

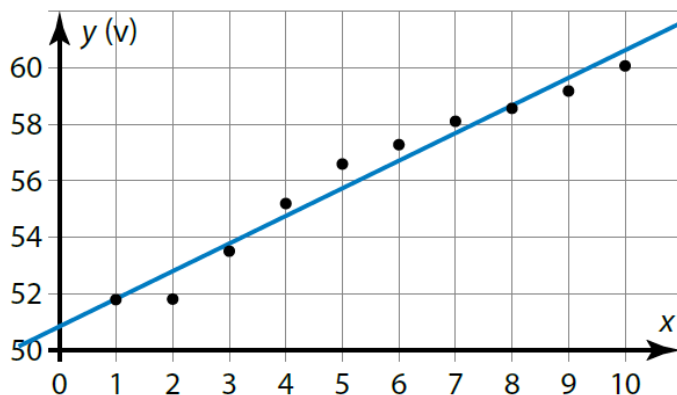
## 6.1

Ratkaistaan tehtävä GeoGebran taulukkolaskennalla.

Valitaan selittäväksi muuttujaksi  $x$  tuloluokka ja selitettäväksi muuttujaksi  $y$  elinajanodote. Taulukoidaan havaintoarvot.

	A	B
1	Tuloluokka $x$	Elinajan odote $y$ (v)
2	1	51.8
3	2	51.8
4	3	53.5
5	4	55.2
6	5	56.6
7	6	57.3
8	7	58.1
9	8	58.6
10	9	59.2
11	10	60.1

Piirretään hajontakuviota ja määritetään korrelaatiokerroin ja selitysaste.



KeskiarvoX	5.5
KeskiarvoY	56.22
Sx	3.0277
Sy	3.0177
r	0.9814
$\rho$	0.997
Sxx	82.5
VarianssiY	81.956
Sxy	80.7
R <sup>2</sup>	0.9632
SSE	3.0167

Korrelaatiokerroin  $r \approx 0,98$  ja selitysaste  $r^2 \approx 0,96$ .

Hajontakuvion ja korrelaatiokertoimen perusteella tuloluokan ja elinajanodotteen välillä on voimakas positiivinen lineaarinen riippuvuus: tuloluokan noustessa elinajanodote kasvaa. Selitysasteen perusteella suomalaisen tuloluokka selittää 96 % elinajanodotteen vaihtelusta.

### Vastaus

$r \approx 0,98$ ; voimakas positiivinen lineaarinen riippuvuus;

$r^2 \approx 0,96$ ; tuloluokka selittää 96 % elinajanodotteen vaihtelusta

## 6.2

- a) Hajontakuvioon pisteet eivät sovi suoralle, joten hajontakuvio ei kuvaa lineaarista riippuvuutta. Vastausvaihtoehto 5 ei lineaarista riippuvuutta sopii kuvioon
- b) Hajontakuvioon voidaan sovittaa nouseva suora, joten kyseessä on positiivinen lineaarinen riippuvuus. Pisteet sopivat suoralle ainoastaan kohtalaisesti, joten kyseessä on heikko lineaarinen riippuvuus. Vastausvaihtoehdot 1 positiivinen lineaarinen riippuvuus ja 4 heikko lineaarinen riippuvuus sopivat kuvioon.
- c) Hajontakuvioon voidaan sovittaa nouseva suora, joten kyseessä on positiivinen lineaarinen riippuvuus. Pisteet sopivat suoralle hyvin, joten kyseessä on voimakas lineaarinen riippuvuus. Vastausvaihtoehdot 1 positiivinen lineaarinen riippuvuus ja 3 voimakas lineaarinen riippuvuus sopivat kuvioon.
- d) Hajontakuvioon voidaan sovittaa laskeva suora, joten kyseessä on negatiivinen lineaarinen riippuvuus. Pisteet sopivat suoralle hyvin, joten kyseessä on voimakas lineaarinen riippuvuus. Vastausvaihtoehdot 1 positiivinen lineaarinen riippuvuus ja 3 voimakas lineaarinen riippuvuus sopivat kuvioon.

### Vastaus

- a) 5
- b) 1, 4
- c) 1, 3
- d) 2, 3

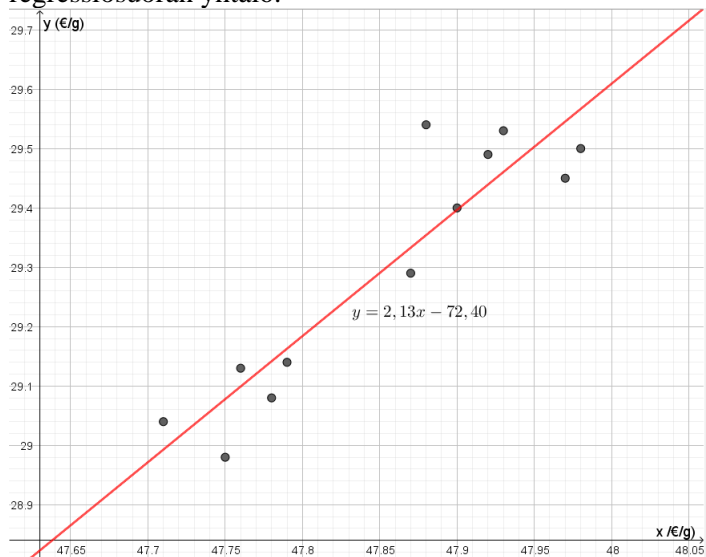
## 6.3

a) Ratkaistaan tehtävä GeoGebran taulukkolaskennalla.

Valitaan selittäväksi muuttujaksi  $x$  kullan hinta (€/g) ja selitettäväksi muuttujaksi  $y$  platinan hinta (€/g). Taulukoidaan havaintoarvot.

A	B
Kullan hinta (€/g)	Platinan hinta (€/g)
47.98	29.5
47.93	29.53
47.88	29.54
47.92	29.49
47.9	29.4
47.97	29.45
47.79	29.14
47.76	29.13
47.75	28.98
47.71	29.04
47.78	29.08
47.87	29.29

Piirretään hajontakuvio ja määritetään korrelaatiokerroin ja regressiosuoran yhtälö.



KeskiarvoX	47.8533
KeskiarvoY	29.2975
Sx	0.0917
Sy	0.2115
r	0.9212
ρ	0.8462
Sxx	0.0925
VarianssiY	0.492
Sxy	0.1965
R <sup>2</sup>	0.8487
SSE	0.0744

Korrelaatiokerroin  $r \approx 0,92$  ja regressiosuoran yhtälö on  $y = 2,13x - 72,40$ .

Hajontakuvion ja korrelaatiokertoimen perusteella kullan ja platinan hinnan välillä on voimakas positiivinen lineaarinen riippuvuus: kullan hinnan noustessa myös hopean hinta nousee.

b) Lasketaan platinan hinta  $y$ , kun kullan hinta  $x = 47,80$  (€/g).

$$\begin{aligned}y &= 2,13x - 72,40 && \text{Sijoitetaan } x = 47,80. \\&= 2,13 \cdot 47,80 - 72,40 \\&\approx 29,41 \text{ (€/g)}\end{aligned}$$

Lasketaan platinan hinta  $y$ , kun kullan hinta  $x = 46,00$  (€/g).

$$\begin{aligned}y &= 2,13x - 72,40 && \text{Sijoitetaan } x = 46,00. \\&= 2,13 \cdot 46,00 - 72,40 \\&\approx 25,58 \text{ (€/g)}\end{aligned}$$

Platinan hinnan ennusteiksi saadaan 29,41 €/g ja 25,58 €/g.

c) Ennuste 26,41 €/g lienee luotettava, sillä kullan hinta 47,80 €/g on havaintoarvojen vaihteluvälin sisäpuolella, missä hintojen välillä on voimakas positiivinen lineaarinen riippuvuus.

Ennuste 25,58 €/kg ei ole yhtä luotettava, sillä kullan hinta 46,90 €/g on havaintoarvojen vaihteluvälin ulkopuolella, eikä tiedetä, kuinka kauas lineaarinen riippuvuus ulottuu.

### Vastaus

- a)  $r \approx 0,92$ ;  $y = 2,13x - 72,40$ ; voimakas positiivinen lineaarinen riippuvuus
- b) 29,41 €/g ja 25,58 €/g
- c) Havaintoarvojen vaihteluvälillä riippuvuus on voimakas ja ennuste luotettava, vaihteluvälin ulkopuolella ennuste ei ole yhtä luotettava.

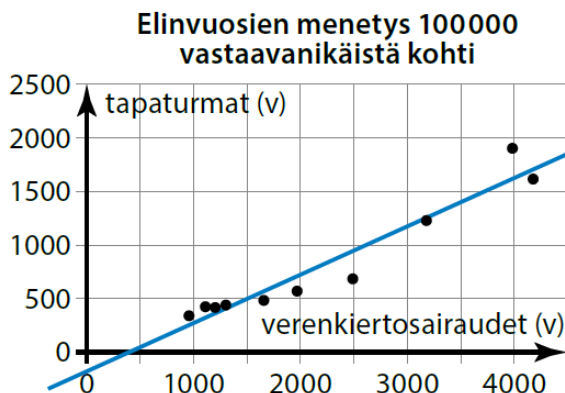
## 6.4

Ratkaistaan tehtävä GeoGebran taulukkolaskennalla.

Valitaan selittäväksi muuttujaksi  $x$  verenkiertosairauksien aiheuttamat elinvuosien menetykset ja selitettäväksi muuttujaksi  $y$  tapaturmien aiheuttamat elinvuosien menetykset. Taulukoidaan havaintoarvot.

	A	B
1	Verenkiertosairaudet	Tapaturmat
2	3973	1900
3	4177	1611
4	3172	1221
5	2485	680
6	1965	570
7	1675	483
8	1306	435
9	1201	419
10	1110	423
11	957	340

Piirretään hajontakuviio ja määritetään korrelaatiokerroin.



KeskiarvoX	2202.1
KeskiarvoY	808.2
Sx	1199.0394
Sy	562.1579
r	0.9624
p	0.9758
Sxx	12939258.9
VarianssiY	2844193.6
Sxy	5838600.8

Korrelaatiokerroin  $r \approx 0,96$

Hajontakuvion ja korrelaatiokertoimen perusteella verenkiertosaireuksien aiheuttamien elinvuosien menetyksien ja tapaturmien aiheuttamien elinvuosien menetyksien välillä on voimakas positiivinen lineaarinen riippuvuus: verenkiertosaireuksien aiheuttamien elinvuosien menetyksien lisääntyessä myös tapaturmien aiheuttamat elinvuosien menetykset lisääntyvät.

### Vastaus

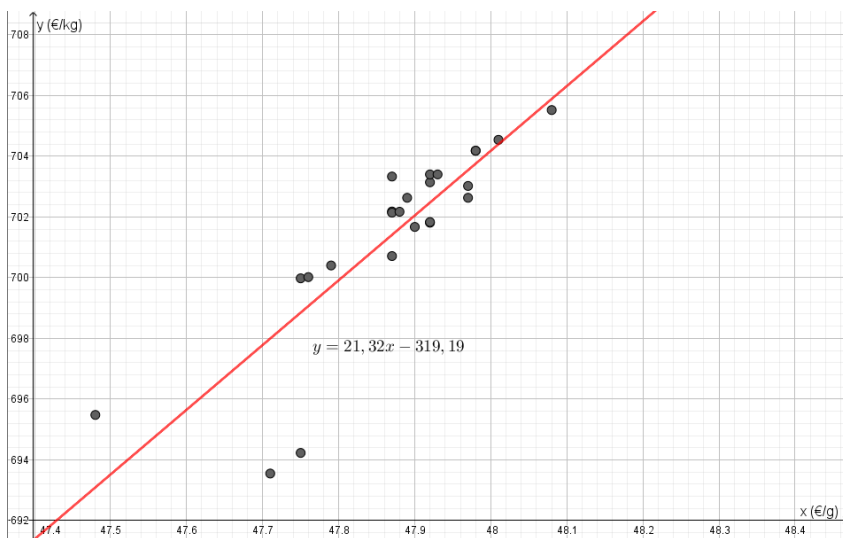
$r \approx 0,96$ ; voimakas positiivinen lineaarinen riippuvuus



## 6.5

- a) Ratkaistaan tehtävä GeoGebran taulukkolaskennalla. Avataan aineisto GeoGebralla. Valitaan selittäväksi muuttujaksi  $x$  kullan hinta (€/kg) ja selitettäväksi muuttujaksi  $y$  hopean hinta (€/kg).

Piirretään hajontakuviio ja määritetään korrelaatiokerroin ja regressiosuoran yhtälö.



KeskiarvoX	47.8746
KeskiarvoY	701.5058
Sx	0.1229
Sy	3.0792
r	0.8509
$\rho$	0.8736
Sxx	0.3474
VarianssiY	218.0742
Sxy	7.4066
R <sup>2</sup>	0.7241
SSE	60.1647

Korrelaatiokerroin  $r \approx 0,85$  ja regressiosuoran yhtälö on  $y = 21,32x - 319,19$ .

Hajontakuvion ja korrelaatiokertoimen perusteella kullan ja hopean hinnan välillä on voimakas positiivinen lineaarinen riippuvuus: kullan hinnan noustessa myös hopean hinta nousee.

- b) Lasketaan hopean hinta  $y$ , kun kullan hinta  $x = 47,80$  (€/g).

$$\begin{aligned} y &= 21,32x - 319,19 && \text{Sijoitetaan } x = 47,80. \\ &= 21,32 \cdot 47,80 - 319,19 \\ &\approx 699,91 \text{ (€/kg)} \end{aligned}$$

Lasketaan hopean hinta  $y$ , kun kullan hinta  $x = 48,60$  (€/g).

$$\begin{aligned} y &= 21,32x - 319,19 && \text{Sijoitetaan } x = 48,60. \\ &= 21,32 \cdot 48,60 - 319,19 \\ &\approx 716,96 \text{ (€/kg)} \end{aligned}$$

Platinan hinnan ennusteiksi saadaan 699,91 €/kg ja 716,96 €/kg.

- c) Korrelaatiokerroin on hieman pienempi kuin esimerkissä 2, joten riippuvuus on hieman heikompi. Ennustetut hinnat ovat kuitenkin samansuuntaiset kuin esimerkissä 2, joten ennusteet tukevat toisiaan.

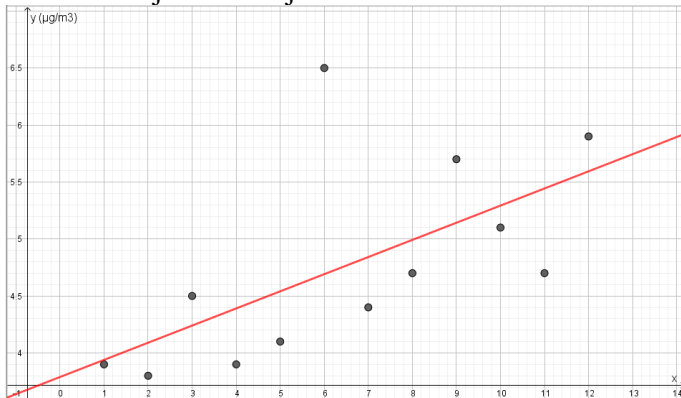
### Vastaus

- a)  $r \approx 0,85$ ;  $y = 21,32x - 319,19$ ; voimakas positiivinen lineaarinen riippuvuus  
b) 699,91 €/kg ja 716,96 €/kg  
c) Riippuvuus on hieman heikompi, mutta ennusteet ovat samansuuntaiset kuin esimerkissä 2.

## 6.6

- a) Ratkaistaan tehtävä GeoGebran taulukkolaskennalla. Avataan aineisto GeoGebralla. Valitaan selittäväksi muuttujaksi  $x$  kuukausi ja selitettäväksi muuttujaksi  $y$  pienhiukkaspitoisuus ( $\mu\text{g}/\text{m}^3$ ).

Piirretään hajontakuvio ja määritetään korrelaatiokerroin.



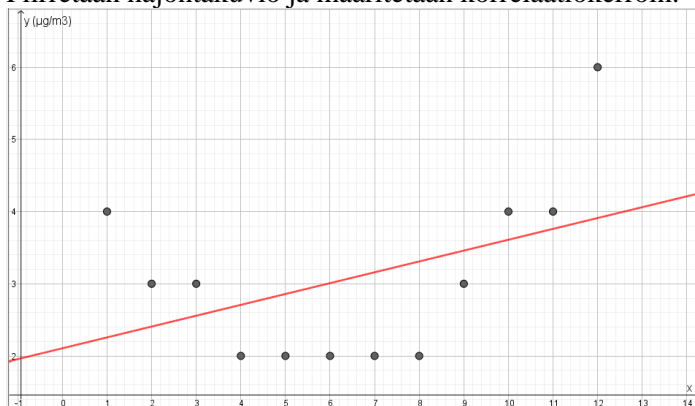
KeskiarvoX	6.5
KeskiarvoY	4.7667
Sx	3.6056
Sy	0.8721
r	0.6216
$\rho$	0.7509
Sxx	143
VarianssiY	8.3667
Sxy	21.5
R <sup>2</sup>	0.3864
SSE	5.1341

Korrelaatiokerroin  $r \approx 0,62$

Hajontakuvion ja korrelaatiokertoimen perusteella kuukauden ja pienhiukkaspitoisuuden välillä on huomattava positiivinen lineaarinen riippuvuus.

- b) Valitaan nyt selittäväksi muuttujaksi  $x$  kuukausi ja selitettäväksi muuttujaksi  $y$  typpidioksidipitoisuus ( $\mu\text{g}/\text{m}^3$ ).

Piirretään hajontakuviota ja määritetään korrelaatiokerroin.



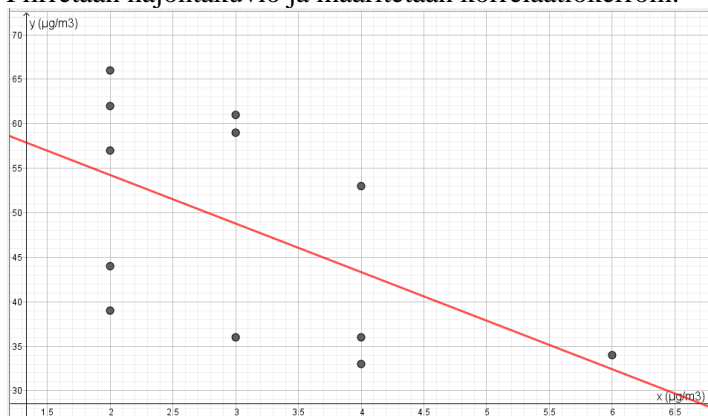
KeskiarvoX	6.5
KeskiarvoY	3.0833
Sx	3.6056
Sy	1.2401
r	0.4371
p	0.3313
Sxx	143
VarianssiY	16.9167
Sxy	21.5
R²	0.1911
SSE	13.6841

Korrelaatiokerroin  $r \approx 0,43$

Hajontakuvion ja korrelaatiokertoimen perusteella kuukauden ja pienhiukkaspitoisuuden välillä on kohtalainen positiivinen lineaarinen riippuvuus.

- c) Valitaan nyt selittäväksi muuttujaksi  $x$  typpidioksidipitoisuus ( $\mu\text{g}/\text{m}^3$ ) ja selitettäväksi muuttujaksi  $y$  otsonipitoisuus ( $\mu\text{g}/\text{m}^3$ ).

Piirretään hajontakuviota ja määritetään korrelaatiokerroin.



KeskiarvoX	3.0833
KeskiarvoY	48.3333
Sx	1.2401
Sy	12.507
r	-0.5412
$\rho$	-0.6232
Sxx	16.9167
VarianssiY	1720.6667
Sxy	-92.3333
R <sup>2</sup>	0.2929
SSE	1216.6995

Korrelaatiokerroin  $r \approx -0,54$

Hajontakuviota ja korrelaatiokertoimen perusteella kuukauden ja pienhiukkaspitoisuuden välillä on kohtalainen negatiivinen lineaarinen riippuvuus.

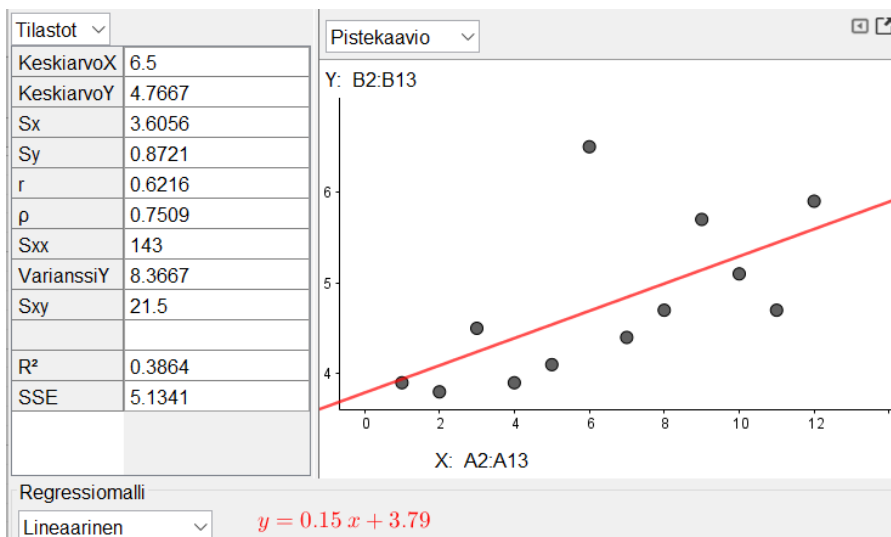
### Vastaus

- $r \approx 0,62$ ; huomattava positiivinen lineaarinen riippuvuus
- $r \approx 0,43$ ; kohtalainen positiivinen lineaarinen riippuvuus
- $r \approx -0,54$ ; kohtalainen negatiivinen lineaarinen riippuvuus

## 6.7

- a) Ratkaistaan tehtävä GeoGebran taulukkolaskennalla. Avataan aineisto GeoGebralla. Valitaan selittäväksi muuttujaksi  $x$  kuukausi ja selitettäväksi muuttujaksi  $y$  pienhiukkaspitoisuus ( $\mu\text{g}/\text{m}^3$ ).

Piirretään hajontakuviota ja kokeillaan erilaisia regressiomalleja.

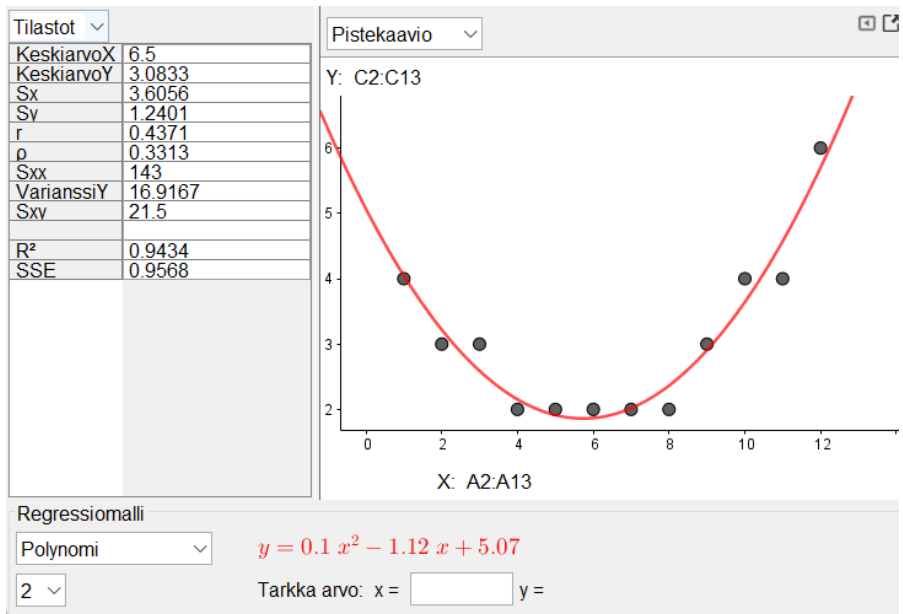


Parhaiten sopiva regressiomalli on lineaarinen. Tällöin regressiokäyrän yhtälö on  $y = 0,15x + 3,79$ .

Korrelaatiokerroin  $r \approx 0,62$

- b) Valitaan nyt selittäväksi muuttujaksi  $x$  kuukausi ja selitettäväksi muuttujaksi  $y$  typpidioksidipitoisuus ( $\mu\text{g}/\text{m}^3$ ).

Piirretään hajontakuviota ja kokeillaan erilaisia regressiomalleja.

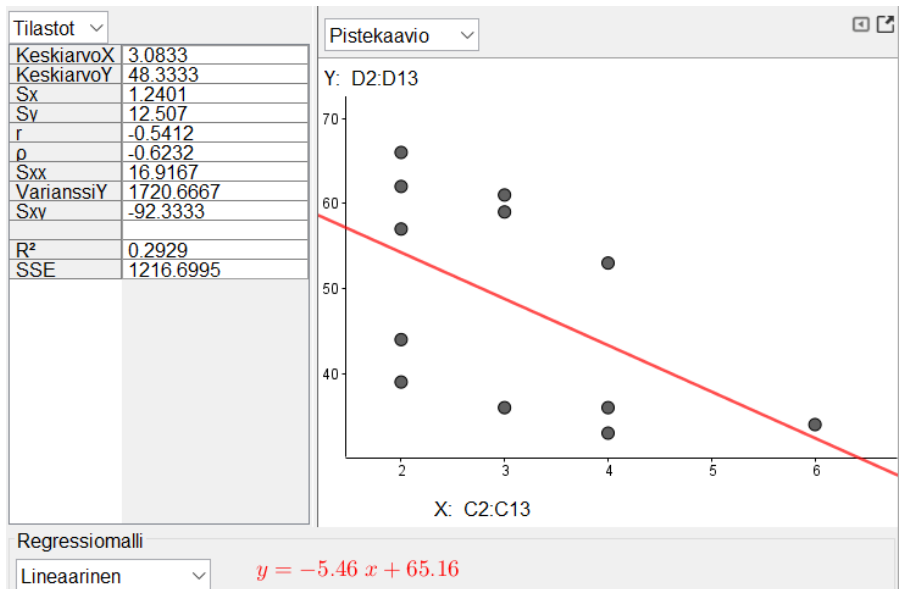


Parhaiten sopiva regressiomalli on polynomi. Tällöin regressiokäyrän yhtälö on  $y = 0,10x^2 - 1,12x + 5,07$ .

Sopivimman polynomin astelukua on silmämääräisesti vaikea arvioida.

- c) Valitaan nyt selittäväksi muuttujaksi  $x$  typpidioksidipitoisuus ( $\mu\text{g}/\text{m}^3$ ) ja selitettäväksi muuttujaksi  $y$  otsonipitoisuus ( $\mu\text{g}/\text{m}^3$ ).

Piirretään hajontakuviota ja kokeillaan erilaisia regressiomalleja.



Parhaiten sopiva regressiomalli on lineaarinen. Tällöin regressiokäyrän yhtälö on  $y = -5,46x + 65,16$ .

Korrelaatiokerroin on kuitenkin ainoastaan  $r \approx -0,54$ , joten riippuvuus on tälläkin mallilla ainoastaan kohtalainen.

### Vastaus

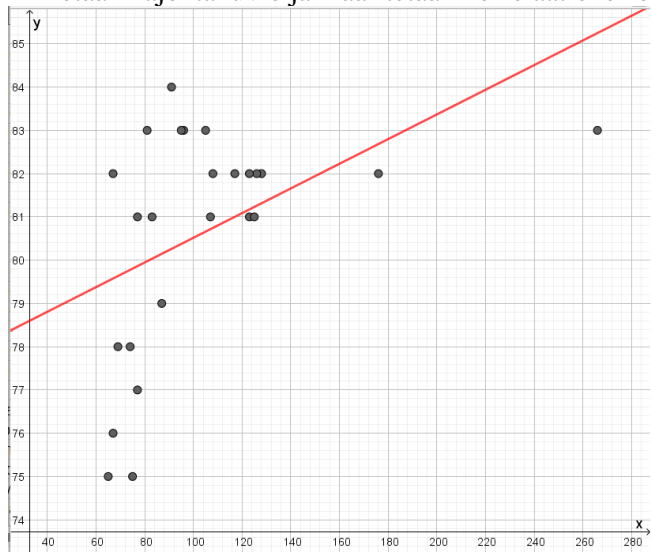
- lineaarinen regressiomalli,  $y = 0,15x + 3,79$
- polynomi regressiomalli, esimerkiksi  $y = 0,10x^2 - 1,12x + 5,07$
- lineaarinen regressiomalli, mutta riippuvuus on vain kohtalainen,  $y = -5,46x + 65,16$



## 6.8

- a) Ratkaistaan tehtävä GeoGebran taulukkolaskennalla. Avataan aineisto GeoGebralla. Valitaan selittäväksi muuttujaksi  $x$  bruttokansantuote ja selitettäväksi muuttujaksi  $y$  elinajanodote.

Piirretään hajontakuviio ja määritetään korrelaatiokerroin.

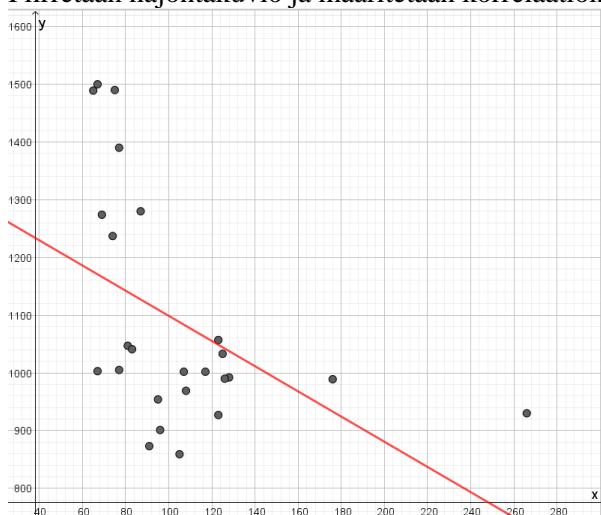


KeskiarvoX	104.32
KeskiarvoY	80.64
Sx	42.8658
Sy	2.6439
r	0.4621
$\rho$	0.5316
Sxx	44099.44
VarianssiY	167.76
Sxy	1256.88
$R^2$	0.2135
SSE	131.9376

Korrelaatiokerroin  $r \approx 0,46$

Hajontakuvion ja korrelaatiokertoimen perusteella bruttokansantuotteen ja elinajanodotteen välillä on kohtalainen positiivinen lineaarinen riippuvuus.

- b) Valitaan nyt selittäväksi muuttujaksi  $x$  bruttokansantuote ja selitettäväksi muuttujaksi  $y$  kuolleisuusaste. Piirretään hajontakuviota ja määritetään korrelaatiokerroin.



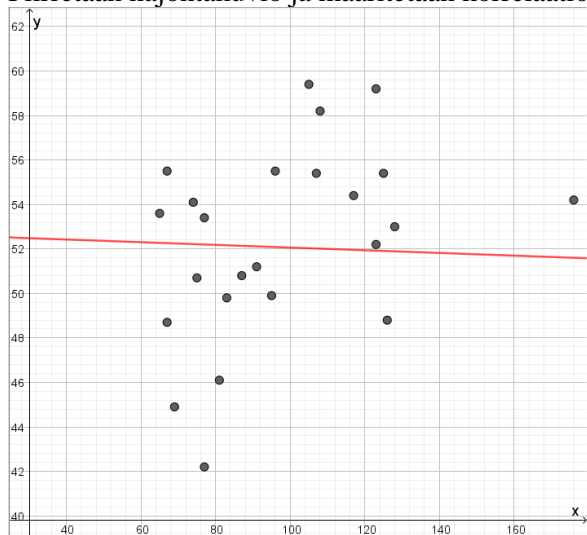
KeskiarvoX	104.32
KeskiarvoY	1089.36
Sx	42.8658
Sy	199.8966
r	-0.4678
p	-0.6282
Sxx	44099.44
VarianssiY	959007.76
Sxy	-96204.88
R <sup>2</sup>	0.2188
SSE	749132.5567

Korrelaatiokerroin  $r \approx -0,46$

Hajontakuvion ja korrelaatiokertoimen perusteella bruttokansantuotteen ja kuolleisuusasteen välillä on kohtalainen negatiivinen lineaarinen riippuvuus.

- c) Valitaan nyt selittäväksi muuttujaksi  $x$  bruttokansantuote ja selitettäväksi muuttujaksi  $y$  huoltosuhte.

Piirretään hajontakuvio ja määritetään korrelaatiokerroin.



KeskiarvoX	104.32
KeskiarvoY	52.036
Sx	42.8658
Sy	4.5088
r	-0.0574
p	0.2081
Sxx	44099.44
VarianssiY	487.8976
Sxy	-266.288
R <sup>2</sup>	0.0033
SSE	486.2897

Korrelaatiokerroin  $r \approx -0,06$

Hajontakuvion ja korrelaatiokertoimen perusteella bruttokansantuotteen ja huoltosuhteen välillä on merkityksetön negatiivinen lineaarinen riippuvuus.

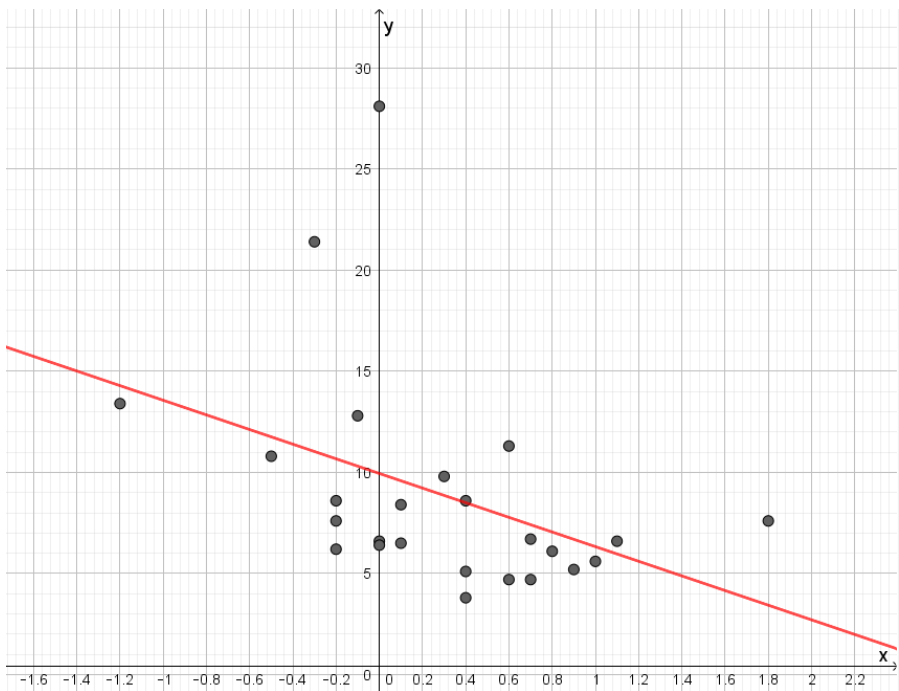
### Vastaus

- a)  $r \approx 0,46$ ; kohtalainen positiivinen lineaarinen riippuvuus  
b)  $r \approx -0,47$ ; kohtalainen negatiivinen lineaarinen riippuvuus  
c)  $r \approx -0,06$ ; merkityksetön negatiivinen lineaarinen riippuvuus

## 6.9

Ratkaistaan tehtävä GeoGebran taulukkolaskennalla. Avataan aineisto GeoGebralla. Valitaan selittäväksi muuttujaksi  $x$  inflaatioaste ja selitettäväksi muuttujaksi  $y$  työttömyysaste.

Piirretään hajontakuvio ja määritetään korrelaatiokerroin.



KeskiarvoX	0.288
KeskiarvoY	8.904
Sx	0.6133
Sy	5.4666
r	-0.4059
p	-0.5373
Sxx	9.0264
VarianssiY	717.2096
Sxy	-32.6588
R <sup>2</sup>	0.1648
SSE	599.0454

Korrelaatiokerroin  $r \approx -0,41$

Hajontakuvion ja korrelaatiokertoimen perusteella inflaation ja työttömyysasteen välillä on kohtalainen negatiivinen lineaarinen riippuvuus. Aineisto siis tukee jossain määrin hypoteesia.

### Vastaus

Inflaation ja työttömyysasteen välillä on kohtalainen negatiivinen lineaarinen riippuvuus,  $r \approx -0,41$ .

Aineisto tukee jossain määrin hypoteesia.

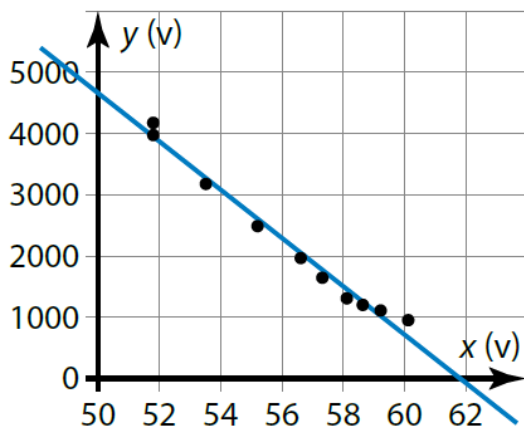
## 6.10

Ratkaistaan tehtävä GeoGebran taulukkolaskennalla.

Valitaan selittäväksi muuttujaksi  $x$  elinajanodote ja selitettäväksi muuttujaksi  $y$  elinvuosien menetys. Taulukoidaan havaintoarvot.

	A	B
1	Elinajanodote (v)	Elinvuosien menetys (v)
2	51.8	3973
3	51.8	4177
4	53.5	3172
5	55.2	2485
6	56.6	1965
7	57.3	1657
8	58.1	1306
9	58.6	1201
10	59.2	1110
11	60.1	957

Piirretään hajontakuvi ja määritetään korrelaatiokerroin ja selitysaste.



KeskiarvoX	56.22
KeskiarvoY	2200.3
Sx	3.0177
Sy	1199.9318
r	-0.9916
$\rho$	-0.997
Sxx	81.956
VarianssiY	12958526.1
Sxy	-32314.66
$R^2$	0.9832
SSE	217088.6102

Korrelaatiokerroin  $r \approx -0,99$  ja selitysaste  $r^2 \approx 0,98$ .

Hajontakuvion ja korrelaatiokertoimen perusteella elinajanodotteen ja elinvuosien menetyksen välillä on voimakas negatiivinen lineaarinen riippuvuus: elinajanodotteen kasvaessa elinvuosien menetys pienenee. Selitysasteen perusteella suomalaisen elinajanodote selittää 98 % elinvuosien menetyksen vaihtelusta.

### Vastaus

$r \approx -0,99$ ; voimakas negatiivinen lineaarinen riippuvuus;

$r^2 \approx 0,98$ ; elinajanodote selittää 98 % elinvuosien menetyksen vaihtelusta

## 6.11

- a) Hajontakuvioon voidaan sovittaa laskeva suora, joten kyseessä on negatiivinen lineaarinen riippuvuus. Pisteet sopivat suoralle hyvin, joten kyseessä on voimakas lineaarinen riippuvuus. Vastausvaihtoehdot 2 negatiivinen lineaarinen riippuvuus, 3 voimakas riippuvuus ja 4 voimakas lineaarinen riippuvuus sopivat kuvioon.
- b) Hajontakuvioon pisteet eivät sovi suoralle, joten hajontakuvio ei kuvaa lineaarista riippuvuutta. Pisteet sopivat kuitenkin hyvin kaarevalle käyrälle, joten hajontakuvio kuvaa voimakasta riippuvuutta. Vastausvaihtoehto 3 voimakas riippuvuus sopii kuvioon
- c) Hajontakuvioon pisteet eivät sovi suoralle, joten hajontakuvio ei kuvaa lineaarista riippuvuutta. Pisteet eivät myöskään sovi muulle käyrälle vaan hajoavat satunnaisesti koordinaatistoon. Vastausvaihtoehto 5 ei riippuvuutta sopii kuvioon
- d) Hajontakuvioon voidaan sovittaa nousevat suora, joten kyseessä on positiivinen lineaarinen riippuvuus. Pisteet sopivat suoralle hyvin, joten kyseessä on voimakas lineaarinen riippuvuus. Vastausvaihtoehdot 1 positiivinen lineaarinen riippuvuus, 3 voimakas riippuvuus ja 4 voimakas lineaarinen riippuvuus sopivat kuvioon.

### Vastaus

- a) 2, 3, 4
- b) 3
- c) 5
- d) 1, 3, 4



## 6.12

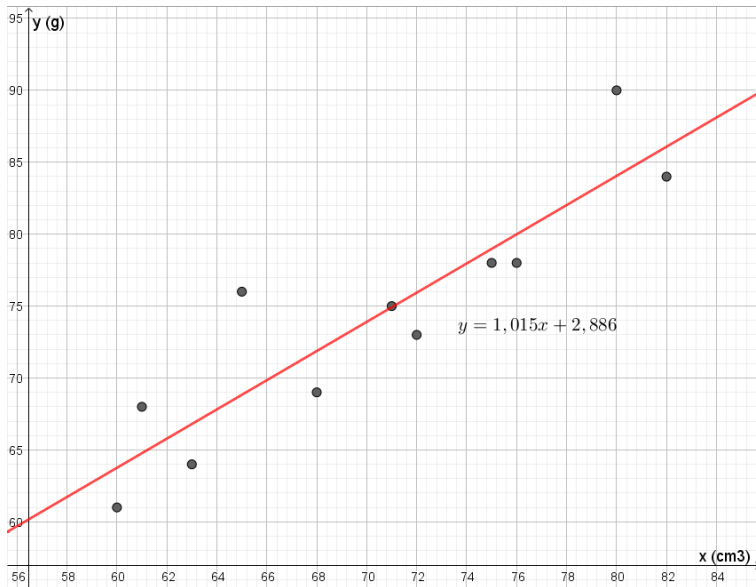
a) Ratkaistaan tehtävä GeoGebran taulukkolaskennalla.

Valitaan selittäväksi muuttujaksi  $x$  tilavuus (cm) ja selitettäväksi muuttujaksi  $y$  massa (g). Taulukoidaan havaintoarvot.

	A	B
1	Tilavuus $x$ (cm <sup>3</sup> )	Massa $y$ (g)
2	60	61
3	65	76
4	75	78
5	61	68
6	71	75
7	80	90
8	63	64
9	68	69
10	72	73
11	76	78
12	82	84

Tehtävässä voi käyttää merkintöjen  $x$  ja  $y$  tilalla myös suureiden tunnuksia  $V$  (tilavuus) ja  $m$  (massa).

Piirretään hajontakuviota ja määritetään korrelaatiokerroin ja regressiosuoran yhtälö.



KeskiarvoX	70.2727
KeskiarvoY	74.1818
Sx	7.5378
Sy	8.5067
r	0.899
ρ	0.9157
Sxx	568.1818
VarianssiY	723.6364
Sxy	576.4545
R²	0.8082
SSE	138.7886

Korrelaatiokerroin  $r \approx 0,90$  ja regressiosuoran yhtälö on  $y = 1,015x + 2,886$ .

Hajontakuvion ja korrelaatiokertoimen perusteella tilavuuden ja massan välillä on voimakas positiivinen lineaarinen riippuvuus: tilavuuden kasvaessa myös massa kasvaa.

b) Lasketaan massa  $y$ , kun tilavuus  $x = 70$  (cm<sup>3</sup>).

$$\begin{aligned} y &= 1,015x + 2,886 && \text{Sijoitetaan } x = 70. \\ &= 1,015 \cdot 70 + 2,886 \\ &\approx 74 \text{ (g)} \end{aligned}$$

Lasketaan massa  $y$ , kun tilavuus  $x = 0$  (cm<sup>3</sup>).

$$\begin{aligned} y &= 1,015x + 2,886 && \text{Sijoitetaan } x = 0. \\ &= 1,015 \cdot 0 + 2,886 \\ &\approx 3 \text{ (g)} \end{aligned}$$

Massoiksi saadaan 74 g ja 3 g.

c) Pelkän korrelaatiokertoimen perusteella regressiomalli ja sen ennusteet vaikuttavat luotettavilta havaintoarvojen vaihteluvälillä  $60 \text{ cm}^3 - 82 \text{ cm}^3$ . Mittaustuloksissa vaikuttaa kuitenkin olevan jokin systemaattinen virhe, sillä suolaliuoksen massa ei voi olla 3 g, kun sen tilavuus on nolla.

### Vastaus

- a)  $r \approx 0,90$ ;  $y = 1,015x + 2,886$ ; voimakas positiivinen lineaarinen riippuvuus
- b) 29,41 €/g ja 25,58 €/g
- c) Pelkän korrelaatiokertoimen perusteella regressiomalli ja sen ennusteet vaikuttavat luotettavilta havaintoarvojen vaihteluvälillä  $60 \text{ cm}^3 - 82 \text{ cm}^3$ . Mittaustuloksissa vaikuttaa kuitenkin olevan jokin systemaattinen virhe, sillä suolaliuoksen massa ei voi olla 3 g, kun sen tilavuus on nolla.

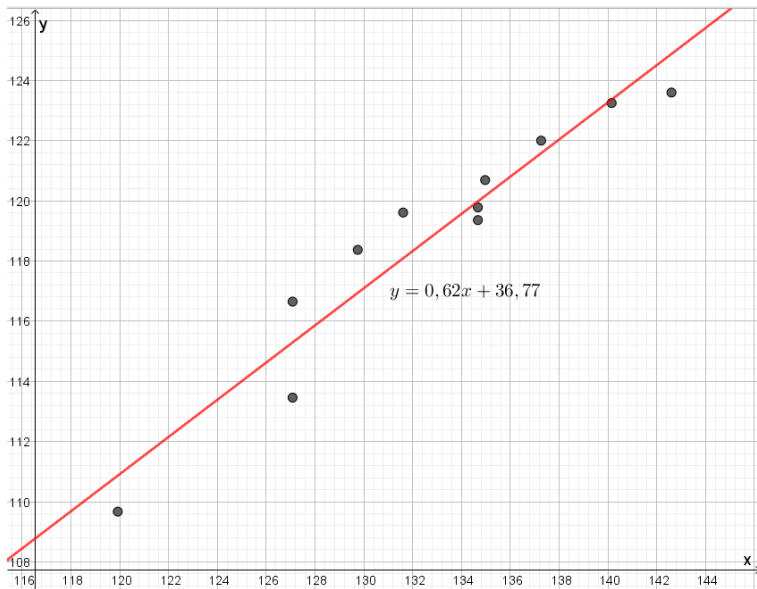
## 6.13

a) Ratkaistaan tehtävä GeoGebran taulukkolaskennalla.

Valitaan selittäväksi muuttujaksi  $x$  ansiotasoindeksi ja selitettäväksi muuttujaksi  $y$  kuluttajahintaindeksi. Taulukoidaan havaintoarvot.

	A	B	C
1	<b>Vuosi</b>	<b>Ansiotasoindeksi x</b>	<b>Kuluttajahintaindeksi y</b>
2	2010	119.92	109.67
3	2011	127.08	113.46
4	2012	127.08	116.65
5	2013	129.75	118.37
6	2014	131.61	119.61
7	2015	134.67	119.36
8	2016	134.67	119.78
9	2017	134.97	120.69
10	2018	137.26	122
11	2019	140.15	123.25
12	2020	142.6	123.6

Piirretään hajontakuvio ja määritetään korrelaatiokerroin ja regressiosuoran yhtälö.



KeskiarvoX	132.7055
KeskiarvoY	118.7673
Sx	6.5085
Sy	4.1863
r	0.9607
ρ	0.9817
Sxx	423.6083
VarianssiY	175.2528
Sxy	261.7514
R <sup>2</sup>	0.9229
SSE	13.5143

Korrelaatiokerroin  $r \approx 0,96$  ja regressiosuoran yhtälö on  $y = 0,62x + 36,77$ .

Hajontakuvion ja korrelaatiokertoimen perusteella ansiotason ja kuluttajahintojen välillä on voimakas positiivinen lineaarinen riippuvuus: ansiotaso ja kuluttajahinnat seuraavat toistensa kehitystä.

- b) Regressiosuoran kulmakerroin kuvaa sitä, kuinka monta yksikköä kuluttajahintaindeksi muuttuu, kun ansiotasoindeksi muuttuu yhden yksikön. Regressiosuoran kulmakerroin on  $k = 0,62$ .

Kulmakertoimesta voidaan siis päätellä, että ansiotason noustessa yhden prosenttiyksikön kuluttajahinnat nousevat 0,62 prosenttiyksikköä.

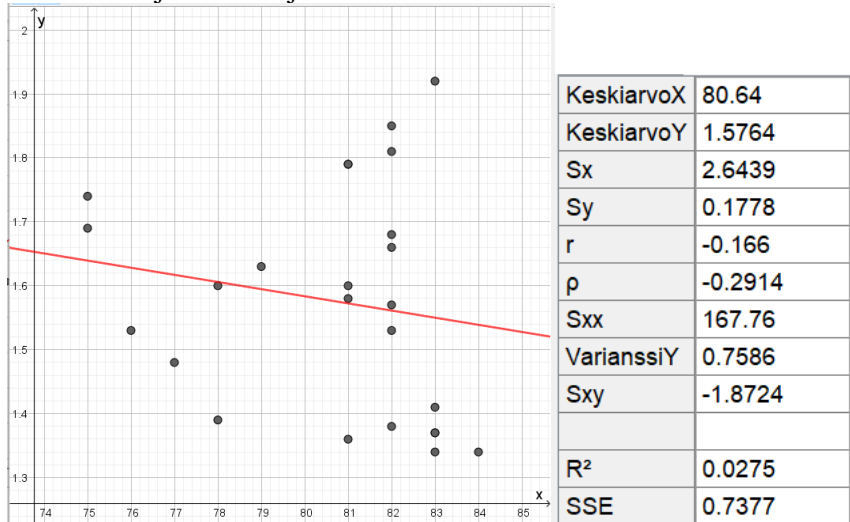
### Vastaus

- a)  $r \approx 0,96$ ;  $y = 0,62x + 36,77$ ; voimakas positiivinen lineaarinen riippuvuus  
b) 0,62 prosenttiyksikköä

## 6.14

- a) Ratkaistaan tehtävä GeoGebran taulukkolaskennalla. Avataan aineisto GeoGebralla. Valitaan selittäväksi muuttujaksi  $x$  elinajanodote ja selitettäväksi muuttujaksi  $y$  hedelmällisyysluku.

Piirretään hajontakuvio ja määritetään korrelaatiokerroin.

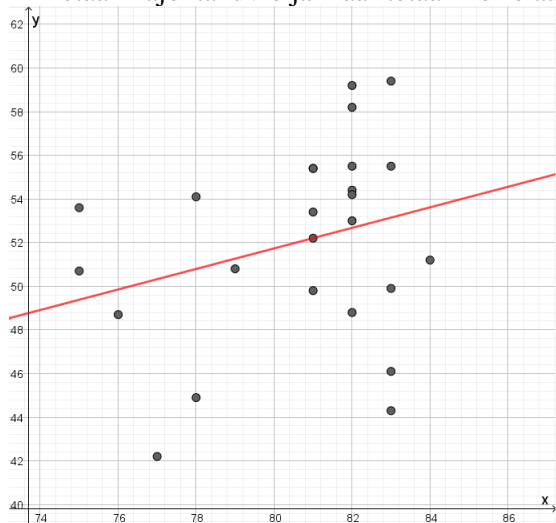


Korrelaatiokerroin  $r \approx -0,17$

Hajontakuvion ja korrelaatiokertoimen perusteella elinajanodotteen ja hedelmällisyysluvun välillä on merkityksetön negatiivinen lineaarinen riippuvuus.

- b) Valitaan nyt selittäväksi muuttujaksi  $x$  elinajanodote ja selitettäväksi muuttujaksi  $y$  huoltosuhte.

Piirretään hajontakuvi ja määritetään korrelaatiokerroin.



KeskiarvoX	80.64
KeskiarvoY	52.036
Sx	2.6439
Sy	4.5088
r	0.2759
$\rho$	0.2279
Sxx	167.76
VarianssiY	487.8976
Sxy	78.924
$R^2$	0.0761
SSE	450.7672

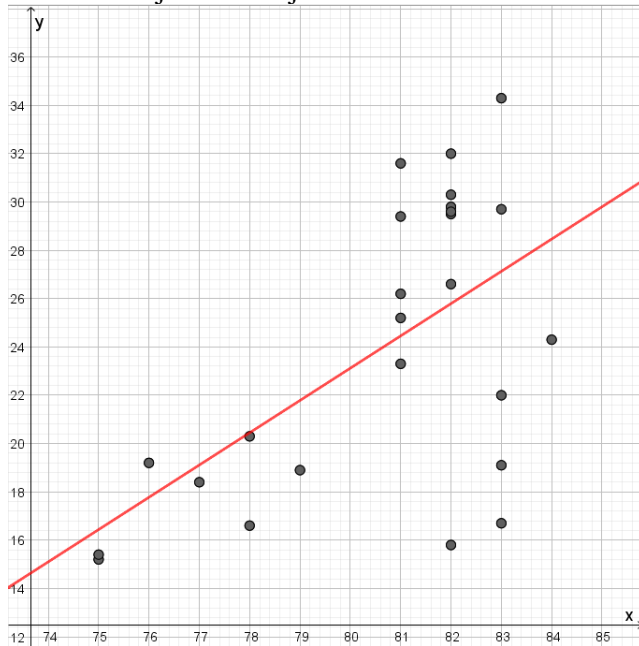
Korrelaatiokerroin  $r \approx 0,28$

Hajontakuvion ja korrelaatiokertoimen perusteella elinajanodotteen ja huoltosuhteen välillä on merkityksetön positiivinen lineaarinen riippuvuus.



- c) Valitaan nyt selittäväksi muuttujaksi  $x$  elinajanodote ja selitettäväksi muuttujaksi  $y$  sosiaalimenot.

Piirretään hajontakuvio ja määritetään korrelaatiokerroin.



KeskiarvoX	80.64
KeskiarvoY	23.976
Sx	2.6439
Sy	6.0788
r	0.5812
p	0.4528
Sxx	167.76
VarianssiY	886.8456
Sxy	224.184
R <sup>2</sup>	0.3378
SSE	587.2601

Korrelaatiokerroin  $r \approx 0,58$

Hajontakuvion ja korrelaatiokertoimen perusteella elinajanodotteen ja sosiaalimenojen välillä on kohtalainen positiivinen lineaarinen riippuvuus.

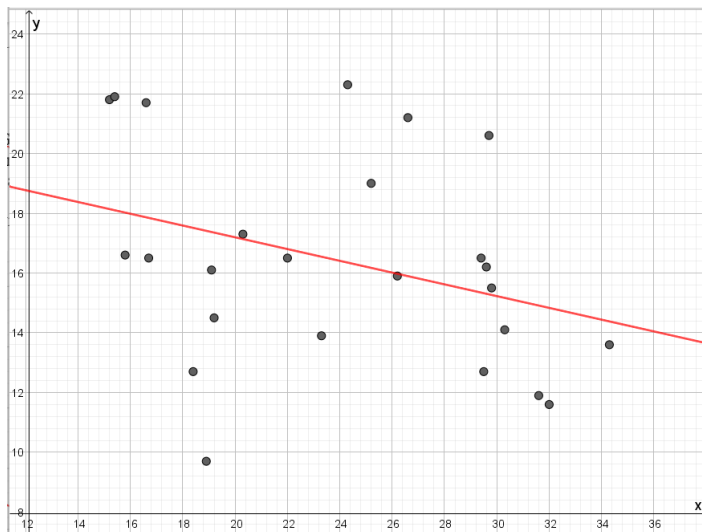
**Vastaus**

- a)  $r \approx -0,17$ ; merkityksetön negatiivinen lineaarinen riippuvuus
- b)  $r \approx 0,28$ ; merkityksetön positiivinen lineaarinen riippuvuus
- c)  $r \approx 0,58$ ; kohtalainen positiivinen lineaarinen riippuvuus

## 6.15

Ratkaistaan tehtävä GeoGebran taulukkolaskennalla. Avataan aineisto GeoGebralla. Valitaan selittäväksi muuttujaksi  $x$  sosiaalimenot ja selitettäväksi muuttujaksi  $y$  pienituloisuusaste.

Piirretään hajontakuviota ja määritetään korrelaatiokerroin.



KeskiarvoX	23.976
KeskiarvoY	16.412
Sx	6.0788
Sy	3.5961
r	-0.3327
$\rho$	-0.4404
Sxx	886.8456
VarianssiY	310.3664
Sxy	-174.5728
$R^2$	0.1107
SSE	276.0023

Korrelaatiokerroin  $r \approx -0,33$

Hajontakuvion ja korrelaatiokertoimen perusteella sosiaalimenojen ja pienituloisuusasteen välillä on kohtalainen negatiivinen lineaarinen riippuvuus. Riippuvuus ei kuitenkaan anna tietoa syy-seuraus-suhteesta, joten voi esimerkiksi olla, että pienempi pienituloisuusaste vähentää sosiaalimenoja.

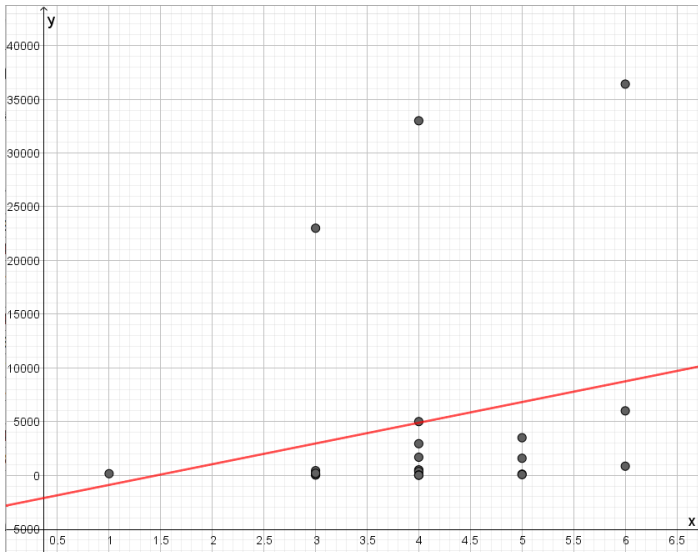
### **Vastaus**

Sosiaalimenojen ja pienituloisuusasteen välillä on kohtalainen negatiivinen lineaarinen riippuvuus,  $r \approx -0,33$ . Mutta riippuvuus ei anna tietoa syy-seuraus-suhteesta.

## 6.16

- a) Ratkaistaan tehtävä GeoGebran taulukkolaskennalla. Avataan aineisto GeoGebralla. Valitaan selittäväksi muuttujaksi  $x$  VEI-indeksi ja selitettäväksi muuttujaksi  $y$  uhrimäärä.

Piirretään hajontakuvio ja määritetään korrelaatiokerroin.



KeskiarvoX	4.087
KeskiarvoY	5058.2609
Sx	1.1644
Sy	10541.0783
r	0.213
p	0.3294
Sxx	29.8261
VarianssiY	244451530...
Sxy	57520.4783
R <sup>2</sup>	0.0454
SSE	233358538...

Korrelaatiokerroin  $r \approx 0,21$

Hajontakuvion ja korrelaatiokertoimen perusteella elinajanodotteen ja hedelmällisyysluvun välillä on merkityksetön negatiivinen lineaarinen riippuvuus.

**b)** Esimerkiksi:

Purkaukset ovat harvoin räjähdysmäisiä ja purkautuvan laavan nopeus ei ole kovin suuri, joten itse laava aiheuttaa harvoin merkittävästi kuolonuhreja.

Monet tulivuoret sijaitsevat harvaan asutulla alueella. Lisäksi tulivuorten aktiivisuutta seurataan eri mittausmenetelmin ja purkausten varalle on tehty evakuointisuunnitelmia.

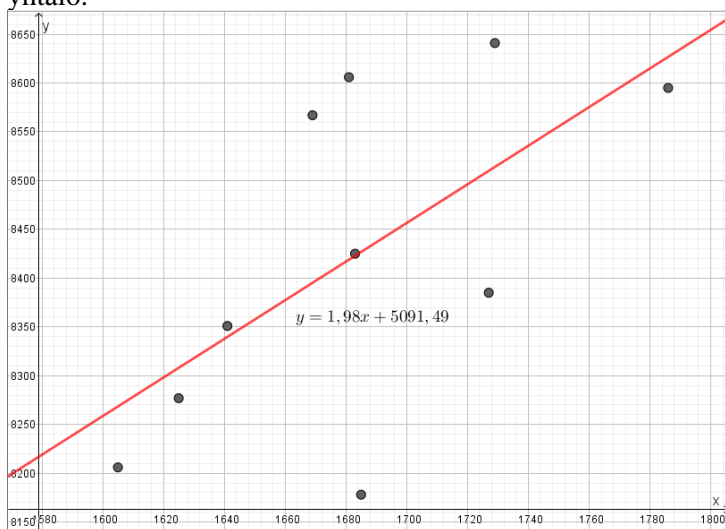
## Vastaus

**a)** Hajontakuvion ja korrelaatiokertoimen  $r \approx 0,21$  perusteella lineaarinen riippuvuus on merkityksetön

## 6.17

Ratkaistaan tehtävä GeoGebran taulukkolaskennalla. Avataan aineisto GeoGebralla. Lasketaan ensiksi kunkin urheilija pistesumma juoksusta ja kuulantönnöstä. Valitaan sitten selittäväksi muuttujaksi  $x$  pistesumma juoksusta ja kuulantönnöstä ja selitettäväksi muuttujaksi  $y$  ottelun kokonaispistemäärä.

Piirretään hajontakuvi ja määritetään korrelaatiokerroin ja regressiosuoran yhtälö.



KeskiarvoX	1683.1
KeskiarvoY	8423.1
Sx	53.8278
Sy	172.0449
r	0.6193
p	0.5273
Sxx	26076.9
VarianssiY	266394.9
Sxy	51617.9
R <sup>2</sup>	0.3835
SSE	164219.8869

Korrelaatiokerroin  $r \approx 0,62$  ja regressiosuora yhtälö  
 $y = 1,98x + 5091,49$ .

Lasketaan ensiksi Hämläläisen pistesumma juoksusta ja kuulantyyönnöstä.

$$858 + 732 = 1590$$

Lasketaan sitten ottelun kokonaispistemäärä  $y$ , kun pistesumma  
 $x = 1590$ .

$$\begin{aligned} y &= 1,98x + 5091,49 && \text{Sijoitetaan } x = 1590. \\ &= 1,98 \cdot 1590 + 5091,49 \\ &\approx 8240 \end{aligned}$$

Hämläläisen odotettavissa oleva kokonaispistemäärä olisi ollut 8240.

### **Vastaus**

$r \approx 0,62$ ; regressiosuoran yhtälö  $y = 1,98x + 5091,49$ ;  
kokonaispistemäärä 8240



## 6.18

Hun huomioidaan koko lapsuusaika, lapsen pituuden kasvaessa lapsen paino kasvaa.

a-1

Kengännumeron ja verenpaineen välillä ei ole korrelaatiota.

b-2

Maan bruttokansantuotteen kasvaessa lapsikuolleisuus yleensä pienenee. Korkean bruttokansantuotteen maissa lapsikuolleisuus on keskimäärin pienempää kuin matalan bruttokansantuotteen maissa.

c-3

Heitetyt tikat eivät yleensä muodosta nousevan tai laskevan suoran näköistä kuviota. Korrelaatiota ei siis ole.

d-2

1. Esimerkiksi:

- Korrelaatio on tilastollinen yhteys kahden muuttujan välillä.
- Positiivinen korrelaatio tarkoittaa, että toinen muuttuja suurenee, kun toinen suurenee.
- Negatiivinen korrelaatio tarkoittaa, että toinen muuttuja suurenee, kun toinen pienenee.
- Korrelaatiokerroin saa arvoja välillä  $-1 \dots 1$ .
- Mitä suurempi korrelaatiokertoimen itseisarvo on, sitä vahvempi on lineaarinen riippuvuus.
- Korrelaatiokerroin on suurin piirtein nolla, mikäli lineaarista riippuvuutta ei ole.

2. Kerro esimerkki kahdesta järkevästä ilmiöstä, joiden välillä on oikeanlainen yhteys. Esitä tämä esimerkki kahden muuttujan avulla. Selitä muuttujien välinen korrelaatio ja syy-yhteyden puuttuminen. Esimerkiksi talvihaalarien myyntimäärien ja liukastumisonnettomuuksien välillä on korrelaatio, mutta ei syy-yhteyttä. Sekä talvihaalarien myyntimäärä että liukastumisonnettomuudet ovat suurimmillaan talvella ja pienimmillään kesällä. Kylmät kelit lisäävät sekä talvihaalarien myyntimääriä että liukastumisonnettomuuksia. Talvihaalarien myynti ei kuitenkaan lisää liukastumistodennäköisyyttä eikä liukastumisonnettomuus saa henkilöä ostamaan talvihaalari