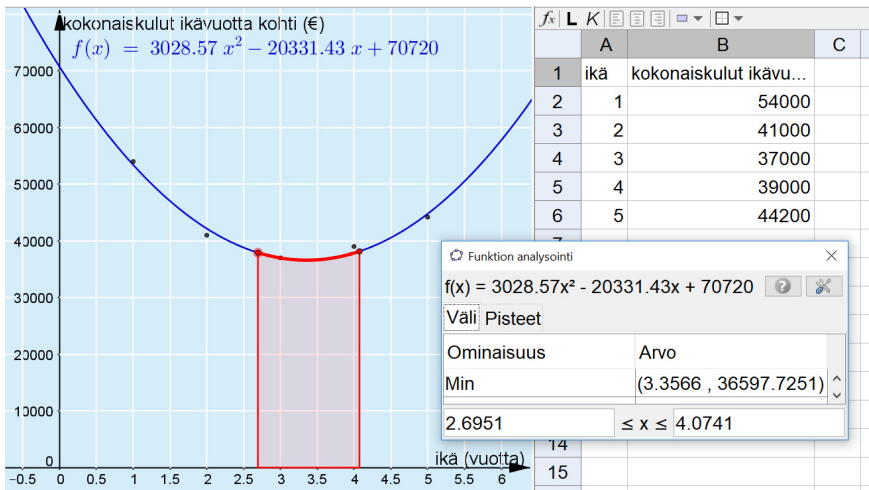


Vastaus:

Laitteen ikä (vuotta)	1	2	3	4	5
Arvon vähenemisen kertymä (€)	50 000	70 000	85 000	90 000	95 000
Muut vuosikulut (€)	4000	8000	14 000	40 000	60 000
Muiden vuosikulujen kertymä (€)	4000	12 000	26 000	66 000	126 000
Kokonaiskulujen kertymä (€)	54 000	82 000	111 000	156 000	221 000
Kokonaiskulut ikävuotta kohti (€)	54 000	41 000	37 000	39 000	44 200

$$f(x) = 3028,57x^2 - 20\,331,43x + 70\,720$$

- b) Laitteen vaihtaminen on edullisinta, kun laitteen kokonaiskulut ikävuotta kohti ovat pienimmillään. Määritetään ohjelman avulla millä muuttujan x arvolla funktio saa pienimmän arvonsa.



Funktio saa pienimmän arvonsa, kun $x = 3,3566$ vuotta eli
 $3 \text{ vuotta} + 0,3566 \cdot 12 \text{ kk} = 3 \text{ vuotta } 4,279... \text{ kk} \approx 3 \text{ vuotta } 4 \text{ kk}$.

Vastaus: 3 vuotta 4 kk

- b)** Määritetään oppilaiden eri kokeiden pisteiden välisen regressiosuoran yhtälö $y = b_1 \cdot x + b_0$, taulukon ja kaavojen avulla.

$$b_1 = \frac{\sum (y - \bar{y})(x - \bar{x})}{\sum (x - \bar{x})^2} = \frac{470}{730} = \frac{47}{73}$$

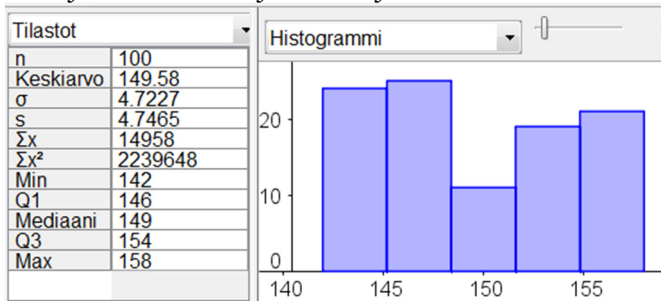
$$b_0 = \bar{y} - b_1 \cdot \bar{x} = 77 - \frac{47}{73} \cdot 78 = \frac{1955}{73}$$

Regressiosuoran yhtälö on $y = b_1 \cdot x + b_0 = \frac{47}{73}x + \frac{1955}{73}$

Vastaus: $y = \frac{47}{73}x + \frac{1955}{73}$

LOPETUSAUKEAAMAAN LIITTYVIÄ TEHTÄVIÄ

1. Kopioidaan taulukko sopivaan ohjelmaan ja määritetään jogurttipurkkien massojen vaihteluväli ja keskihajonta.



Havaitaan, että aineiston pienin arvo on 142 g ja suurin arvo 158 g, joten vaihteluväli on [142 g, 158 g].
Keskihajonta on $s = 4,7465 \text{ g} \approx 4,75 \text{ g}$

Vastaus: [142 g, 158 g], $s \approx 4,75 \text{ g}$

2. Pienimmän arvon ja keskiarvon erotus on $149,58 - 142 = 7,58$. Lasketaan erotuksen ja keskihajonnan osamäärä.

$$\frac{7,58}{4,746...} = 1,596... \approx 1,6$$
 eli pienin arvo poikkeaa keskiarvosta 1,6 keskihajonnan verran.

Suurimman arvon ja keskiarvon erotus on $158 - 149,58 = 8,42$.

Lasketaan erotuksen ja keskihajonnan osamäärä.

$$\frac{8,42}{4,746...} = 1,774... \approx 1,8$$
 eli suurin arvo poikkeaa keskiarvosta 1,8 keskihajonnan verran.

Kumpikaan ei poikkeaa vähintään kahta keskihajontaa keskiarvosta, joten poikkeamat eivät ole merkittäviä.

Vastaus: Pienin paino on 1,6 keskihajonnan päässä, suurin paino 1,8 keskihajonnan päässä. Ei poikkeaa.