

5 TILASTOAINEISTON ANALYSOINTIA

5.1 Hajontalukuja

ALOITA PERUSTEISTA

- 501.** a) Suurin aallonkorkeus on merkitty kuvaan kohtaan 24, ja se 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.

Jakaumassa C havaintoarvot ovat jakautuneet melko symmetrisesti keskimmäisen arvon ympärille, mutta eivät ole keskittyneet niin keskelle kuin jakaumassa A, joten keskihajonta on siinä jakaumien A ja B välissä.

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. a)** Järjestetään havaintoarvot suuruusjärjestykseen: 10, 15, 15 ja 20. Pienin havaintoarvo on 10 sekuntia ja suurin 20 sekuntia, joten vaihteluväli on [10, 20].

Vaihteluvälin pituus on suurimman ja pienimmän havaintoarvon erotus 20 sekuntia – 10 sekuntia = 10 sekuntia.

Vastaus: [10, 20], 10 s

- b)** Keskiarvo saadaan laskemalla havaintoarvot yhteen ja jakamalla havaintoarvojen lukumäärällä.

$$\bar{x} = \frac{10 + 15 + 15 + 20}{4} = 15$$

Viestin lukuaikojen keskiarvo on 15 sekuntia.

Vastaus: 15 s

- c)** Lasketaan havaintoarvojen keskihajonta.

$$\begin{aligned}s &= \sqrt{\frac{\sum (x - \bar{x})^2}{n}} \\&= \sqrt{\frac{(10 - 15)^2 + (15 - 15)^2 + (15 - 15)^2 + (20 - 15)^2}{4}} \\&= 3,535... \\&\approx 3,5\end{aligned}$$

Havaintoarvojen keskihajonta on noin 3,5 sekuntia.

Vastaus: 3,5 s

- d) Lasketaan merkittävän poikkeaman rajat lisäämällä ja vähentämällä keskiarvosta kaksinkertainen keskihajonta.

$$\text{yläraja: } 15 \text{ g} + 2 \cdot 3,5 \text{ g} = 22 \text{ g}$$

$$\text{alaraja: } 15 \text{ g} - 2 \cdot 3,5 \text{ g} = 8 \text{ g}$$

Kaikki havaintoarvot ovat rajojen sisäpuolella, joten yksikään havaintoarvo ei ole merkitsevästi poikkeava.

Vastaus: ei poikkeaa

505. Lasketaan merkittävän poikkeaman rajat lisäämällä ja vähentämällä keskiarvosta kaksinkertainen keskihajonta.

$$\text{yläraja: } 177,5 \text{ cm} + 2 \cdot 10,6 \text{ cm} = 198,7 \text{ cm}$$

$$\text{alaraja: } 177,5 \text{ cm} - 2 \cdot 10,6 \text{ cm} = 156,3 \text{ cm}$$

Havaintoarvojen suurin arvo 201 cm on rajojen ulkopuolella, joten se on merkitsevästi poikkeava havaintoarvo.

Vastaus: on, pidempi

506. Lasketaan ensin kahvintuotanto kilogrammoina.

Vuosi	2015	2016	2017	2018
Tuotanto (60 kg:n säkkeinä)	53 850	55 776	55 838	63 098
Tuotanto (kg)	$60 \cdot 53\,850 =$ 3 231 000	$60 \cdot 55\,776 =$ 3 346 560	$60 \cdot 55\,838 =$ 3 350 280	$60 \cdot 63\,098 =$ 3 785 880

Tuotannon keskiarvo kilogrammoina

$$\bar{x} = \frac{3\,231\,000 + 3\,346\,560 + 3\,350\,280 + 3\,785\,880}{4} = 3\,428\,430 \approx 3\,430\,000.$$

Keskiarvo kilogrammoina on noin 3 430 000 kg.

Keskihajonta kilogrammoina

$$\begin{aligned}s &= \sqrt{\frac{(3\,231\,000 - \bar{x})^2 + (3\,346\,560 - \bar{x})^2 + (3\,350\,200 - \bar{x})^2 + (3\,785\,880 - \bar{x})^2}{4}} \\&= 211\,872,147... \\&\approx 212\,000\end{aligned}$$

Keskihajonta kilogrammoina on noin 212 000 kg.

Vastaus: $\bar{x} \approx 3\,430\,000$ kg, $s \approx 212\,000$ kg

- 507.** Määritetään ohjelman avulla patenttien vuosittaisen määrän keskiarvo, vaihteluväli ja keskihajonta.

	A	B	C
1	Vuosi	Patenttien määrä	
2	2000	1939	
3	2001	2047	
4	2002	2056	
5	2003	2402	
6	2004	2075	
7	2005	1757	
8	2006	1059	
9	2007	921	
10	2008	998	
11	2009	1055	
12	2010	923	
13	2011	841	
14	2012	835	
15	2013	711	
16	2014	786	
17	2015	931	
18	2016	815	
19	2017	704	
20	2018	533	
21	2019	505	
22	keskiarvo	1194,65	=KESKIARVO(B2:B21)
23	keskihajonta	583,94	=KESKIHAJONTA.P(B2:B21)
24	suurin arvo	2402,00	=MAKS(B2:B21)
25	pienin arvo	505,00	=MIN(B2:B21)

Ohjelman tilastotoiminnoilla saadaan patenttien määrän vuosittaiseksi keskiarvoksi noin 1195 ja keskihajonnaksi noin 584. Ohjelman toiminnolla tai taulukosta saadaan suurimmaksi arvoksi 2402 ja pienimmäksi 505, joten vaihteluväli on [505, 2402].

Vastaus: $\bar{x} \approx 1195$, vaihteluväli [505, 2402], $s \approx 584$

- 508.** 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

$$\begin{aligned}s &= \sqrt{\frac{3 \cdot (5,5 - \bar{x})^2 + 5 \cdot (15,5 - \bar{x})^2 + \dots + 240 \cdot (55,5 - \bar{x})^2}{300}} \\ &= 9,281... \\ &\approx 9,3.\end{aligned}$$

Vastaus: $s \approx 9,3$ pistettä

- 509.** Määritetään suurin ja pienin arvo sekä keskihajonta ohjelman avulla. Ajat ovat sekunteina 65, 60, 57, 62, 67, 64, 66, 63, 55, 68, 66, 66 ja 65.

n	13
Keskiarvo	63.38462
σ	3.77315
s	3.92722
Σx	824
Σx^2	52414
Min	55
Q1	61
Mediaani	65
Q3	66
Max	68

- a) Pienin havaintoarvo on 55 s ja suurin 68 s, joten vaihteluväli on $68 \text{ s} - 55 \text{ s} = 13 \text{ sekuntia}$.

Väite on epätosi.

Vastaus: epätosi, 13

- b) Keskihajonnaksi saadaan ohjelmalla $3,773... \text{ s} \approx 3,8 \text{ s}$.

Väite on epätosi.

Vastaus: epätosi, 3,8

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

Keskiarvoksi saadaan ohjelmalla $63,384... \text{ s}$. Merkittävän poikkeaman alaraja on $63,384... \text{ s} - 2 \cdot 3,773... \text{ s} = 55,838... \text{ s}$.

Kauden paras tulos 55 s alittaa alarajan, joten se poikkeaa merkittävästi muista tuloksista.

Väite on tosi.

Vastaus: tosi

- d) Kauden huonoin tulos on 68 s. Jotta se poikkeaisi merkittävästi muista tuloksista, on sen oltava yli kahden keskihajonnan päässä keskiarvosta.

Merkittävän poikkeaman yläraja on
 $63,384... s + 2 \cdot 3,773... s = 70,930... s$.

Kauden huonoin tulos 68 s ei ylitä ylärajaa, joten se ei poikkea merkittävästi muista tuloksista.

Väite on epätosi.

Vastaus: epätosi, ei poikkea

VAHVISTA OSAAMISTA

510. a) Määritetään keskiarvo ja keskihajonta taulukkolaskentaohjelman avulla.

	A	B	C
1	Jokerit	9173	
2	HIFK	8266	
3	Kärpät	5821	
4	TPS	5534	
5	Tappara	5359	
6	Ilves	5177	
7			
8	keskiarvo	6555,0000	'=KESKIARVO(B1:B6)
9	keskihajonta	1564,8185	'=KESKIHAJONTA.P(B1:B6)

Keskiarvo on 6555.

Keskihajonta on $1564,818... \approx 1565$.

Vastaus: $\bar{x} = 6555$, $s \approx 1565$

- b) Poikkeama keskiarvosta ylöspäin saadaan keskiarvon ja keskihajonnan summan avulla

$$6555 + 1564,818... = 8119,818...$$

Jokereiden ja HIFK:n katsojaluvut ovat tätä suurempi.

Poikkeama keskiarvosta alaspäin saadaan keskiarvon ja keskihajonnan erotuksen avulla

$$6555 - 1564,818... = 4990,181...$$

Yhdenkään joukkueen katsojaluvut eivät ole tätä pienempiä.

Vastaus: Jokereiden ja HIFK:n

- 511.** 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

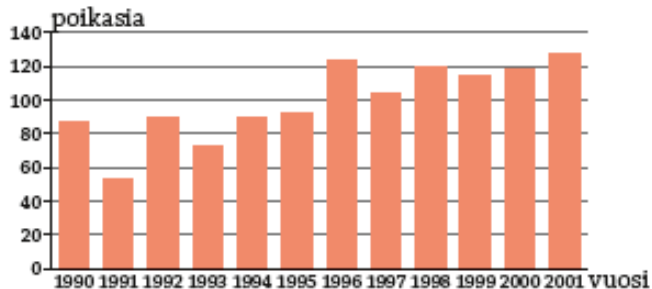
- c) Jamin tulosten keskihajonta on 0,937 m ja Jimin 4,851 m, joten Jami heittää tasaisemmin.

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.

- 512.** Piirretään pylväsdiagrammi.



Määritetään poikasten määrän keskiarvo ja keskihajonta taulukkolaskentaohjelman avulla.

	A	B	C
1	Vuosi	Poikasia	
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	99	'=KESKIARVO(B2:B13)
16	keskihajonta	21,5832	'=KESKIHAJONTA.P(B2:B13)

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

Keskihajonta on $s = 21,583... \approx 22$.

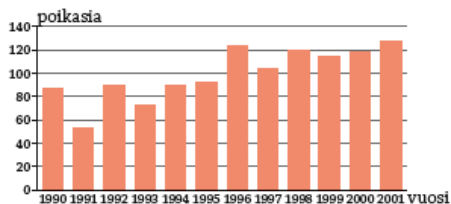
Poikasmäärä poikkeaa merkitsevästi ylöspäin keskiarvosta, jos se on vähintään $99 + 2 \cdot 21,583... = 142,166...$

Vuoden 2001 poikasmäärä ei ole näin suuri, joten se ei poikkeaa merkitsevästi keskiarvosta.

Poikasmäärä poikkeaa merkitsevästi alaspäin keskiarvosta, jos se on korkeintaan $99 - 2 \cdot 21,583... = 55,833...$

Vuoden 1991 poikasmäärä alittaa tämän, joten poikkeama on merkittävä.

Vastaus:



$\bar{x} = 99$, $s \approx 22$,

vuoden 1991 poikkeaa, vuoden 2001 ei poikkeaa

- 513. a)** Määritetään jogurttipurkkien massojen keskiarvo, vaihteluväli ja keskihajonta.

n	100
Keskiarvo	149.58
σ	4.72267
s	4.74646
Σx	14958
Σx^2	2239648
Min	142
Q1	146
Mediaani	149
Q3	154
Max	158

Massojen keskiarvo on 149,59 g $\approx$ 150 g.

Aineiston pienin arvo on 142 g ja suurin arvo 158 g, joten vaihteluväli on [142 g, 158 g].

Keskihajonta on $s = 4,722... \approx 4,72$ g

Vastaus: $\bar{x} \approx 150$ g, [142 g, 158 g], $s \approx 4,72$ g

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

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

1,6 keskihajonnan verran.

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

Lasketaan erotuksen ja keskihajonnan osamäärä.

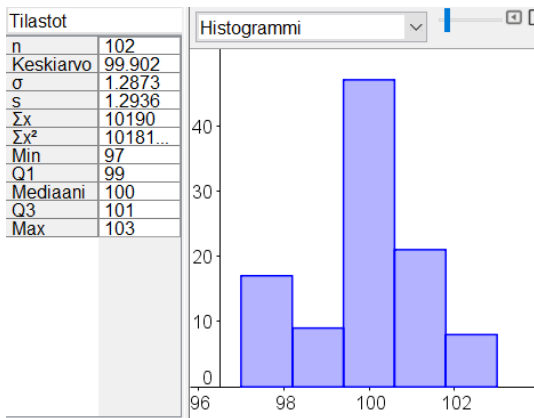
$$\frac{8,42}{4,746...} = 1,774... \approx 1,8 \text{ eli suurin arvo poikkeaa keskiarvosta}$$

1,8 keskihajonnan verran.

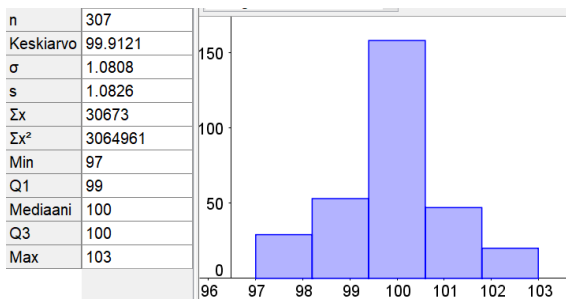
Kumpikaan ei poikkeaa vähintään kahta keskihajontaa keskiarvosta, joten poikkeamat eivät ole merkittäviä.

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

514. Määritetään ohjelman avulla koneen A pussien massojen tunnusluvut GeoGebralla.



Määritetään koneen B tunnusluvut GeoGebralla.



- 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 $99,9 - 2 \cdot 1,08 = 97,74$

Koska molemmilla koneilla pakkauksen massa 103 g on rajojen ulkopuolella, voidaan sanoa, että molempien koneiden pakkaamat 103 g:n pakkaukset poikkeavat merkittävästi keskiarvosta. Vastaavasti myös koneiden pakkaamat kevyimmät 97 g:n pakkaukset poikkeavat merkittävästi keskiarvosta.

Vastaus: poikkeavat

515. Määritetään kasvisliemikuutioiden massojen keskiarvo ja otoskeskihajonta GeoGebralla.

n	145
Keskiarvo	10.01724
σ	0.36395
s	0.36521
Σx	1452.5
Σx^2	14569.25
Min	9
Q1	10
Mediaani	10
Q3	10
Max	11

Havaintoaineiston keskiarvo on $\bar{x} = 10,017\dots$ g ja otoskeskihajonta $s_{n-1} = 0,365\dots$ g.

Otoksen perusteella tuotantolinjan kasvisliemikuutioiden massojen keskiarvo on $10,017\dots$ g $\approx 10,0$ g ja keskihajonta on $0,365\dots$ g $\approx 0,4$ g.

Vastaus: 10,0 g ja 0,4 g

- 516.** a) Määritetään tilavuuden keskiarvo ja keskihajonta taulukkolaskentaohjelman avulla.

	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	keskiarvo	87,975
23	otoskeskihajonta	4,181522639

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

Otoskeskihajonta on $s_{n-1} = 4,181 \dots \text{ cl} \approx 4,2 \text{ cl}$.

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

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

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

517. Lasketaan jakauman keskiarvo ja keskihajonta GeoGebralla käyttämällä
 suhteellisia osuuksia frekvensseinä.

	A	B
1	Puoltoäännet	Osuus (%)
2	7	2.8
3	6	9.1
4	5	18.8
5	4	28
6	3	22.9
7	2	13.8
8	0	4.6

n	100
Keskiarvo	3.765
σ	1.50525
s	1.51283
Σx	376.5
Σx^2	1644.1
Min	0
Q1	3
Mediaani	4
Q3	5
Max	7

Keskiarvo on $\bar{x} = 3,765 \dots \approx 3,8$.

Keskihajonta on $s = 1,505 \dots \approx 1,5$.

Vastaus: $\bar{x} \approx 3,8$ puoltoääntä, $s \approx 1,5$ puoltoääntä

518. a) Lasketaan kaikkina kuukausina syntyneiden lasten määrän keskiarvo taulukkolaskentaohjelman avulla.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Vuosi	Koko vuosi	Tammikuu	Helmikuu	Maaliskuu	Huhtikuu	Toukokuu	Kesäkuu	Heinäkuu	Elokuu	Syyskuu	Lokakuu	Marraskuu	Joulukuu
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	2016	52814	4416	4094	4533	4458	4622	4507	4619	4727	4604	4411	3984	3839
9	2017	50321	4152	3814	4180	4062	4338	4421	4615	4559	4204	4275	3878	3823
10	2018	47577	4088	3657	3974	3989	4113	4051	4222	4351	3944	3915	3719	3554
11	2019	45613	3867	3288	3778	3750	3735	3817	4128	4122	4052	3848	3649	3579
12														
13	Kaikkien kuukausien keskiarvo													
14	4563,3083'=KESKIARVO(C2:N11)													

Keskiarvoksi saadaan noin 4563 syntymää kuukaudessa.

Vastaus: 4563 lasta

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

	A	B	C	D	E
1	Vuosi	Koko vuosi	Tammikuu	Helmikuu	Maaliskuu
2	2010	60980	5007	4733	5297
3	2011	59961	5070	4592	5072
4	2012	59493	4991	4722	5061
5	2013	58134	4837	4463	4970
6	2014	57232	4800	4333	4806
7	2015	55472	4659	4209	4603
8	2016	52814	4416	4094	4533
9	2017	50321	4152	3814	4180
10	2018	47577	4088	3657	3974
11	2019	45613	3867	3288	3778
12			4588,7'=KESKIARVO(C2:C11)		

Kopioidaan keskiarvon laskukaava muiden kuukausien kohdalle.

C	D	E	F	G	H	I	J	K	L	M	N
Tammikuu	Helmikuu	Maaliskuu	Huhtikuu	Toukokuu	Kesäkuu	Heinäkuu	Elokuu	Syyskuu	Lokakuu	Marraskuu	Joulukuu
5007	4733	5297	4956	5095	5111	5441	5272	5280	5251	4716	4821
5070	4592	5072	4790	5091	5057	5309	5434	5274	4888	4803	4581
4991	4722	5061	4822	5003	5047	5384	5388	4990	4832	4769	4484
4837	4463	4970	4840	5014	4866	5247	5134	5025	4987	4400	4351
4800	4333	4806	4683	4997	4880	5227	5052	5065	4781	4232	4376
4659	4209	4603	4629	4633	4696	5010	4882	4756	4615	4500	4280
4416	4094	4533	4458	4622	4507	4619	4727	4604	4411	3984	3839
4152	3814	4180	4062	4338	4421	4615	4559	4204	4275	3878	3823
4088	3657	3974	3989	4113	4051	4222	4351	3944	3915	3719	3554
3867	3288	3778	3750	3735	3817	4128	4122	4052	3848	3649	3579
4588,7	4190,5	4627,4	4497,9	4664,1	4645,3	4920,2	4892,1	4719,4	4580,3	4265	4168,8

Vastaus:

Kuukausi	Syntyneitä keskimäärin
tammikuu	4589
helmikuu	4191
maaliskuu	4627
huhtikuu	4450
toukokuu	4664
kesäkuu	4645
heinäkuu	4920
elokuu	4892
syyskuu	4719
lokakuu	4580
marraskuu	4265
joulukuu	4169

- c) Suurin syntyneiden määrän keskiarvo on heinäkuussa, 4920.
 Sen poikkeama keskiarvosta on $4920 - 4563 = 357$.

Pienin syntyneiden määrän keskiarvo on joulukuussa, 4169.
 Sen poikkeama keskiarvosta on $4169 - 4563 = -394$.

Lasketaan kaikkien tutkittujen kuukausien syntyneiden määrän keskihajonta taulukkolaskentaohjelman avulla.

	A	B	C	D
1	Vuosi	Koko vuosi	Tammikuu	Helmikuu
2	2010	60980	5007	4733
3	2011	59961	5070	4592
4	2012	59493	4991	4722
5	2013	58134	4837	4463
6	2014	57232	4800	4333
7	2015	55472	4659	4209
8	2016	52814	4416	4094
9	2017	50321	4152	3814
10	2018	47577	4088	3657
11	2019	45613	3867	3288
12			4588,7	4190,5
13	Kaikkien kuukausien keskiarvo			
14	4563,3083'=KESKIARVO(C2:N11)			
15	Kaikkien kuukausien keskihajonta			
16	497,3201'=KESKIHAJONTA.P(C2:N11)			

Keskihajonnaksi saadaan 497,320...

Syntymien määrä, joka on kaksi keskihajontaa keskiarvoa suurempi,
 on $4563,308... + 2 \cdot 497,320... = 5557,948...$

Syntymien määrä, joka on kaksi keskihajontaa keskiarvoa pienempi,
 on $4563,308... - 2 \cdot 497,320... = 3568,668...$

Yksikään keskiarvoista ei ylitä arvoa 5557,948... eikä alita arvoa 3568,668..., joten poikkeamat eivät ole merkittäviä.

Vastaus: ei poikkea

- 519. a)** 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 $\bar{x} = 6,760... \text{ €} \approx 6,76 \text{ €}$.

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

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

520. 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: $a = 0$

Kerro vakiolla: $b = 10$

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.

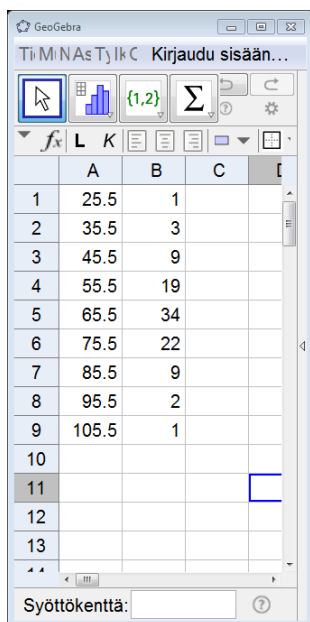
Vastaus: ei muutu

- 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.

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

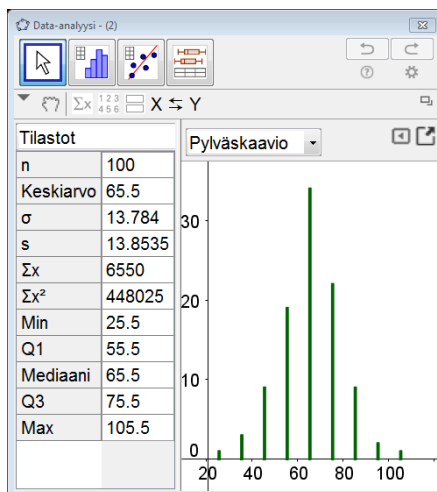
- 522.** Puhelinten käyttöikien moodi on 61–70 kk, koska sen frekvenssi on suurin. Muiden sijaintilukujen määrittämiseksi lasketaan luokkakeskukset ja kirjoitetaan frekvenssit ohjelmaan. Ensimmäisen luokan 21–30 todellinen alaraja on 20,5 ja todellinen yläraja 30,5, joten luokkakeskus on $\frac{20,5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

Lasketaan kysytyt tunnusluvut GeoGebralla.



	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:

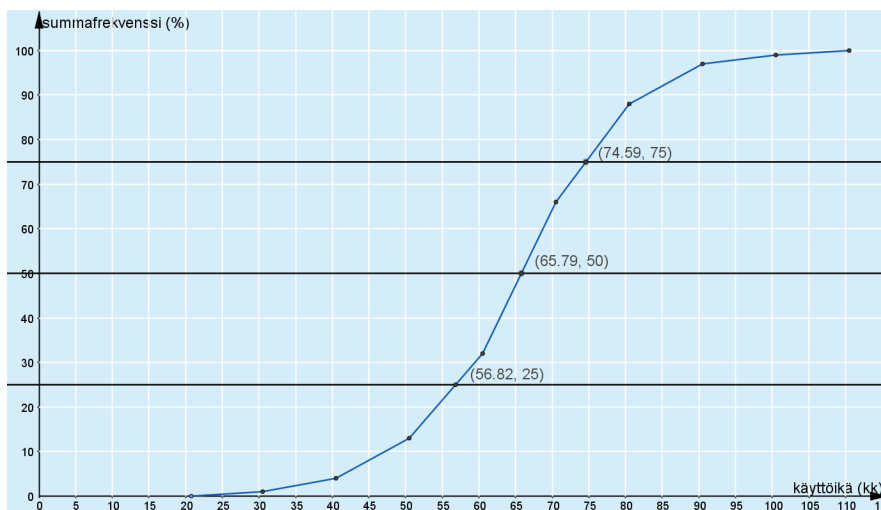


Ohjelma antaa alakvartiiliksi $Q_1 = 55,5$ kk, mediaaniksi $Q_2 = 65,5$ kk ja yläkvartiiliksi $Q_3 = 75,5$ kk sekä keskiarvoksi 65,5 kk ja keskihajonnaksi $s = 13,784 \dots$ kk $\approx 13,8$ kk

Lasketaan suhteelliset summafrekvenssit mediaanin ja kvartiilien määrittämiseksi tarkemmin.

Käyttöikä (kk)	Frekvenssi	Suhteellinen frekvenssi	Suhteellinen summafrekvenssi	Piste (x, y)
21-30	1	1,00 %	1,00 %	(30,5; 1)
31-40	3	3,00 %	4,00 %	(40,5; 4)
41-50	9	9,00 %	13,00 %	(50,5; 13)
51-60	19	19,00 %	32,00 %	(60,5; 32)
61-70	34	34,00 %	66,00 %	(70,5; 66)
71-80	22	22,00 %	88,00 %	(80,5; 88)
81-90	9	9,00 %	97,00 %	(90,5; 97)
91-100	2	2,00 %	99,00 %	(100,5; 99)
101-110	1	1,00 %	100,00 %	(110,5; 100)
	100			

Piirretään summafrekvenssin kuvaaja.



Kuvaajasta nähdään, että $Q_1 \approx 56,8$ kk, $Q_2 \approx 66,8$ kk ja $Q_3 \approx 74,6$ kk.

Vastaus: $Mo = 61-70$ kk, $Q_1 \approx 56,8$ kk, $Q_2 \approx 66,8$ kk, $Q_3 \approx 74,6$ kk,
 $\bar{x} = 65,5$ kk, $s \approx 13,8$ kk

- 523. a)** Naisia on ikäluokassa 70–74 eniten, joten naisten moodi on 70–74 vuotta. Vastaavasti miesten moodi on 35–39 vuotta. Muiden tunnuslukujen määrittämiseksi lasketaan luokkakeskukset ja kirjoitetaan frekvenssit 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	130884	125040
3	7.5	156654	149633
4	12.5	157712	151113
5	17.5	152230	144441
6	22.5	161679	152265
7	27.5	183092	172593
8	32.5	181115	169653
9	37.5	186122	174660
10	42.5	177928	168547
11	47.5	159982	154391
12	52.5	179182	176612
13	57.5	183719	185152
14	62.5	176283	183336
15	67.5	171275	185685
16	72.5	163697	186034
17	77.5	93987	118190
18	82.5	65140	96256
19	87.5	33733	63133
20	92.5	11789	31652
21	97.5	1907	7888
22	102.5	152	756

Määritetään naisten ikäjakauman tunnusluvut.

Naisten iän jakaumassa

$Md = 42,5$ vuotta

$\bar{x} \approx 44,8$ vuotta

$Q_1 = 22,5$ vuotta

$Q_3 = 67,5$ vuotta

$s \approx 24,6$ vuotta

n	2801227
Keskiarvo	44.7539
σ	24.5891
s	24.5891
Σx	125365857.5
Σx^2	7304305768.75
Min	2.5
Q1	22.5
Mediaani	42.5
Q3	67.5
Max	97.5

Vastaavasti saadaan miesten jakauman tunnusluvut.

Miesten iän jakaumassa

$Md = 42,5$ vuotta

$\bar{x} \approx 42,1$ vuotta

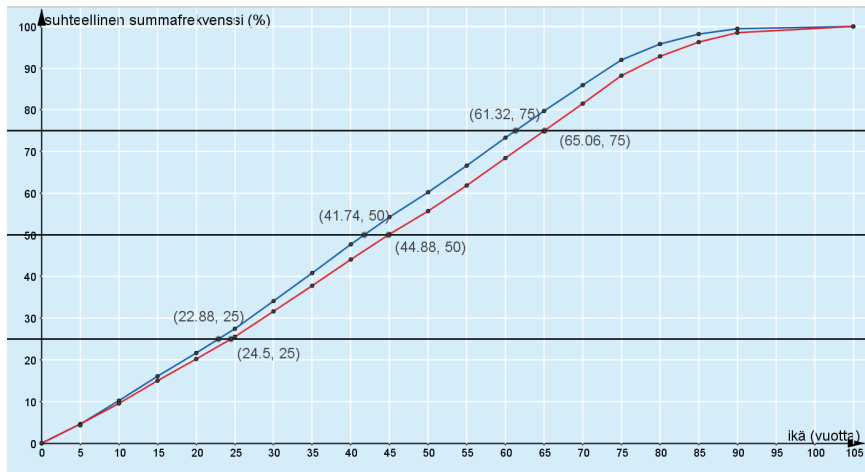
$Q_1 = 22,5$ vuotta

$Q_3 = 62,5$ vuotta

$s \approx 23,4$ vuotta

n	2735819
Keskiarvo	42.1212
σ	23.3614
s	23.3614
Σx	115235867.5
Σx^2	6346950718.75
Min	2.5
Q1	22.5
Mediaani	42.5
Q3	62.5
Max	97.5

Piirretään suhteellisten summafrekvenssien kuvaajat mediaanin ja kvartiilien määrittämiseksi tarkemmin.



Kuvaajasta nähdään kvartiilit:

Miehet: $Q_1 \approx 22,9$ vuotta, $Q_2 \approx 41,7$ vuotta ja $Q_3 \approx 61,3$ vuotta

Naiset: $Q_1 \approx 24,5$ vuotta, $Q_2 \approx 44,9$ vuotta ja $Q_3 \approx 65,1$ vuotta

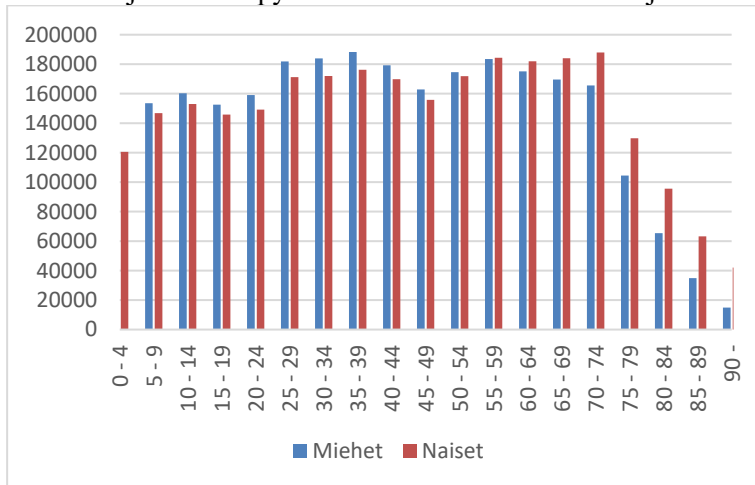
Vastaus: Naiset: $Mo = 70-74$ vuotta $Md \approx 44,9$ vuotta,

$Q_1 = 24,5$ vuotta, $Q_3 = 65,1$ vuotta, $\bar{x} \approx 44,8$ vuotta, $s \approx 24,6$ vuotta,

miehet: $Mo = 35-39$ vuotta, $Md \approx 41,7$ vuotta, $Q_1 = 22,9$ vuotta,

$Q_3 = 61,3$ vuotta, $\bar{x} \approx 42,1$ vuotta, $s \approx 23,4$ vuotta

b) Piirretään jakaumien pylväskaavio taulukkolaskentaohjelmalla.



- 524.** a) Määritetään Anteron ja Laurin tulosten keskiarvo ja keskihajonta GeoGebralla.

Antero

n	13
Keskiarvo	633.07692
σ	12.73675
s	13.25683
Σx	8230
Σx^2	5212332
Min	602
Q1	629
Mediaani	633
Q3	643
Max	657

Lauri

n	13
Keskiarvo	631.23077
σ	37.03461
s	38.54684
Σx	8206
Σx^2	5197710
Min	560
Q1	598
Mediaani	634
Q3	670.5
Max	682

Anteron tulokeskiarvo on noin 633 cm ja Laurin 631 cm.

Anteron tulosten keskihajonta on noin 13 cm ja Laurin noin 37 cm.

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

Lauri: $\bar{x} \approx 631$ cm, $s \approx 37$ 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

525. Esimerkiksi jakauma

x	Frekvenssi
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
Frekvenssi	5	10	5

526. Merkitään luokkien ylärajat taulukkoon ja luokitellaan aineisto niiden mukaan TAAJUUS-komennon avulla.

D	E	F	G	H	I
Sikiäimen pituus (cm)	Luokan yläraja	Frekvenssi	Kypsän kurkun pituus (cm)	Luokan yläraja	Frekvenssi
1,0–1,5	1,5	2	12–16	16	10
1,5–2,0	2	10	16–20	20	19
2,0–2,5	2,5	31	20–24	24	36
2,5–3,0	3	28	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.

2,9979	24,8123'=KESKIARVO(B2:B137)
0,8115	5,9566'=KESKIHJONTA.P(B2:B137)

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

Kypsä kurkku: keskiarvo $\bar{x} \approx 24,8$ cm ja keskihajonta $s \approx 5,96$ 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 $\frac{0,95 + 1,45}{2} = 1,2$.

	A	B
1	Luokkakeskus	Frekvenssi
2	1.2	2
3	1.7	10
4	2.2	31
5	2.7	28
6	3.2	27
7	3.7	23
8	4.2	8
9	4.7	4
10	5.2	1
11	5.7	1

n	135
Keskiarvo	2.94815
σ	0.82971
s	0.8328
Σx	398
Σx^2	1266.3
Min	1.2
Q1	2.2
Mediaani	2.7
Q3	3.7
Max	5.7

Sikiäimen keskiarvo $\bar{x} \approx 2,95$ cm ja keskihajonta $s \approx 0,83$ cm

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

	A	B
1	Luokkakeskus	Frekvenssi
2	14	10
3	18	19
4	22	36
5	26	31
6	30	23
7	34	10
8	38	5
9	42	1

n	135
Keskiarvo	24.75556
σ	6.09002
s	6.11271
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 $s \approx 6,09$ 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, $s \approx 5,96$ cm

Luokiteltu aineisto

sikiäin: $\bar{x} \approx 2,95$ cm, $s \approx 0,83$ cm

kypsä kurkku: $\bar{x} \approx 24,8$ cm, $s \approx 6,09$ cm,

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

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

$$\frac{x + 4x + 3 + 5x + 1 + 4 - 2x - 6x}{5} = 1$$

$$\frac{2x + 8}{5} = 1 \quad \parallel \cdot 5$$

$$2x + 8 = 5$$

$$2x = 5 - 8$$

$$2x = -3 \quad \parallel : 2$$

$$x = -1,5$$

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 ohjelmalla.

Tilastot	Σx	123 345	
n	5		
Keskiarvo	1		
σ	5.9749		
s	6.6802		
Σx	5		
Σx^2	183.5		
Min	-6.5		
Q1	-4.75		
Mediaani	-1.5		
Q3	8		
Max	9		

Keskihajonta on $s = 5,974... \approx 6,0$.

Vastaus: $s \approx 6,0$

528. Määritetään lukumääräarvioille keskiarvot ja otoskeskihajonnat.

Aineisto luokiteltuna			
Lukumäärän 5 arviot	<i>f</i>	Keskiarvo	Keskihajonta
5	102		
6	4		
7	1		
		5,1	0,27
Lukumäärän 6 arviot	<i>f</i>		
5	7		
6	120		
7	20		
		6,09	0,42
Lukumäärän 7 arviot	<i>f</i>		
6	18		
7	113		
8	25		
		7,04	0,52
Lukumäärän 8 arviot	<i>f</i>		
7	30		
8	76		
9	28		
10	1		
		8	0,68
Lukumäärän 9 arviot	<i>f</i>		
7	2		
8	24		
9	76		
10	18		
11	2		
		8,95	0,69
Lukumäärän 10 arviot	<i>f</i>		
8	6		
9	37		
10	46		
11	16		
12	2		
		9,73	0,85
Lukumäärän 11 arviot	<i>f</i>		
8	1		
9	11		
10	19		
11	26		
12	12		
		10,54	1,01

Lukumäärän 12 arviot	<i>f</i>		
9	1		
10	4		
11	17		
12	19		
13	3		
14	1		
		11,49	0,92
Lukumäärän 13 arviot	<i>f</i>		
11	7		
12	11		
13	6		
14	1		
15	1		
		12,15	1,01
Lukumäärän 14 arviot	<i>f</i>		
11	2		
12	3		
13	3		
14	4		
15	2		
		13,07	1,33
Lukumäärän 15 arviot	<i>f</i>		
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

- 529.** Kopioidaan taulukko ohjelmaan. Määritetään ohjelman avulla koneiden A ja B pussien keskiarvot ja otoskeskihajonnat.

	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

- 530.** 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

- 531.** 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

- 532.** 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: $0,123$.

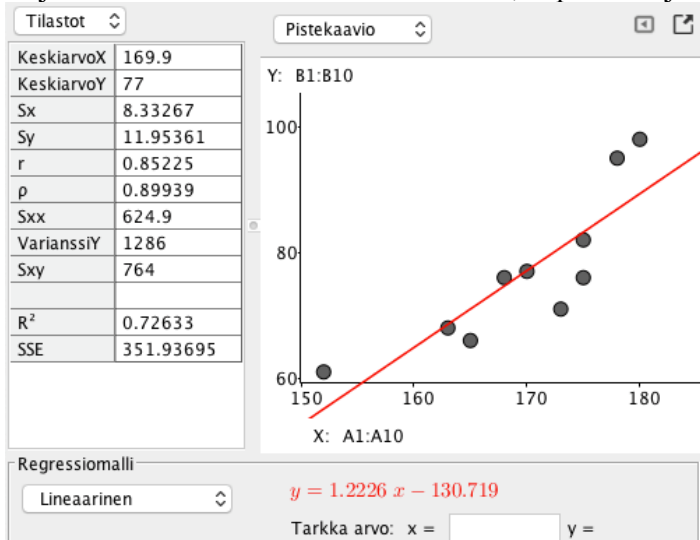
Kuviosta III havaitaan, että kyseessä on voimakas positiivinen korrelaatio, joten siihen sopii korrelaatiokerroin B: $0,995$.

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

- 533.** 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

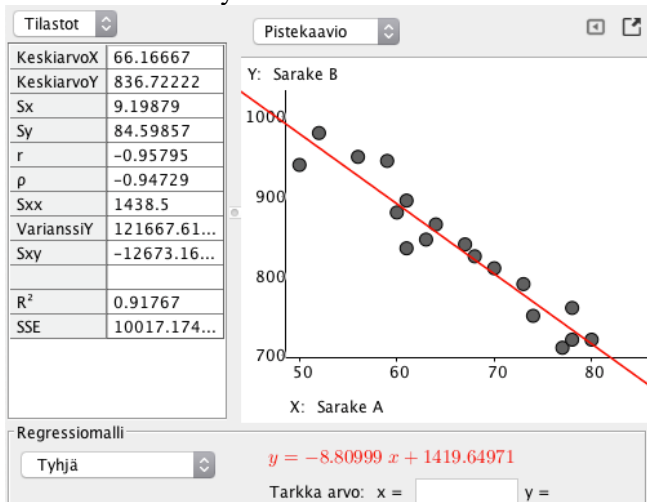
534. Ohjelma antaa korrelaatiokertoimeksi $r \approx 0,85$ pituuden ja massan välille.



Pituuden ja massan välillä voimakas positiivinen korrelaatio eli mitä pidempi henkilö on, sitä suurempi on hänen massansa.

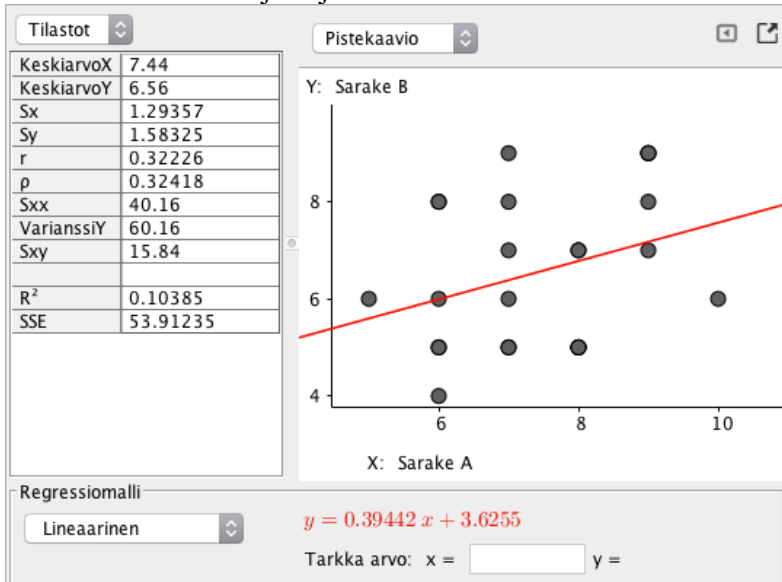
Vastaus: $r \approx 0,85$, voimakas positiivinen korrelaatio

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



Vastaus: n. 92 %

536. Määritetään taulukon ja ohjelman avulla tilastolliset tunnusluvut.



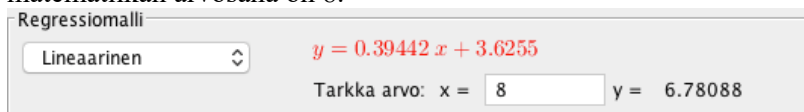
- a) Matematiikan ja ruotsin kielen arvosanojen riippuvuutta kuvaavan regressiosuoran yhtälö on $y = 0,39x + 3,63$, jossa x on matematiikan arvosana ja y on ruotsin kielen arvosana.

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

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

Vastaus: 0,32, kohtalainen positiivinen

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



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

Vastaus: arvosana 7

- d) Sijoitetaan ruotsin kielen arvosana 7 regressiosuoran yhtälöön $y = 0,39x + 3,63$ muuttujan y paikalle ja ratkaistaan yhtälöstä matematiikan arvosana x .

$$0,39x + 3,63 = 7$$

$$0,39x = 3,37 \quad \| :0,39$$

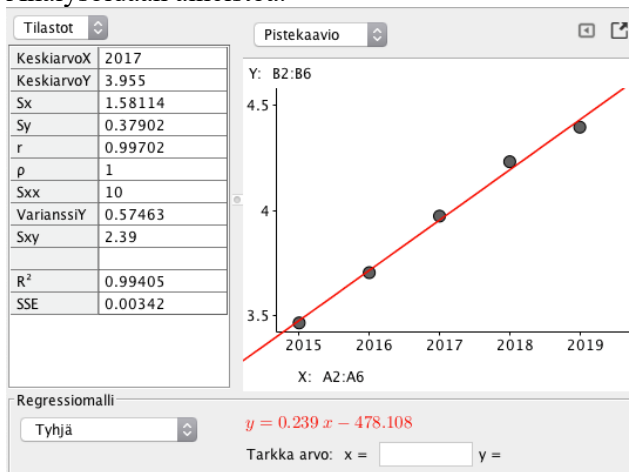
$$x = 8,641...$$

$$x \approx 9$$

Mallin mukaan matematiikan arvosana on 9, jos ruotsin kielen arvosana on 7.

Vastaus: arvosana 9

537. Analysoidaan aineistoa.



- a) Vuosiluvun ja matkustajamäärän välinen korrelaatiokerroin $r = 0,9970... \approx 0,997$. Korrelaatio on voimakas ja positiivinen.

Vastaus: $r \approx 0,997$, voimakas positiivinen

- b) Vuosiluvun ja matkustajamäärän riippuvuutta kuvaavan regressiosuoran yhtälö on $y = 0,239x - 478,108$, jossa x on vuosiluku ja y on lentomatkustajien määrä miljardeina.

Vastaus: $y = 0,239x - 478,108$, jossa x on vuosiluku ja y on lentomatkustajien määrä miljardeina

- c) Lasketaan ohjelman avulla arvio lentomatkustajien määrälle vuonna 2030.

Regressiomalli

Tyhjä

$y = 0.239x - 478.108$

Tarkka arvo: x = 2030 y = 7.062

Lentomatkustajien määrä vuonna 2030 on mallin mukaan 7,06 miljardia.

Vastaus: 7,06 mrd

VAHVISTA OSAAMISTA

538.

	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.

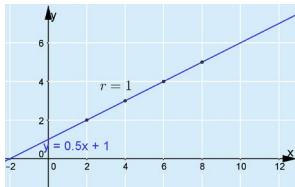
3. rivin 4. ruutu: äidin ja isän pituuden korrelaatiokerroin on 0,248 eli merkityksetön positiivinen riippuvuus.

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. Vanhempien keskinäisten pituuksien välinen riippuvuus on merkityksetön positiivinen

- 539. a)** Korrelaatiokerroin r on 1, kun kaikki pisteet sijaitsevat samalla suoralla. Esimerkiksi aineisto

x	y
2	2
4	3
6	4
8	5

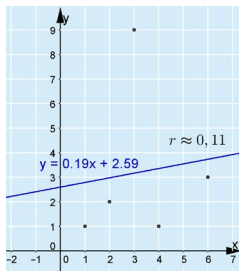


Vastaus: esim.

x	2	4	6	8
y	2	3	4	5

- b)** Korrelaatio on lähellä nollaa, jos pisteet sijaitsevat niin hajallaan, ettei pistejoukkoon voi asettaa suoraa. Esimerkiksi aineisto

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

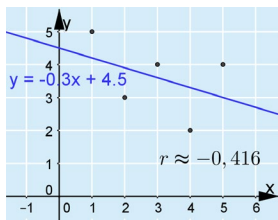


Vastaus: esim.

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

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

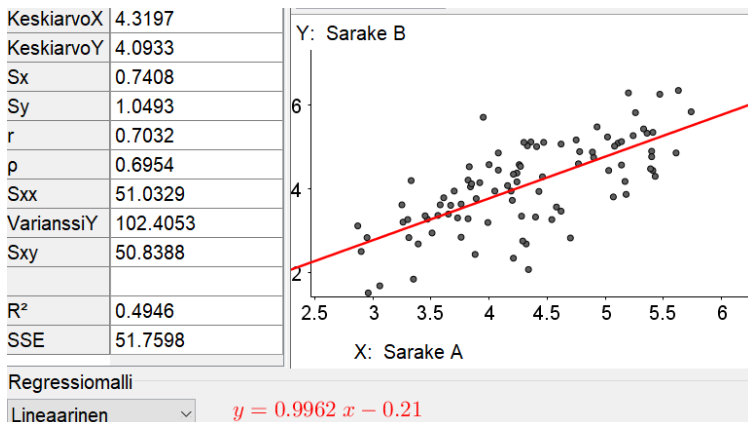
x	y
1	5
2	3
3	4
4	2
5	4



Vastaus: esim.

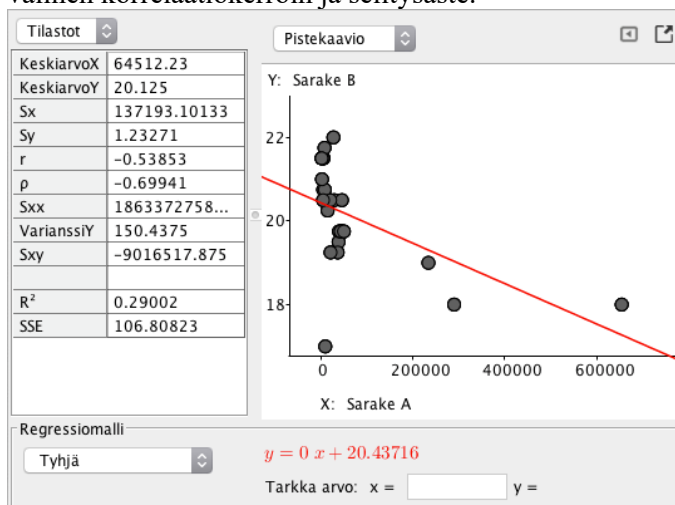
x	1	2	3	4	5
y	5	3	4	2	4

540. Korrelaatiokertoimeksi saadaan $r = 0,70$, joka on huomattava positiivinen korrelaatio.



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

541. Määritetään ohjelman avulla kunnan väkiluku ja kunnallisveroprosentin välinen korrelaatiokerroin ja selitysaste.



Ohjelma antaa korrelaatiokertoimeksi $r \approx -0,54$ ja selitysasteeksi $r^2 \approx 0,29 \approx 29\%$.

Korrelaatio on kohtalaisen negatiivinen eli väkiluvun kasvaessa veroprosentti pienenee. Väkiluku selittää noin 29 % veroprosentin vaihtelusta.

Vastaus: $r = -0,54, 29 \%$

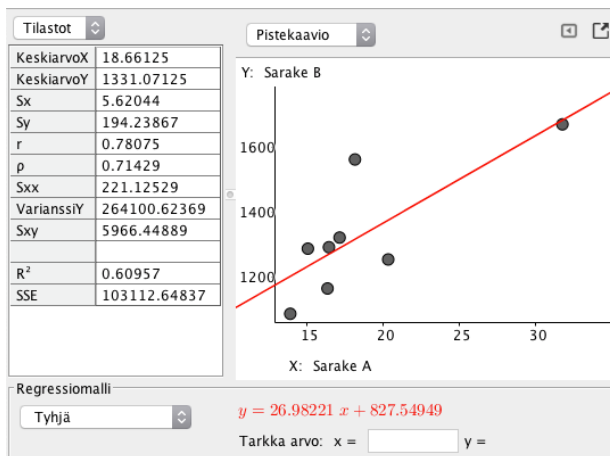
542. Määritetään kaikki kolme riippuvuussuhdetta: mobiili ja kiinteä, kiinteä ja SMS sekä mobiili ja SMS.

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	$-0,985... \approx -0,99$	$0,971... \approx 0,97$	
lankapuhelin ja SMS	$-0,590... \approx -0,59$	$0,348... \approx 0,35$	
mobiili ja SMS	$0,467... \approx 0,47$	$0,218... \approx 0,22$	

Vastaus: Mobiililaajakaistaliittymien ja lankapuhelinliittymien välinen korrelaatiokerroin on $-0,99$. Mobiililaajakaistaliittymien määrän kasvu selittää 97 % lankapuhelinliittymien määrän kasvusta.

Mobiililaajakaistaliittymien ja viestien välinen korrelaatiokerroin on $0,47$ ja selitysaste 22 %, lankapuhelinliittymien ja viestien välillä korrelaatiokerroin on $-0,59$ ja selitysaste 35 %.

543. Analysoidaan aineistoa ohjelman avulla.



Tuloksiksi saadaan seuraavat:

$$r = 0,78075$$

$$r^2 = 0,60957 \approx 61 \%$$

$$\text{regressiosuoran yhtälö } y = 26,98x + 827,55$$

Saatujen tulosten perusteella:

- a) Väite on tosi, koska hopean ja kullan hinnoilla on huomattava positiivinen korrelaatio.

Vastaus: tosi

- b) Väite on epätosi, koska hopean hinta selittää 61 % kullan hinnan vaihtelusta.

Vastaus: epätosi, 61 %

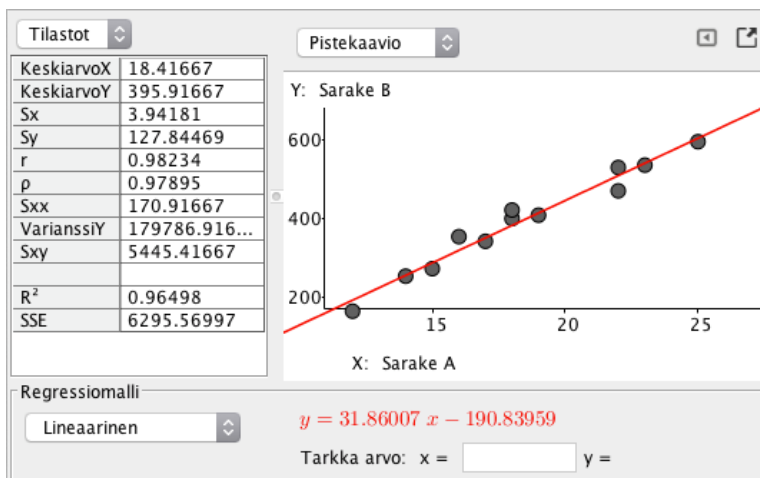
- c) Väite on tosi, koska kullan hinta = $26,98 \cdot \text{hopean hinta} + 827,55$.

Vastaus: tosi

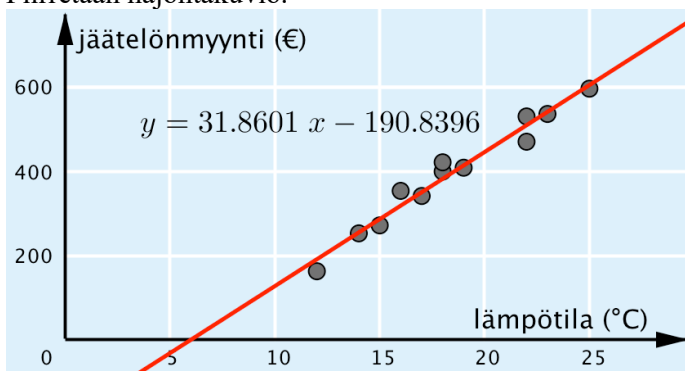
- d) Väite on epätosi, koska kullan hinnan noustessa hopean hinta nousee, koska regressiosuora on nouseva.

Vastaus: epätosi, nousee

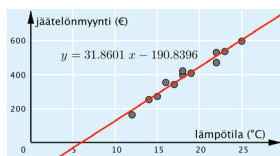
544. Analysoidaan aineistoa ohjelman avulla.



a) Piirretään hajontakuviio.



Lämpötilan ja jäätelönmyynnin riippuvuutta kuvaavan regressiosuoran yhtälö on $y = 31,8601x - 190,840$.



Vastaus: $y = 31,8601x - 190,840$

- b) Selitysaste on $r^2 = 0,964... \approx 96 \%$, joten lämpötila selittää 96 % jäätelönmyynnin vaihtelusta.

Vastaus: 96 %

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

Regressiomalli

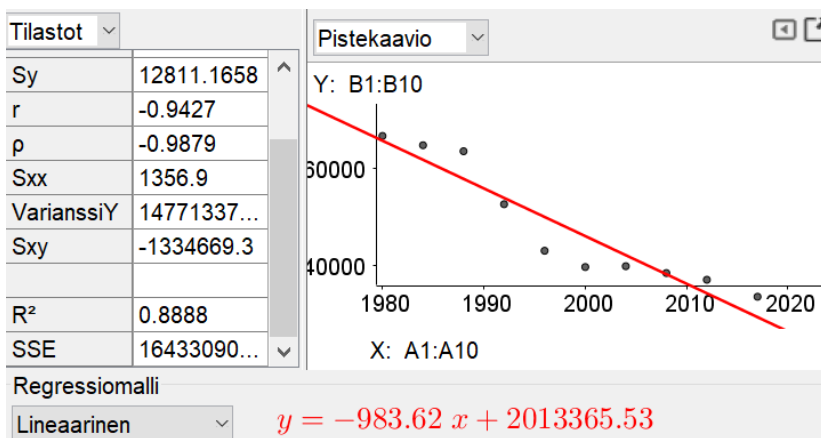
Lineaarinen

$y = 31.8601x - 190.8396$

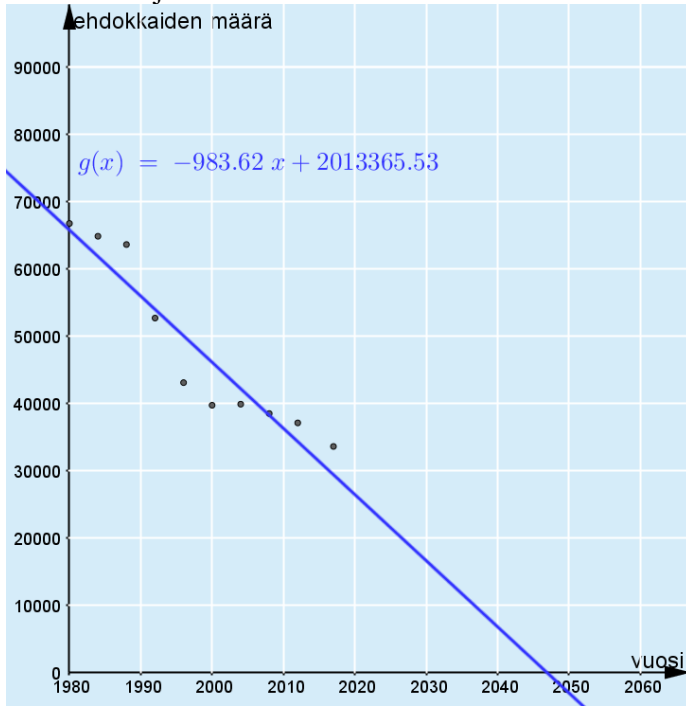
Tarkka arvo: x = 30 y = 764.9625

Vastaus: 765 €

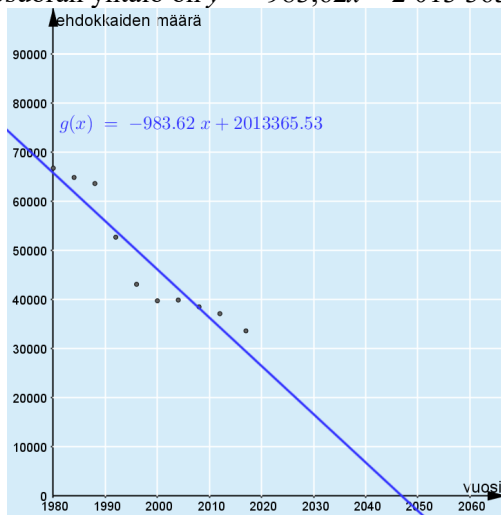
545. Analysoidaan aineistoa ohjelman avulla.



a) Piirretään hajontakuviio.



Vuosiluvun ja ehdokkaiden määrän riippuvuutta kuvaavan regressiosuoran yhtälö on $y = -983,62x + 2\,013\,365,53$.



Vastaus:

$$y = -983,62x + 2\,013\,365,53$$

- b) Muuttujien välinen korrelaatiokerroin on $r = -0,942 \dots \approx -0,94$. Selitysaste on $r^2 = 0,888 \dots \approx 89 \%$, joten vuosiluku selittää 89 % ehdokasmäärän vaihtelusta.

Vastaus: $-0,94$ ja 89 %

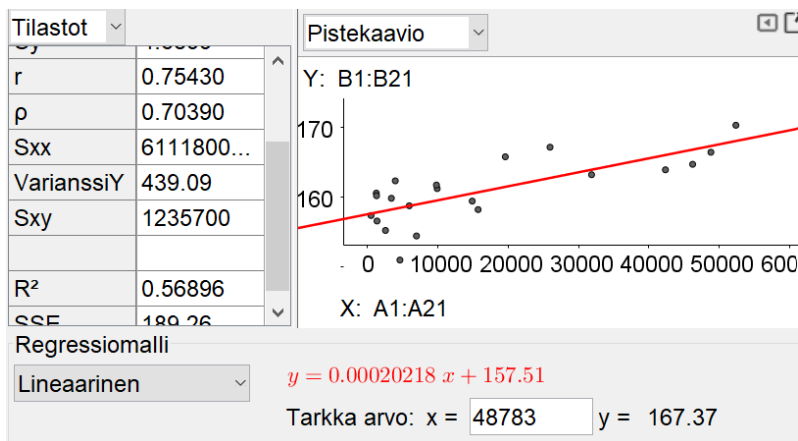
- c) Muodostetaan yhtälö ja ratkaistaan se ohjelmalla. Ohjelma antaa vuodeksi 2037,75. Jos kunnallisvaalit järjestetään joka neljäs vuosi, on vuosi 2037 kunnallisvaalivuosi, joten mallin mukaan ehdokkaita on vuonna 2037 yhtä monta kuin vuoden 2017 vaaleissa valittiin valtuutettuja.

CAS	
1	RatkaiseNumeerisesti[g(x)=8999]
☐	$\approx \{x = 2037.75\}$

Vastaus: vuonna 2037

546. Analysoidaan aineistoa ohjelman avulla.

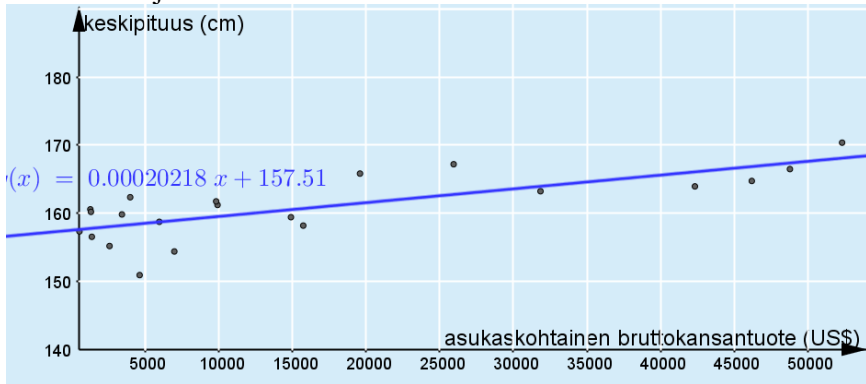
n	21
Keskiarvo	160.8746
σ	4.5726
s	4.6855
Σx	3378.3663
Σx^2	543932.3727
Min	150.914
Q1	157.761
Mediaani	160.575
Q3	164.3345
Max	170.3612



- a) Keskipituuksien keskiarvo on $160,87... \text{ cm} \approx 160,9 \text{ cm}$.
 Vaihteluväli on $[150,91... \text{ cm} ; 170,36... \text{ cm}] \approx [150,9 \text{ cm} ; 170,4 \text{ cm}]$.
 Vaihteluvälin pituus on $170,36... \text{ cm} - 150,91... \text{ cm} \approx 19,5 \text{ cm}$.
 Keskihajonta on $4,572... \text{ cm} \approx 4,6 \text{ cm}$.

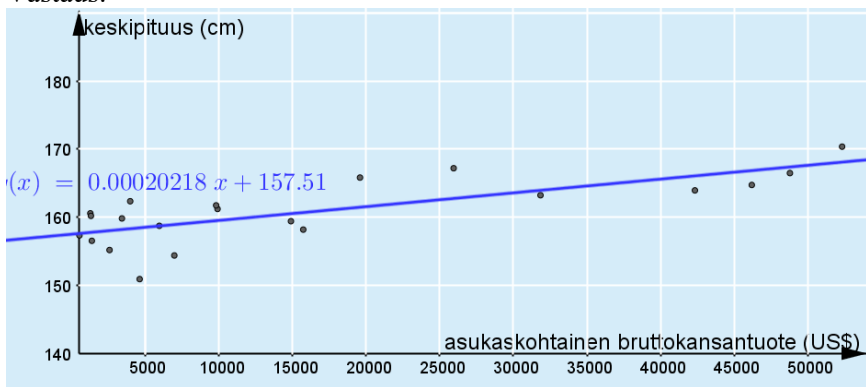
Vastaus: keskiarvo $160,9 \text{ cm}$ vaihteluväli $[150,9 \text{ cm} ; 170,4 \text{ cm}]$,
 vaihteluvälin pituus on $19,5 \text{ cm}$, keskihajonta $4,6 \text{ cm}$

b) Piirretään hajontakuvio.



Asukaskohtaisen bruttokansantuotteen ja tyttöjen keskipituuden riippuvuutta kuvaavan regressiosuoran yhtälö on $y = 0,00020218x + 157,51$.

Vastaus:



, $y = 0,00020218x + 157,51$

- c)** Muuttujien välinen korrelaatiokerroin on $r = 0,754... \approx 0,75$. Selitysaste on $r^2 = 0,568... \approx 57\%$, joten bruttokansantuote selittää 57 % keskipituuden vaihtelusta.

Vastaus: 0,75 ja 57 %

- d) Ohjelma antaa mallin mukaiseksi keskipituudeksi 167,37 cm $\approx$ 167,4 cm, kun asukaskohtainen bruttokansantuote on 48 783 dollaria. Mallin antama pituus on 167,4 cm – 166,5 cm = 0,9 cm pidempi kuin tilastoarvo.

Regressiomalli

Lineaarinen $y = 0.00020218 x + 157.51$

Tarkka arvo: x = 48783 y = 167.37

Vastaus: malli antaa 0,9 cm pidemmän keskipituuden

547. Analysoidaan aineistoa ohjelman avulla.

Vuosi ja hedelmällisyysluku:		Vuosi ja elinajanodote:		Hedelmällisyysluku ja elinajanodote:	
KeskiarvoX	1989	KeskiarvoX	1989	KeskiarvoX	3.4788
KeskiarvoY	3.4788	KeskiarvoY	64.33402	KeskiarvoY	64.33402
Sx	17.17556	Sx	17.17556	Sx	0.92985
Sy	0.92985	Sy	5.52316	Sy	5.52316
r	-0.96913	r	0.98287	r	-0.9765
p	-0.99778	p	1	p	-0.99778
Sxx	17110	Sxx	17110	Sxx	50.14828
VarianssiY	50.14828	VarianssiY	1769.30918	VarianssiY	1769.30918
Sxy	-897.704	Sxy	5407.81	Sxy	-290.87267
R ²	0.9392	R ²	0.96603	R ²	0.95356
SSE	3.04877	SSE	60.10935	SSE	82.17419

- a) Määritetään korrelaatiokertoimet muuttujien välille.
 vuosi ja hedelmällisyysluku: $r = -0,969$
 vuosi ja elinajanodote: $r = 0,983$
 hedelmällisyysluku ja elinajanodote: $r = -0,977$

Vastaus: vuosiluku ja hedelmällisyysluku $r \approx -0,969$, vuosiluku ja elinajanodote $r \approx 0,983$, hedelmällisyysluku ja elinajanodote $r \approx -0,977$

- 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

548. Määritetään muuttujien väliset korrelaatiokertoimet.

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

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

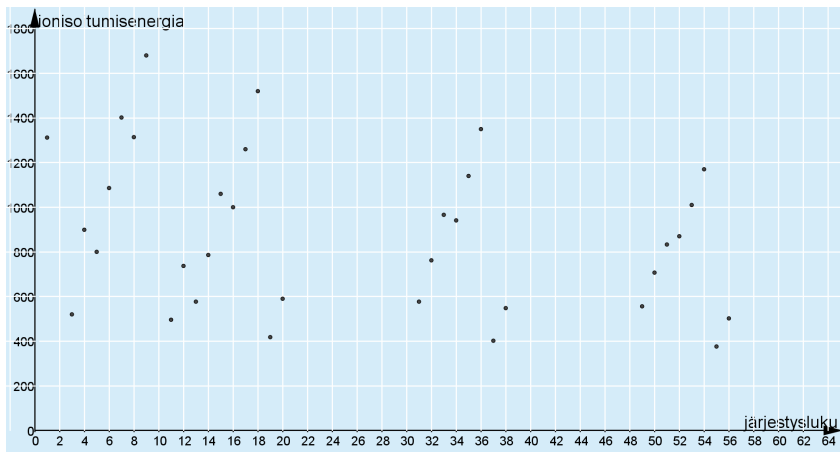
Kuolemien ja väestömäärän välillä $r \approx 0,96$

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 $r \approx 0,96$.

- 549. 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.

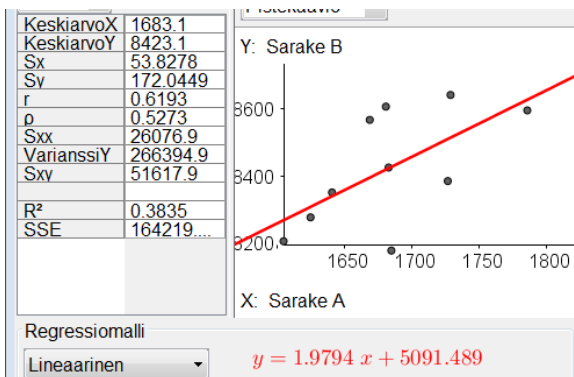
Vastaus: Ionisoitumisenergian ja järjestysluvun riippuvuutta voidaan kuvata kahdeksalla suoralla.

- 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 $y = 198,61x - 68,32$

550. Analysoidaan aineistoa.

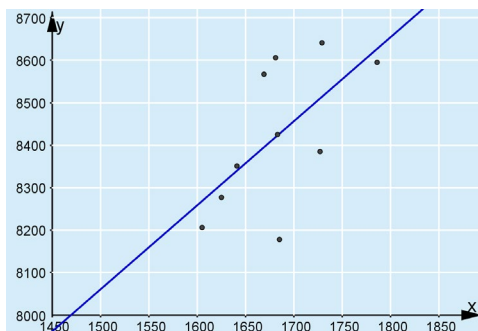
Määritetään
100 m juoksun sekä
kuulantäynnön
yhteispistemäärän ja
ottelun
kokonaispistemäärän
korrelaatiokerroin ja
regressiosuoran
yhtälö.



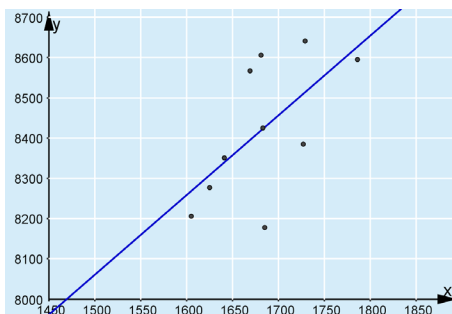
Korrelaatiokerroin on $r = 0,6193 \approx 0,619$ ja regressiosuoran yhtälö on $y = 1,9794x + 5091,489$.

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,



551. a) Täydennetään taulukko. Lasketaan ensin sarakkeiden x_i ja y_i keskiarvot $\bar{x}$ ja $\bar{y}$. Keskiarvot ovat $\bar{x} = 78$ ja $\bar{y} = 77$.

	A	B	C	D	E	F	G	H
1		Koepisteet						
2		Taidehistoria	Tilastotiede					
3	Oppilas	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
4	1	95	85	=B4-SB\$10	=C4-SC\$10	=D4^2	=E4^2	=D4*E4
5	2	85	95	=B5-SB\$10	=C5-SC\$10	=D5^2	=E5^2	=D5*E5
6	3	80	70	=B6-SB\$10	=C6-SC\$10	=D6^2	=E6^2	=D6*E6
7	4	70	65	=B7-SB\$10	=C7-SC\$10	=D7^2	=E7^2	=D7*E7
8	5	60	70	=B8-SB\$10	=C8-SC\$10	=D8^2	=E8^2	=D8*E8
9	Summa Σ	=SUMMA(B4:B8)	=SUMMA(C4:C8)			=SUMMA(F4:F8)	=SUMMA(G4:G8)	=SUMMA(H4:H8)
10	Keskiarvo	=KESKIARVO(B4:B8)	=KESKIARVO(C4:C8)					
11								

	A	B	C	D	E	F	G	H	I
1		Koepisteet							
2		Taidehistoria	Tilastotiede						
3	Oppilas	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$	
4	1	95	85	17	8	289	64	136	
5	2	85	95	7	18	49	324	126	
6	3	80	70	2	-7	4	49	-14	
7	4	70	65	-8	-12	64	144	96	
8	5	60	70	-18	-7	324	49	126	
9	Summa Σ	390	385			730	630	470	
10	Keskiarvo	78	77						
11									

Määritetään oppilaiden eri kokeiden pistemäärien välinen korrelaatiokerroin.

$$\begin{aligned}
 r &= \frac{\sum (y - \bar{y})(x - \bar{x})}{\sqrt{\sum (x - \bar{x})^2 \cdot \sum (y - \bar{y})^2}} \\
 &= \frac{470}{\sqrt{730 \cdot 630}} \\
 &= 0,693... \\
 &\approx 0,69
 \end{aligned}$$

Vastaus: $r \approx 0,69$

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