

Binomi 5 – Luku 7 – Tehtävien malliratkaisut

7.1

a)

Kuviossa I pisteet sijaitsevat likimain nousevan suoran suuntaisesti. Tämä tarkoittaa, että muuttujan x arvon kasvaessa myös muuttujan y arvo kasvaa, ja muuttujien välillä on positiivinen korrelaatio. Kuvioon I liittyvät siis kuvaus 2 ja ilmaisu A.

Kuviossa II pisteet eivät edes likimain sijaitse minkään suoran suuntaisesti, joten muuttujien välillä ei ole korrelaatiota. Muuttujan x arvon kasvaessa ei muuttujan y arvojen muutoksessa ilmene mitään säännönmukaisuutta. Kuvioon II liittyvät siis kuvaus 3 ja ilmaisu C.

Kuviossa III pisteet sijaitsevat likimain laskevan suoran suuntaisesti. Tämä tarkoittaa, että muuttujan x arvon kasvaessa muuttujan y arvo pienenee, ja muuttujien välillä on negatiivinen korrelaatio. Kuvioon III liittyvät siis kuvaus 1 ja ilmaisu B.

Vastaus:

1-III-B

2-I-A

3-II-C

7.2

Tehtävä voidaan ratkaista taulukkolaskentaohjelmalla tai GeoGebran taulukko-sovelluksella.

Tapa 1. Ratkaistaan tehtävä GeoGebran taulukko-sovelluksella

Katso tarvittaessa ohjeet Esimerkin 1 videolta.

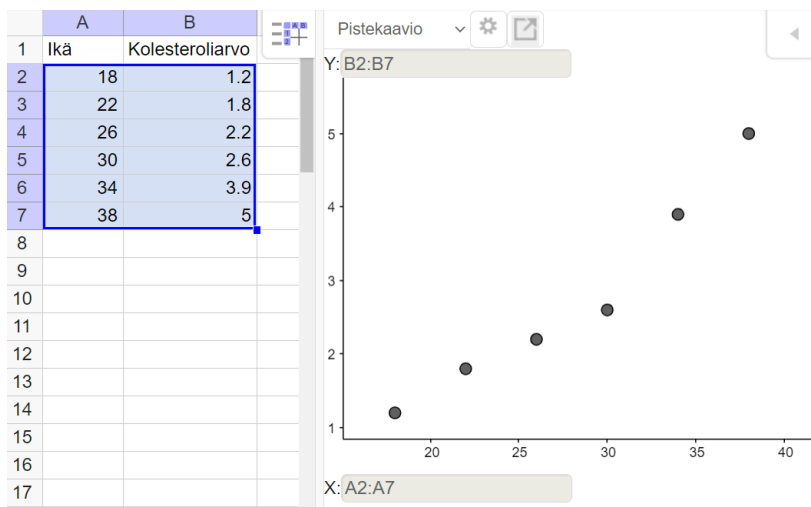
x = ikä

y = kolesteroliarvo

a)

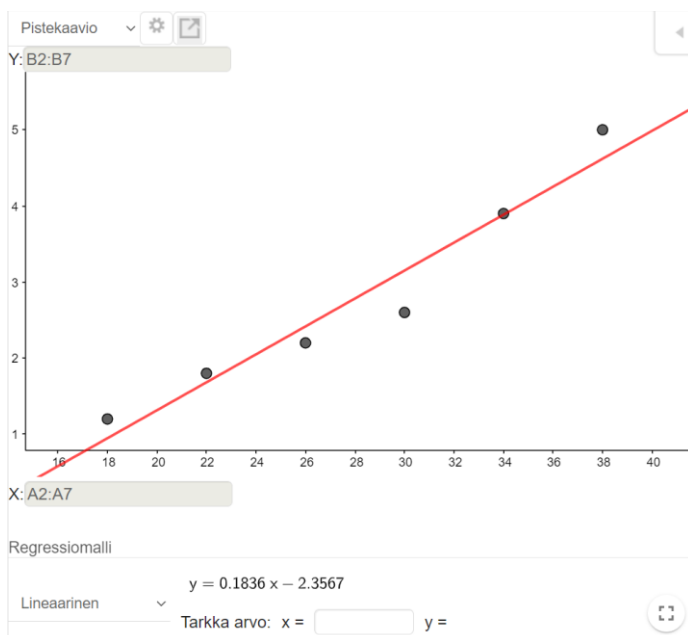
Kirjoitetaan tai kopioidaan havaintoarvot taulukko-sovellukseen.

Valitaan muuttujien arvot ja sitten Kahden muuttujan regressioanalyysi -työväline.



b)

Valitaan regressiomalliksi "lineaarinen".



Pyöristä lukuarvot kahden merkitsevän numeron tarkkuuteen.

Ohjelman antama regressiosuora on $y = 0,1836x - 2,3567$.

Kolesteroliarvon (y) riippuvuutta iästä (x) kuvaa siis yhtälö $y = 0,18x - 2,4$.

Tapa 2. Ratkaistaan tehtävä taulukkolaskentaohjelmalla
Katso tarvittaessa ohjeet Esimerkin 1 videolta.

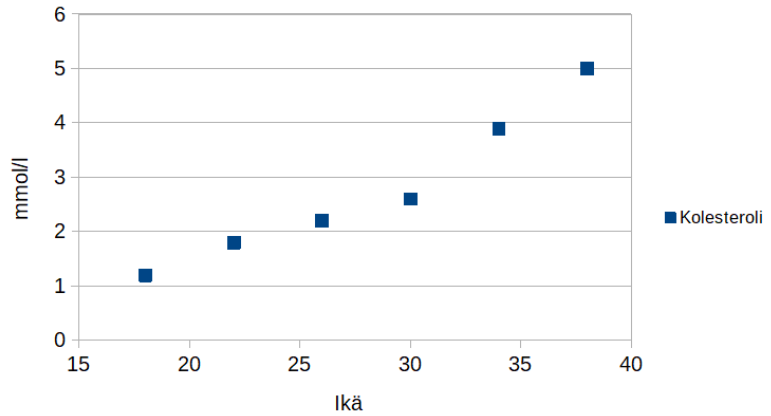
a)

Piirretään hajontakuviio ohjelmalla.

Ohje:

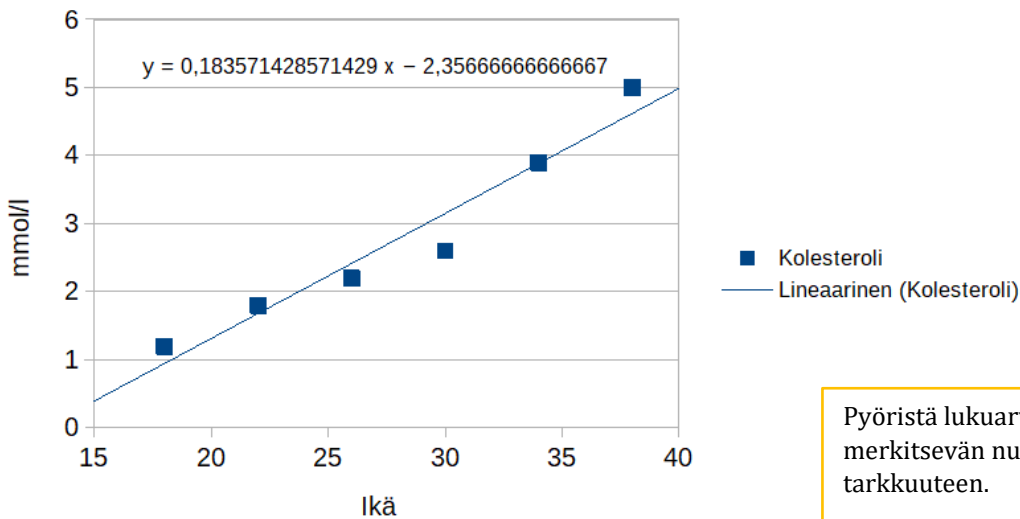
- Valitse iät ja kolesteroliarvot.
- Lisää Kaavio - XY (hajonta)
- Lisää akseleille otsikot.

	A	B
1	Ikä	Kolesteroliarvo (mmol/l)
2	18	1,2
3	22	1,8
4	26	2,2
5	30	2,6
6	34	3,9
7	38	5



b)

Lisätään trendiviiva. Valitaan "lineaarinen" ja "Näytä yhtälö".



Pyöristä lukuarvot kahden merkitsevän numeron tarkkuuteen.

Ohjelman antama regressiosuora on $y = 0,1835...x - 2,3566....$
Kolesteroliarvon (y) riippuvuutta iästä (x) kuvaa siis yhtälö $y = 0,18x - 2,4$.

Vastaus:

b) $y = 0,18x - 2,4$

7.3

Tehtävä voidaan ratkaista taulukkolaskentaohjelmalla tai GeoGebran taulukko-sovelluksella.

Ratkaistaan tehtävä GeoGebran taulukko-sovelluksella.

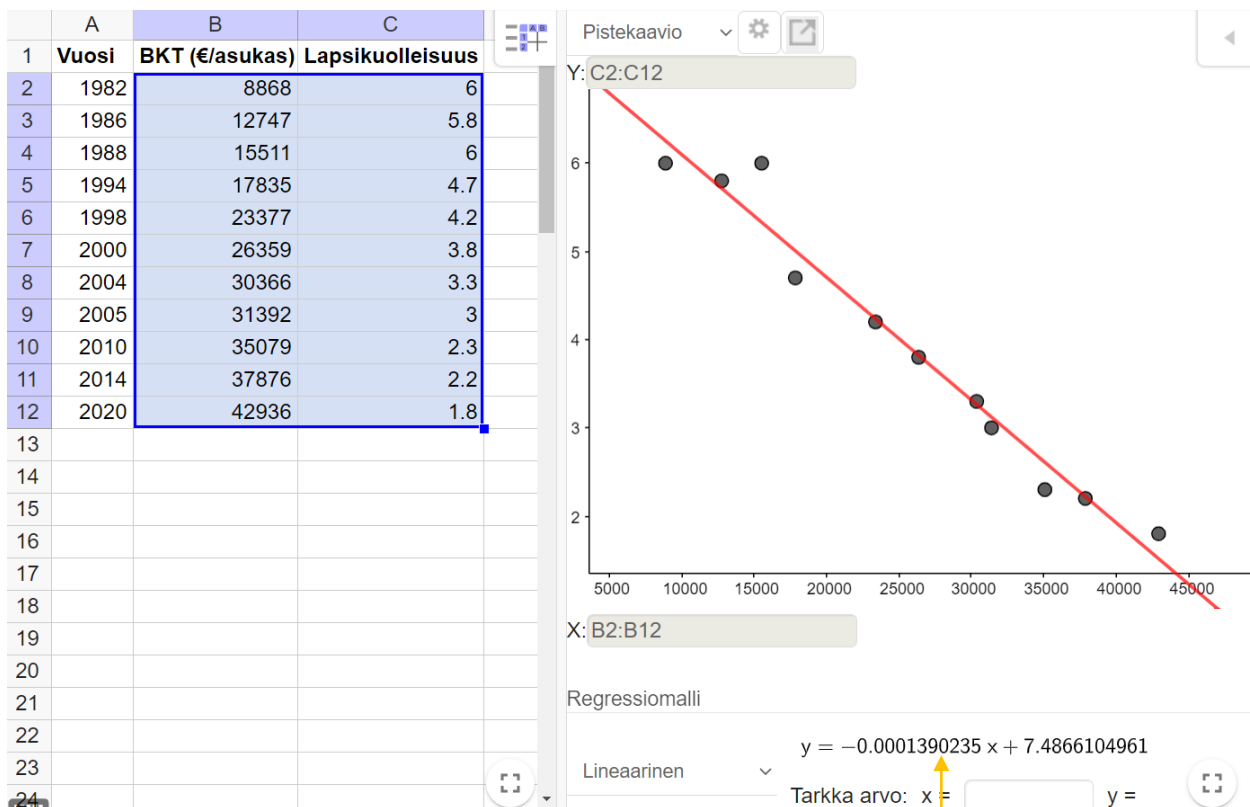
Katso tarvittaessa ohjeet Esimerkin 2 videolta.

$x = \text{BKT}$

$y = \text{lapsikuolleisuusluku}$

a)

Valitaan BKT-arvot ja lapsikuolleisuusluvut ja tehdään kahden muuttujan regressioanalyysi. Valitaan regressiomalliksi "lineaarinen".



Muuta asetuksista pyöristystarkkuudeksi "10 desimaalia".

Pyöristä lukuarvot kahden merkitsevän numeron tarkkuuteen.

Ohjelman antama regressiosuora on $y = -0,000139...x + 7,486...$

Lapsikuolleisuusluvun (y) riippuvuutta BKT-arvosta (x) kuvaa siis yhtälö

$$y = -0,00014x + 7,5.$$

b)

BKT oli 28 552 €/asukas eli $x = 28\ 552$. Lasketaan muuttujan $y = -0,00014x + 7,5$ arvo.

$$y = -0,00014 \cdot 28\ 552 + 7,5 = 3,50272 \approx 3,5$$

Vuonna 2002 lapsikuolleisuus oli noin 3,5 kuollutta lasta jokaista tuhatta syntynyttä kohden.

Huomautus:

GeoGebran taulukko-sovelluksessa on laskuri, jolla voidaan laskea ennusteita muuttujan y arvolle syöttämällä muuttujan x arvo laskuriin. Laskuri käyttää mallissa oleville kertoimille tarkkoja arvoja, mistä syystä laskurin antama tulos voi poiketa pyöristetyillä arvoilla tehdyn laskun tuloksesta.

$$y = -0.0001390235x + 7.4866104961$$

Tarkka arvo: $x = 28552$ $y = 3.5172129069$

c)

Lapsikuolleisuus oli 1,7 eli $y = 1,7$.

Muodostetaan mallin $y = -0,00014x + 7,5$ avulla yhtälö ja ratkaistaan x .

$$-0,0014x + 7,5 = 1,7 \quad \| -7,5$$

$$-0,0014x = -5,8 \quad \| :(-0,00014)$$

$$x = 41\,428,57 \dots$$

$$x \approx 41\,430$$

Mallin mukaan BKT oli vuonna 2015 noin 41 430 €/asukas.

Vastaus:

a) $y = -0,00014x + 7,5$

b) 3,5 kuollutta lasta 1000 syntynyttä kohden

c) 41 430 €/asukas

7.4

Kuviossa I pisteet sijaitsevat likimain laskevan suoran suuntaisesti. Tämä tarkoittaa, että muuttujien välillä on negatiivinen korrelaatio. Korrelaatiokerroin on negatiivinen ja lähellä lukua -1 , joten $r \approx -0,96$.

Kuviossa II pisteet sijaitsevat likimain nousevan suoran suuntaisesti. Tämä tarkoittaa, että muuttujien välillä on positiivinen korrelaatio. Korrelaatiokerroin on positiivinen ja lähellä lukua 1 , joten $r \approx 0,97$.

Kuviossa III pisteet eivät edes likimain sijaitse minkään suoran suuntaisesti, joten muuttujien välillä ei ole korrelaatiota. Korrelaatiokertoimen arvo on likimain nolla, joten $r \approx 0,01$.

Vastaus:

A-II

B-III

C-I

7.5

a)

Punnuksen massa vaikuttaa sen voiman suuruuteen, joka tarvitaan kappaleen liikuttamiseen. Kun punnuksen massa kasvaa, myös tarvittava voima kasvaa.

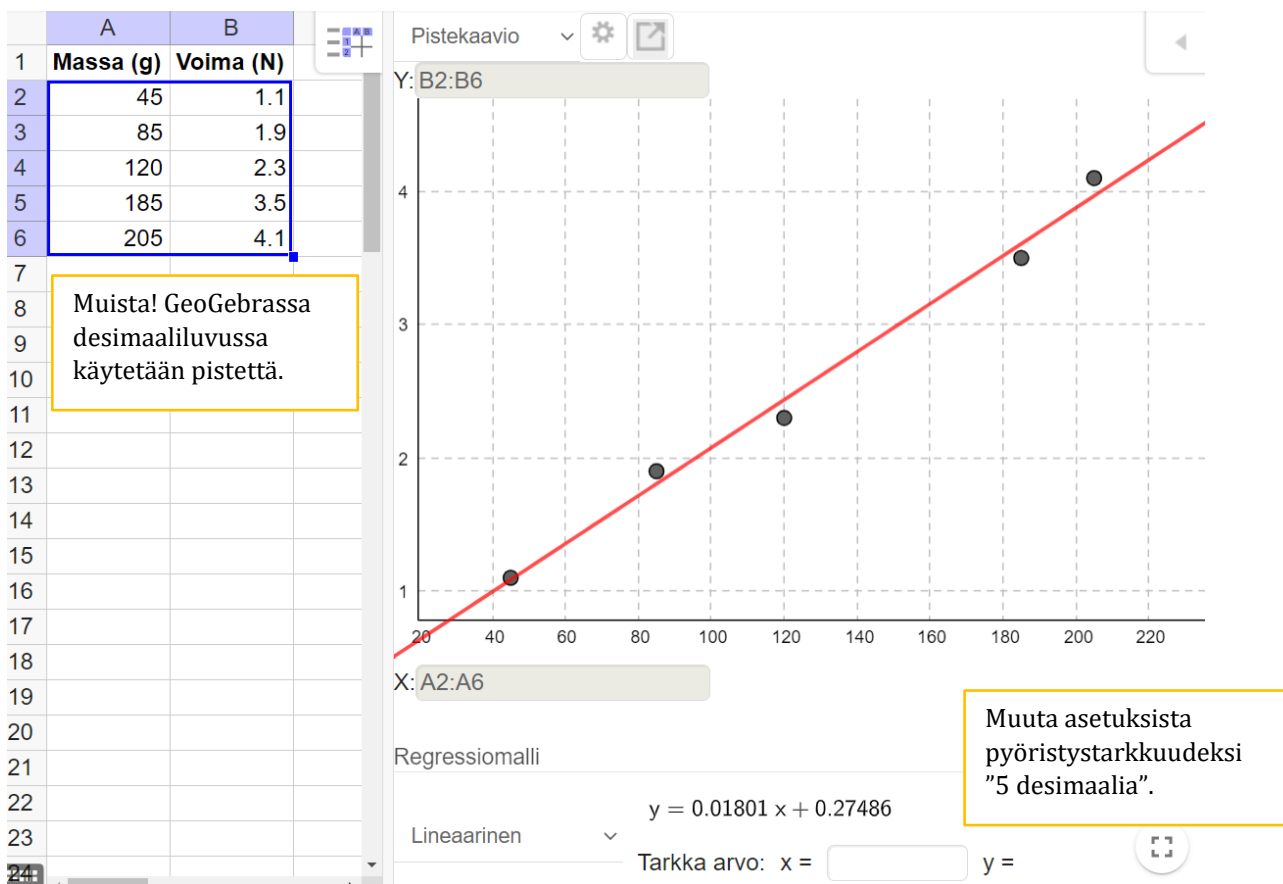
Massa on siis selittävä x-muuttuja, voima on selitettävä y-muuttuja.

b)

Ratkaistaan tehtävä GeoGebran taulukko-sovelluksella.

Katso tarvittaessa ohjeet Esimerkin 1 videolta.

Kirjoitetaan massat (x) ja voimat (y) taulukkoon. Tehdään kahden muuttujan regressioanalyysi. Valitaan regressiomalliksi "lineaarinen".



Ohjelman antama regressiosuora on $y = 0,0180\dots x + 0,274\dots$

Voiman (y) riippuvuutta punnuksen massasta (x) kuvaa siis yhtälö

$$y = 0,018x + 0,27.$$

Pyöristä lukuarvot kahden merkitsevän numeron tarkkuuteen.

Vastaus:

a) Punnuksen massa on selittävä muuttuja. Voima on selitettävä muuttuja.

b) $y = 0,018x + 0,27$

7.6

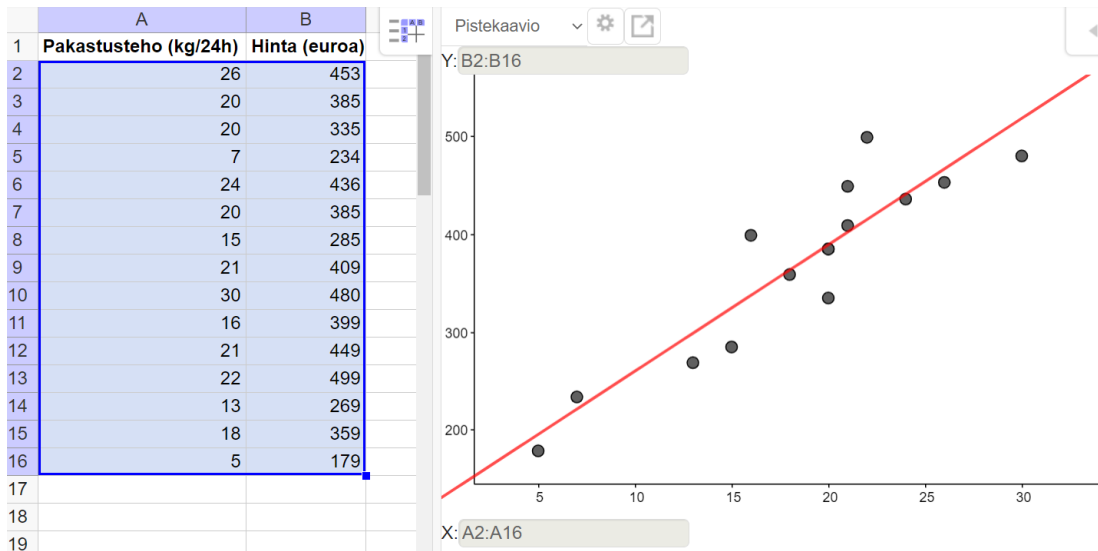
Tehtävä voidaan ratkaista taulukkolaskentaohjelmalla tai GeoGebran taulukko-sovelluksella.

Tapa 1. Ratkaistaan tehtävä GeoGebran taulukko-sovelluksella.

Katso tarvittaessa ohjeet Esimerkin 3 videolta.

a)

Valitaan aineisto, tehdään kahden muuttujan regressioanalyysi ja valitaan regressiomalliksi "lineaarinen".



Pisteet sijoittuvat likimain nousevan suoran suuntaisesti. Pakastustehon x kasvaessa myös pakastimen hinta y nousee. Muuttujien välillä on siis positiivinen yhteys.

b)

Luetaan korrelaatiokerroin arvo ohjelman yhteenvetotaulukosta.

A	B
Pakastusteho (kg/24h)	Hinta (euroa)
26	453
20	385
20	335
7	234
24	436
20	385
15	285
21	409
30	480
16	399
21	449
22	499
13	269
18	359
5	179

Tilastot	
KeskiarvoX	18.5333
KeskiarvoY	370.4
Sx	6.621
Sy	93.6504
r	0.9121
p	0.9228
Sxx	613.7333
VarianssiY	122785.6
Sxy	7917.8
R ²	0.8319
SSE	20637.7233

Korrelaatiokerroin on $r = 0,912... \approx 0,91$.

Korrelaatiokerroin on positiivinen. Kerroimen itseisarvo on $|r| \approx 0,91$.

Tulos kertoo voimakkaasta positiivisesta korrelaatiosta. Tämä tarkoittaa, että mitä suurempi arkkupakastimen pakastusteho on, sitä korkeampi myös sen hinta on.

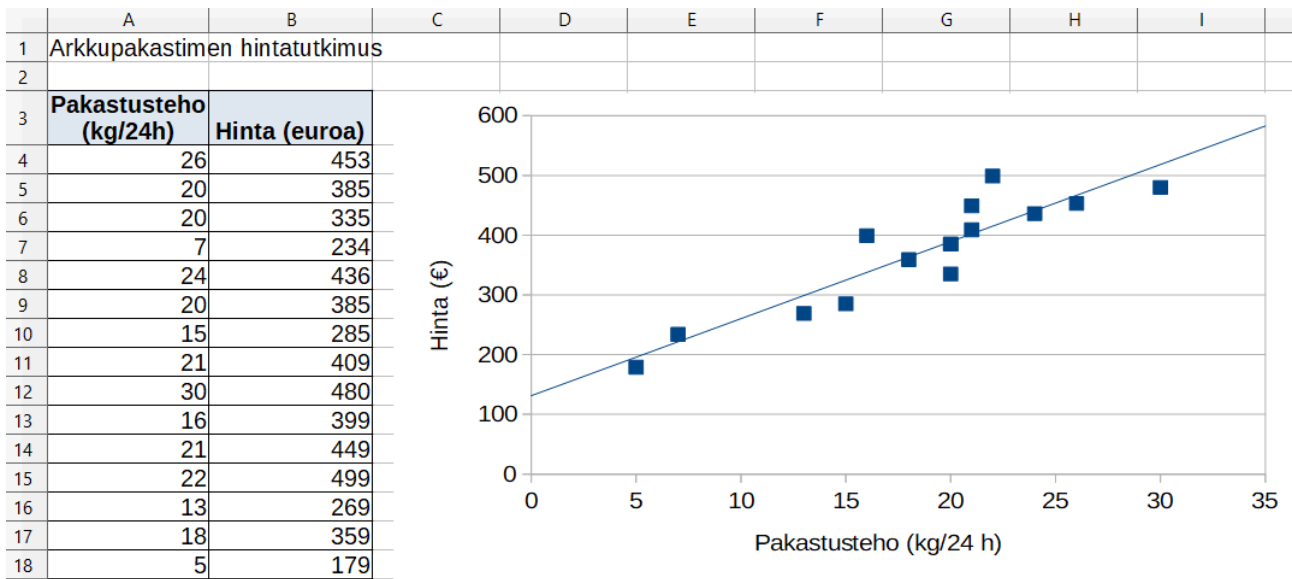
Tapa 2. Ratkaistaan tehtävä taulukkolaskentaohjelmalla
Katso tarvittaessa ohjeet Esimerkin 3 videolta.

a)

Piirretään hajontakuvio ja määritetään regressiosuora.

Ohje:

- Valitse pakastustehot ja hinnat.
- Lisää Kaavio – XY (hajonta). Lisää akseleille otsikot.
- Lisää trendiviiva. Valitse "lineaarinen".



Pisteet sijoittuvat likimain nousevan suoran suuntaisesti. Pakastustehon x kasvaessa myös hinta y nousee. Muuttujien välillä on siis positiivinen yhteys.

b)

Määritetään muuttujien välinen korrelaatiokerroin tilastotoiminnoilla.

	A	B	C	D	E	F
1	Arkkupakastimen hintatutkimus					
2						
3	Pakastusteho (kg/24h)	Hinta (euroa)				
4	26	453	Korrelaatiot Sarake 1	Sarake 2		
5	20	385	Sarake 1	1		
6	20	335	Sarake 2	0,9120968	1	
7	7	234				
8	24	436				
9	20	385				
10	15	285				
11	21	409				
12	30	480				
13	16	399				
14	21	449				
15	22	499				
16	13	269				
17	18	359				
18	5	179				

Ohje:

- Valitse aineisto.
- Valitse Data
 - Tilastotiedot
 - Korrelaatio

Korrelaatiokerroin on $r = 0,9120... \approx 0,91$.

Korrelaatiokertoimen arvo on positiivinen. Kertoimen itseisarvo on $|r| \approx 0,91$.

Tulos kertoo voimakkaasta positiivisesta korrelaatiosta. Tämä tarkoittaa, että mitä suurempi arkkupakastimen pakastusteho on, sitä korkeampi myös sen hinta on.

Vastaus:

a) Pakastustehon kasvaessa hinta nousee. Muuttujien välillä on positiivinen yhteys.

b) $r \approx 0,91$. Tulos kertoo voimakkaasta positiivisesta korrelaatiosta.

7.7

Ratkaistaan tehtävä GeoGebran taulukko-sovelluksessa.

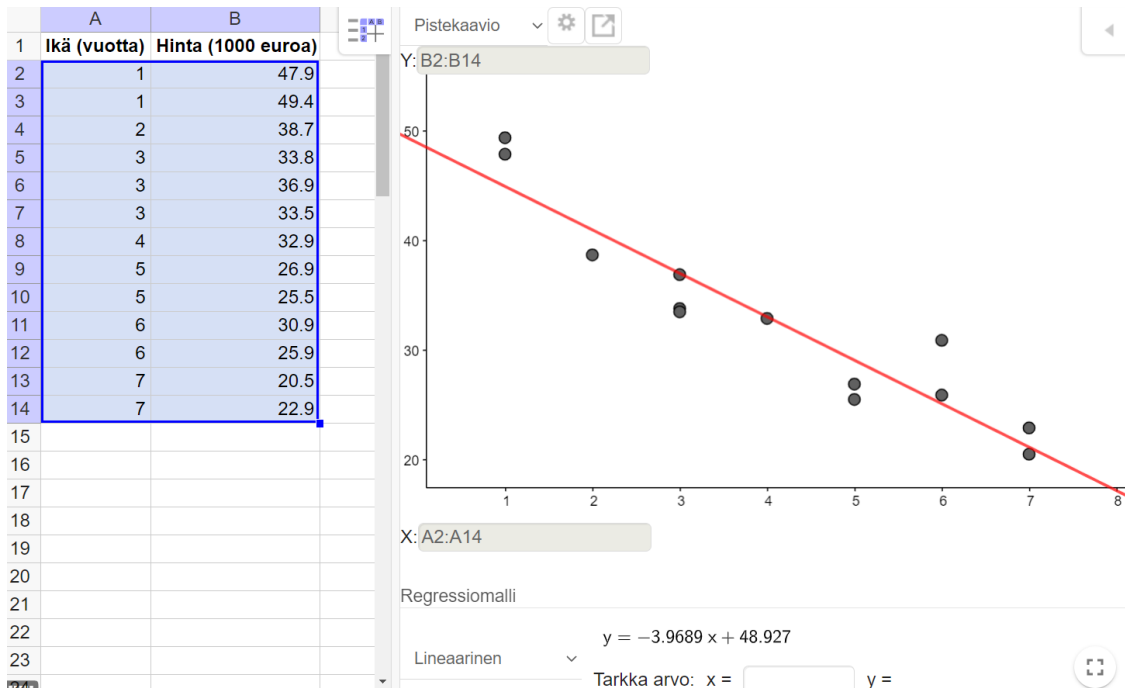
x = auton ikä

y = auton hinta

a)

Valitaan aineisto ja tehdään kahden muuttujan regressioanalyysi.

Valitaan regressiomalliksi "lineaarinen".



Ohjelman antama regressiosuora on $y = -3,968...x + 48,92....$

Auton hinnan (y) riippuvuutta iästä (x) kuvaa siis yhtälö $y = -4,0x + 49$.

b)

Luetaan korrelaatiokerroin ohjelman tulosteesta.

Korrelaatiokerroin on $r = -0,939... \approx -0,94$.

Korrelaatiokerroin on negatiivinen. Tämä kertoo, että auton iän (x) kasvaessa sen hinta (y) laskee.

Korrelaatiokertoimen itseisarvo on $|r| \approx |-0,94| = 0,94$.

Tulos kertoo, että muuttujien välillä on voimakas negatiivinen korrelaatio.

c)

Auton ikä on 3,5 vuotta eli $x = 3,5$. Lasketaan muuttujan $y = -4,0x + 49$ arvo.

$$y = -4,0 \cdot 3,5 + 49 = 35$$

Kun auto on 3,5 vuotta vanha, sen ikä on mallin mukaan 35 000 euroa.

Tilastot	
KeskiarvoX	4.0769
KeskiarvoY	32.7462
Sx	2.1001
Sy	8.8754
r	-0.9391
ρ	-0.9556
Sxx	52.9231
VarianssiY	945.2723
Sxy	-210.0462
R^2	0.8819
SSE	111.6211

Huomautus:

GeoGebran taulukko-sovelluksessa on laskuri, jolla voidaan laskea ennusteita muuttujan y arvolle syöttämällä muuttujan x arvo laskuriin. Laskuri käyttää mallissa oleville kertoimille tarkkoja arvoja, mistä syystä laskurin antama tulos voi poiketa pyöristetyillä arvoilla tehdyn laskun tuloksesta.

$$y = -3.9689x + 48.927$$

Tarkka arvo: $x =$ $y = 35.0359$

c)

Hinta on 15 000 euroa eli $y = 15$.

Muodostetaan mallin $y = -4,0x + 49$ avulla yhtälö ja ratkaistaan x .

$$-4,0x + 49 = 15 \quad \| -49$$

$$-4,0x = -34 \quad \| :(-4,0)$$

$$x = 8,5$$

Mallin mukaan 15 000 eurolla saa 8,5 vuotta vanhan auton.

Vastaus:

a) $y = -4,0x + 49$

b) $r \approx -0,94$. Auton iän ja hinnan välillä on voimakas negatiivinen korrelaatio.

c) 35 000 euroa

d) 8,5 vuotta vanhan

7.8

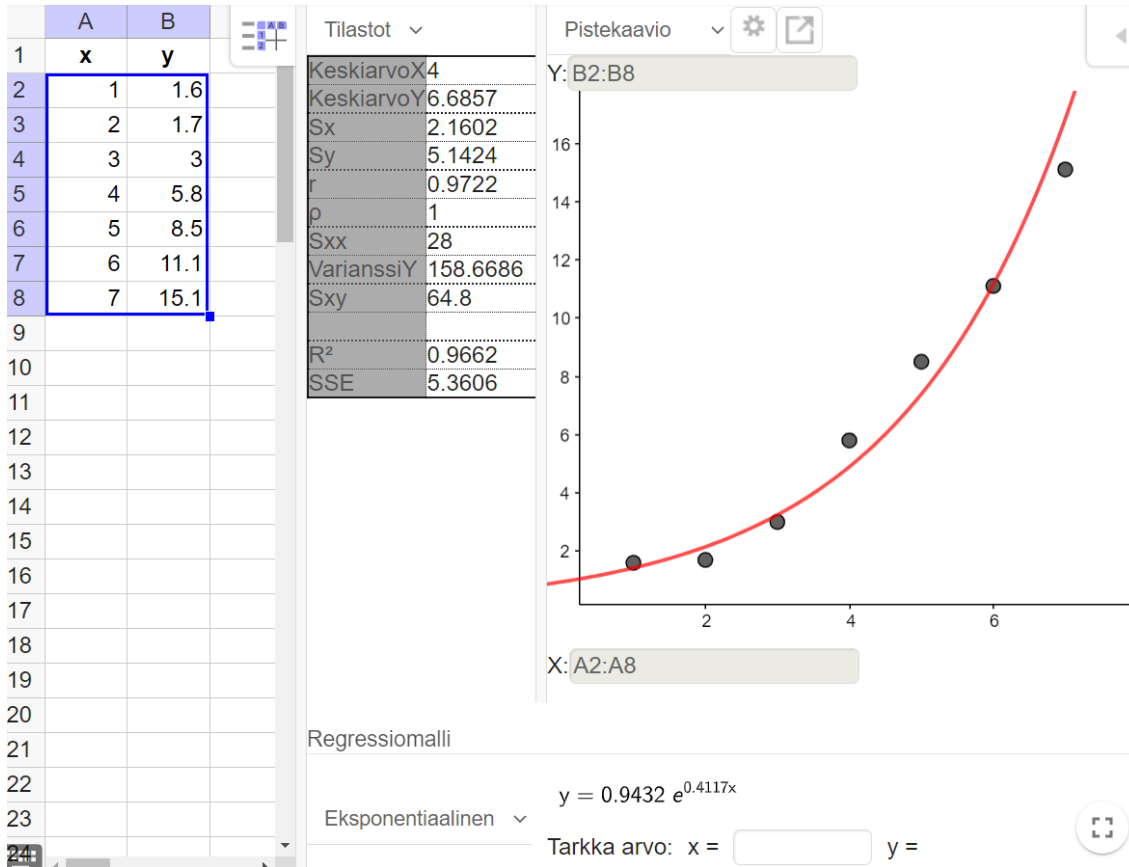
Tehtävä voidaan ratkaista GeoGebran taulukko-ohjelmalla tai taulukkolaskentaohjelmalla.

Tapa 1. Ratkaistaan tehtävä GeoGebran taulukko-sovelluksessa.

a)

Valitaan arvot ja tehdään kahden muuttujan regressioanalyysi.

Valitaan regressiomalliksi "eksponentiaalinen". Valitaan "Näytä tilastot".

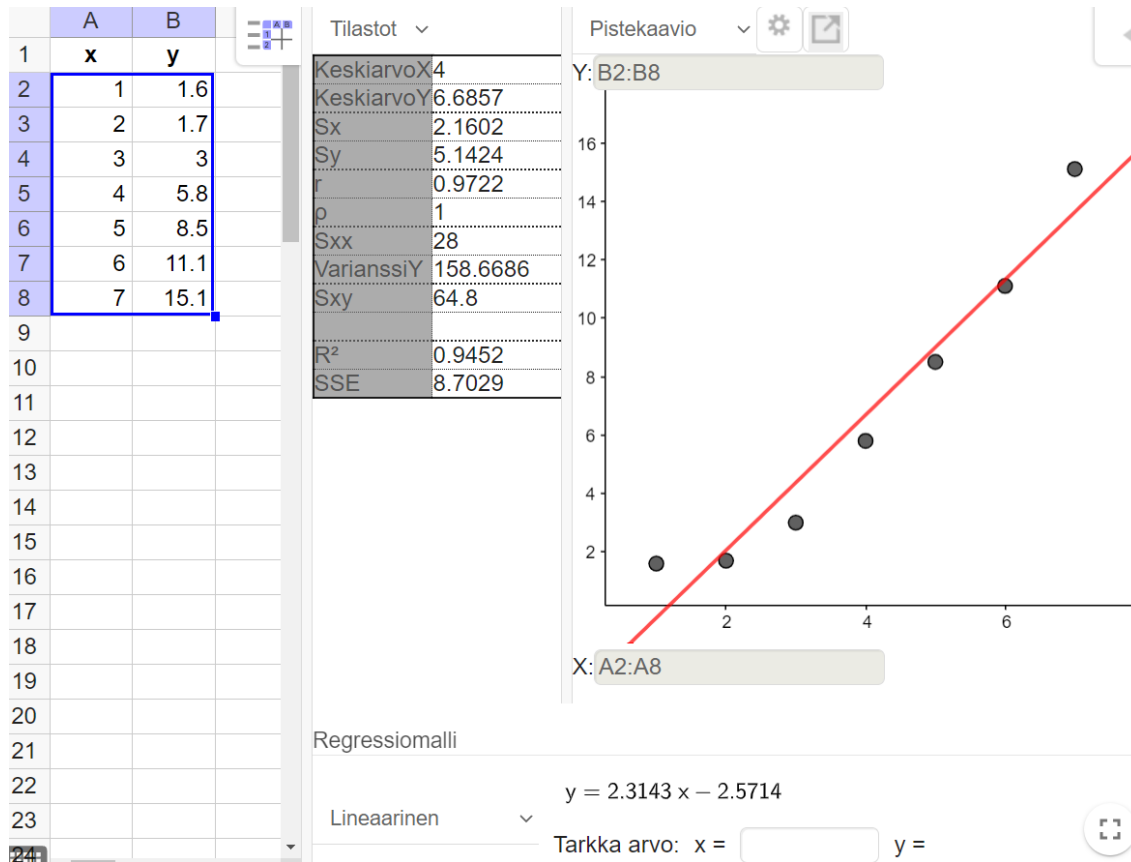


Eksponentiaalinen malli on $y = 0,94 \cdot e^{0,41x}$.

Mallin selitysaste on $R^2 = 0,966... \approx 0,97 = 97\%$.

b)

Vaihdetaan regressiomalliksi "lineaarinen".



Lineaarinen malli on $y = 2,3x - 2,6$.

Mallin selitysaste on $R^2 = 0,945... \approx 0,95 = 95 \%$.

c)

a-kohdan mallin selitysaste on hieman parempi (97 % > 95 %), joten a-kohdan malli kuvaa muuttujien välistä riippuvuutta hieman paremmin.

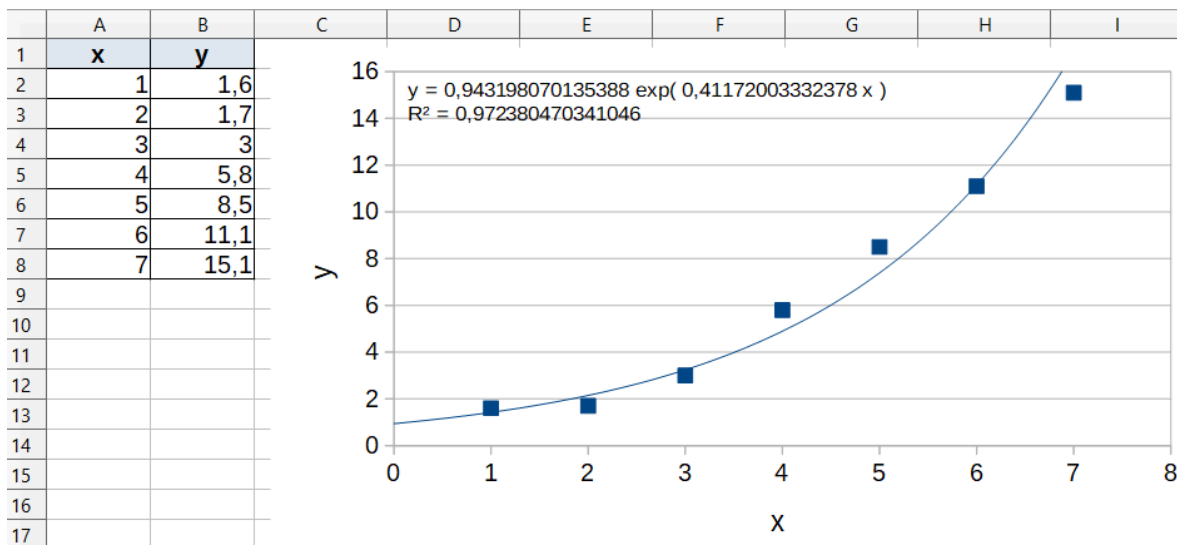
Tapa 2. Ratkaistaan tehtävä taulukkolaskentaohjelmalla.

a)

Piirretään hajontakuviota ja sovitetaan eksponentiaalinen malli.

Ohje:

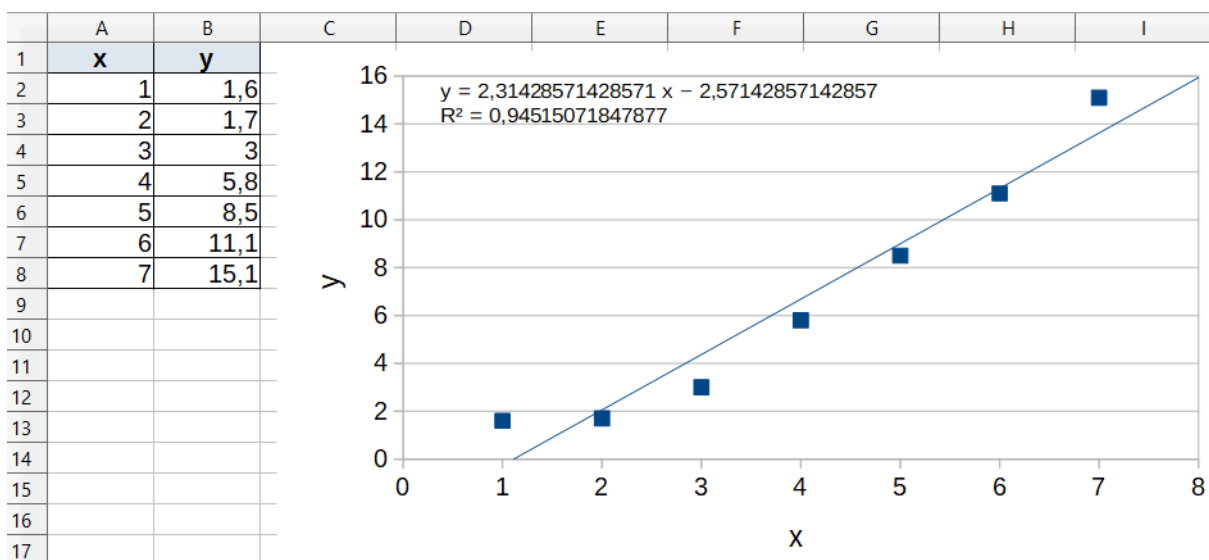
- Valitse muuttujien arvot.
- Lisää Kaavio - XY (hajonta). Lisää akseleille otsikot.
- Lisää trendiviiva. Valitse "eksponentiaalinen".
Valitse "Näytä yhtälö" ja "Näytä selitysaste", ja valitse muuttujan y nimeksi y.



Eksponentiaalinen malli on $y = 0,94 \cdot e^{0,41x}$. Mallin selitysaste on $R^2 = 0,972 \dots \approx 0,97$.

b)

Vaihdetaan trendiviivaksi "lineaarinen".



Lineaarinen malli on $y = 2,3x - 2,6$. Mallin selitysaste on $R^2 = 0,945 \dots \approx 0,95$.

c)

a-kohdan mallin selitysaste on hieman parempi ($0,97 > 0,95$), joten a-kohdan malli kuvaa muuttujien välistä riippuvuutta hieman paremmin.

Vastaus:

a) $y = 0,94 \cdot e^{0,41x}$ ja $R^2 \approx 0,97$

b) $y = 2,3x - 2,6$ ja $R^2 \approx 0,95$

c) a-kohdan eksponentiaalinen malli

7.9

a)

Täydellinen positiivinen korrelaatio tarkoittaa, että havaintopisteet asettuvat täydellisesti jonkin nousevan suoran suuntaisesti.

Esimerkki:

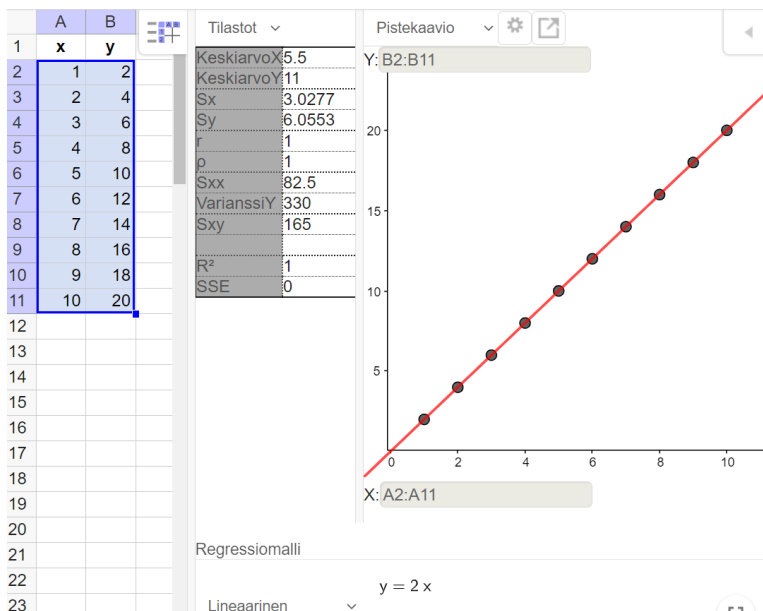
x	y
1	2
2	4
3	6
4	8
5	10
6	12
7	14
8	16
9	18
10	20

Perustellaan, että muuttujien x ja y välillä on täydellinen korrelaatio ($r = 1$).

Tapa 1. Muuttujan y arvo on joka tilanteessa kaksinkertainen muuttujan x arvoon verrattuna eli $y = 2x$. Havaintopisteet ovat siis suoran $y = 2x$ pisteitä, joten korrelaatio on täydellinen.

Tapa 2.

Tehdään aineistolle kahden muuttujan analyysi GeoGebran taulukko-sovelluksessa.



Hajontakuviosta nähdään, että pisteet asettuvat täydellisesti suoralle $y = 2x$. Muuttujien välinen korrelaatiokerroin on $r = 1$.

b)

Muuttujien välillä ei ole korrelaatiota silloin, jos havaintopisteet eivät asetu minkään nousevan tai laskevan suoran suuntaisesti.

Esimerkki:

x	y
1	2
2	2
3	2
4	2
5	2
6	2
7	2
8	2
9	2
10	2

Muuttujan x arvon kasvaessa muuttujan y arvo ei muutu mihinkään. Havaintopisteet asettuvat suoralle $y = 2$.

Muuttujan x arvon muutoksella ei ole yhteyttä muuttujan y arvon muutokseen eli muuttujien välillä ei ole korrelaatiota ($r = 0$).

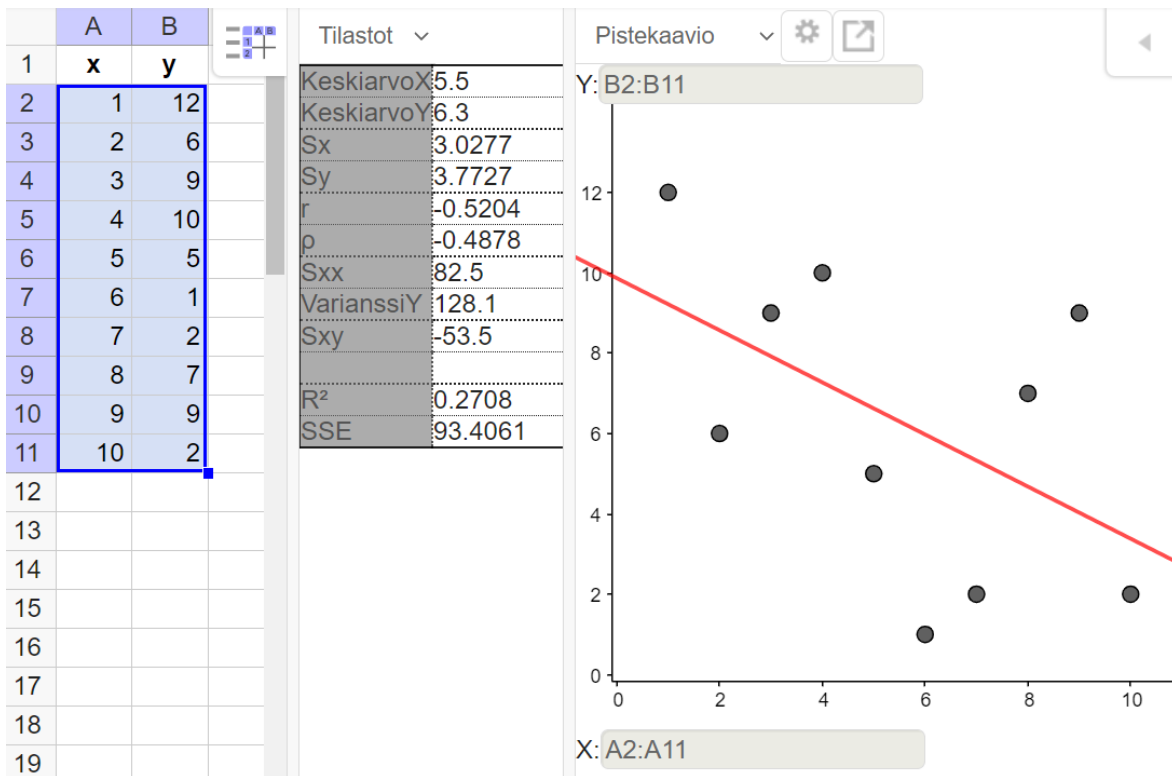
c)

Muuttujien välillä on kohtalainen negatiivinen korrelaatio, jos muuttujan x arvon kasvaessa muuttujan y arvo pienenee, ja hajontakuviassa pisteparven suunta on laskeva.

Kohtalainen korrelaatio tarkoittaa, että korrelaatiokertoimen itseisarvo on välillä 0,3–0,6.

Kokeilemalla esimerkiksi GeoGebran taulukko-sovelluksessa löydetään ehdot toteuttava esimerkki:

x	y
1	12
2	6
3	9
4	10
5	5
6	1
7	2
8	7
9	9
10	2



Hajontakuviosta nähdään, että esimerkin havaintopisteet asettuvat likimain laskevan suoran suuntaisesti.

Muuttujien välinen korrelaatiokerroin on $r = -0,520 \dots \approx -0,52$.

Korrelaatiokertoimen itseisarvo on $|r| \approx 0,52$ eli välillä 0,3–0,6.

Tämä tarkoittaa, että muuttujien välinen korrelaatio on kohtalainen.

7.10

a)

Sidokseen sitoutuneen energian suuruus riippuu sidoksen pituudesta. Kun sidoksen pituus kasvaa, sidokseen sitoutuneen energian määrä pienenee.

x = sidoksen pituus (selittävä muuttuja)

y = sidokseen sitoutunut energia (selitettävä muuttuja)

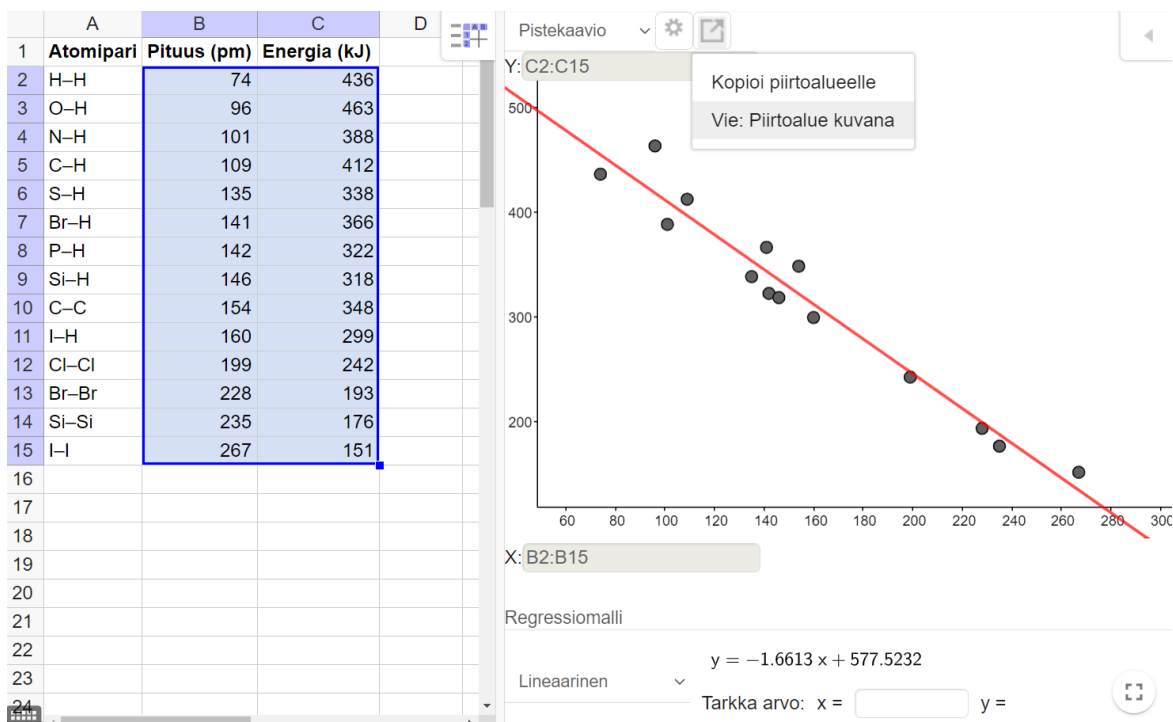
b)

Ratkaistaan tehtävä GeoGebran taulukko-sovelluksella.

Katso tarvittaessa ohjeet Esimerkin 2 videolta.

Valitaan pituudet ja energiat ja tehdään kahden muuttujan regressioanalyysi.

Valitaan regressiomalliksi "lineaarinen".



Ohjelman antama regressiosuora on $y = -1,661...x + 577,53....$

Energian (y) riippuvuutta sidoksen pituudesta (x) kuvaa siis yhtälö

$$y = -1,7x + 580.$$

Pyöristä lukuarvot kahden merkitsevän numeron tarkkuuteen.

c)

Sidoksen pituus on 128 pm eli $x = 128$. Lasketaan muuttujan $y = -1,7x + 580$ arvo.

$$y = -1,7 \cdot 128 + 580 = 362,4 \approx 362$$

Kloorin ja vedyn väliseen sidokseen sitoutuu mallin mukaan 362 kJ energiaa.

Huomautus:

GeoGebran taulukko-sovelluksessa on laskuri, jolla voidaan laskea ennusteita muuttujan y arvolle syöttämällä muuttujan x arvo laskuriin. Laskuri käyttää mallissa oleville kertoimille tarkkoja arvoja, mistä syystä laskurin antama tulos voi poiketa pyöristetyillä arvoilla tehdyn laskun tuloksesta.

$$y = -1.6613x + 577.5232$$

Tarkka arvo: $x =$ $y =$ 364.8732

c)

Sidoksessa on energiaa 163 kJ eli $y = 163$.

Muodostetaan mallin $y = -1,7x + 580$ avulla yhtälö ja ratkaistaan x .

$$-1,7x + 580 = 163 \quad || -580$$

$$-1,7x = -417 \quad ||: (-1,7)$$

$$x = 245,29 \dots$$

$$x \approx 245$$

N-N-sidoksen pituus on mallin mukaan 245 pm.

Vastaus:

a) Sidoksen pituus on selittävä muuttuja. Sidokseen sitoutuneen energian määrä on selitettävä muuttuja.

b) $y = -1,7x + 580$

c) 362 kJ

d) 245 pm

7.11

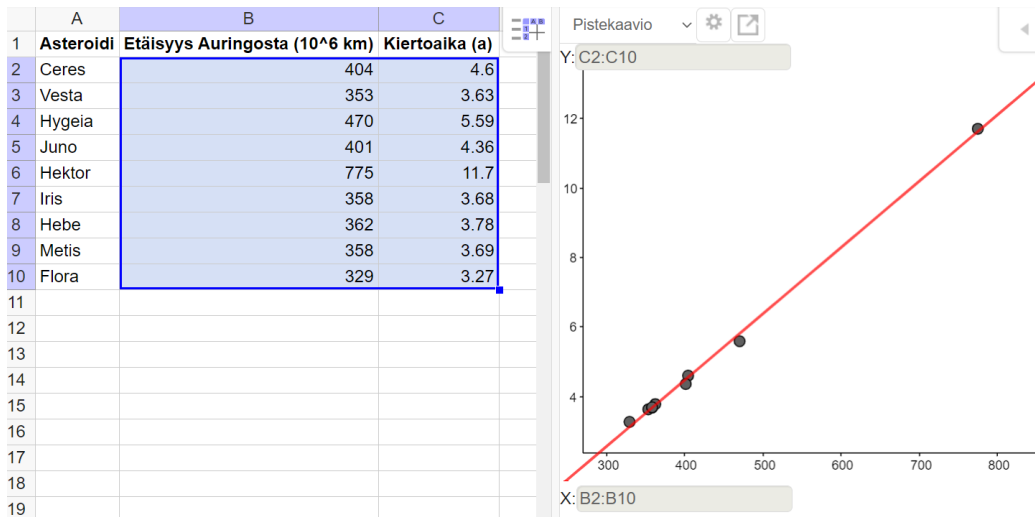
Tehtävä voidaan ratkaista taulukkolaskentaohjelmalla tai GeoGebran taulukko-sovelluksella.

Tapa 1. Ratkaistaan tehtävä GeoGebran taulukko-sovelluksella

Katso tarvittaessa ohjeet Esimerkin 3 videolta.

a)

Valitaan etäisyydet ja kiertoaajat, tehdään kahden muuttujan regressioanalyysi ja valitaan regressiomalliksi "lineaarinen".



Pisteet sijoittuvat melko tarkasti nousevan suoran suuntaisesti. Etäisyyden x kasvaessa myös kiertoaika y kasvaa. Muuttujien välillä on siis positiivinen yhteys.

b)

Luetaan korrelaatiokerroin ohjelman yhteenvotaulukosta.

Asteroidi	Etäisyys Auringosta (10 ⁶ km)	Kiertoaika (a)
Ceres	404	4.6
Vesta	353	3.63
Hygeia	470	5.59
Juno	401	4.36
Hektor	775	11.7
Iris	358	3.68
Hebe	362	3.78
Metis	358	3.69
Flora	329	3.27

Tilastot	
KeskiarvoX	423.3333
KeskiarvoY	4.9222
Sx	138.2317
Sy	2.6356
r	0.9991
p	0.9958
Sxx	152864
VarianssiY	55.57
Sxy	2911.9333
R ²	0.9982
SSE	0.1

Korrelaatiokerroin on $r = 0,999... \approx 1,00$. Korrelaatiokertoimen arvo on lähes 1.

Tulos kertoo lähes täydellisestä lineaarisesta riippuvuudesta eli pisteet sijoittuvat lähes täydellisesti suoralle. Tämä tarkoittaa, että kiertoaika voidaan ennustaa melko tarkasti etäisyyden perusteella lineaarisella mallilla (regressiosuoran avulla).

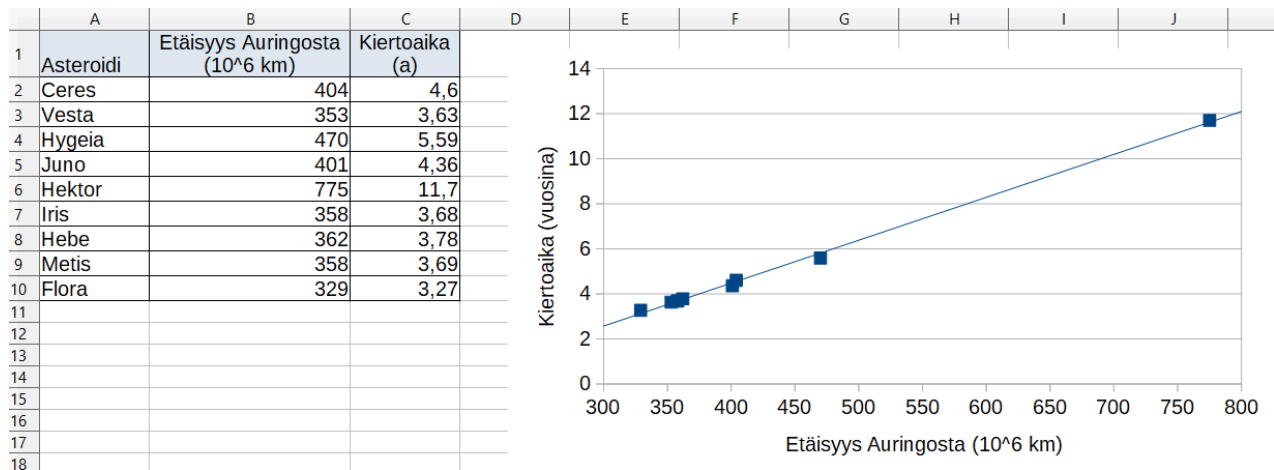
Tapa 2. Ratkaistaan tehtävä taulukkolaskentaohjelmalla
Katso tarvittaessa ohjeet Esimerkin 3 videolta.

a)

Piirretään hajontakuviota ja määritetään regressiosuora.

Ohje:

- Valitse etäisyydet ja kiertoaajat.
- Lisää Kaavio - XY (hajonta). Lisää akselleille otsikot.
- Lisää trendiviiva. Valitse "lineaarinen".



Pisteet sijoittuvat melko tarkasti nousevan suoran suuntaisesti. Etäisyyden x kasvaessa myös kiertoaika y kasvaa. Muuttujien välillä on siis positiivinen yhteys.

b)

Määritetään muuttujien välinen korrelaatiokerroin tilastotoiminnoilla.

	A	B	C	D	E	F	G
1	Asteroidi	Etäisyys Auringosta (10 ⁶ km)	Kiertoaika (a)				
2	Ceres	404	4,6			Korrelaatiot Sarake 1	Sarake 2
3	Vesta	353	3,63			Sarake 1	1
4	Hygeia	470	5,59			Sarake 2	0,9990996
5	Juno	401	4,36				1
6	Hektor	775	11,7				
7	Iris	358	3,68				
8	Hebe	362	3,78				
9	Metis	358	3,69				
10	Flora	329	3,27				

Ohje:

- Valitse aineisto.
- Valitse Data
 - Tilastotiedot
 - Korrelaatio

Korrelaatiokerroin on $r = 0,999... \approx 1,00$. Korrelaatiokertoimen arvo on lähes 1.

Tulos kertoo lähes täydellisestä lineaarisesta riippuvuudesta. Tämä tarkoittaa, että kiertoaika voidaan ennustaa melko tarkasti etäisyydestä lineaarisella mallilla (regressiosuoran avulla).

Vastaus:

a) Etäisyyden kasvaessa kiertoaika kasvaa. Muuttujien välillä on positiivinen yhteys.

b) $r \approx 1,00$. Tulos kertoo täydellisestä lineaarisesta yhteydestä, joten kiertoaika voidaan arvioida etäisyyden perusteella lineaarisella mallilla.

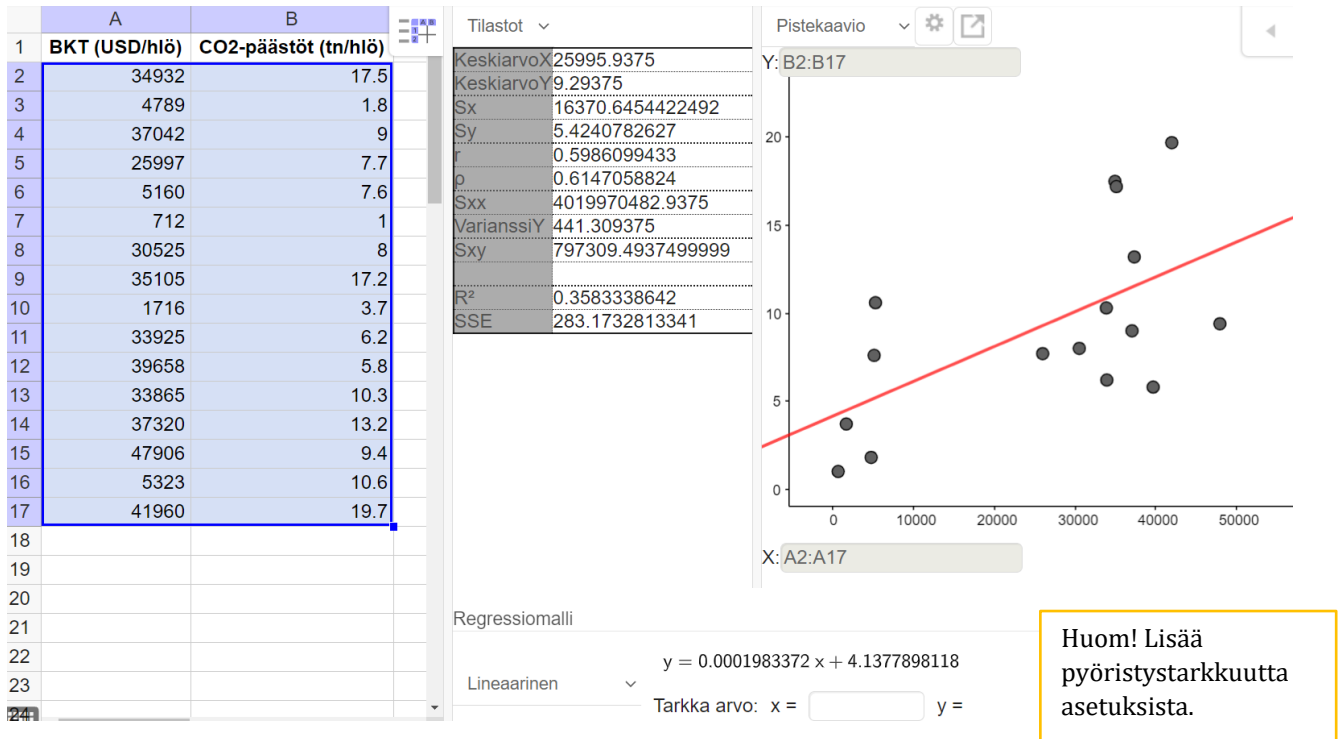
7.12

Ratkaistaan tehtävä GeoGebran taulukko-sovelluksella.

$x = \text{BKT}$

$y = \text{CO}_2\text{-päästöt}$

Valitaan aineisto ja tehdään kahden muuttujan regressioanalyysi.



a)

Hajontakuvio on näkyvissä ohjelman tulosteessa.

Ohjelman antama regressiosuora on $y = 0,000198...x + 4,137...$

Riippuvuutta kuvaa siis yhtälö $y = 0,00020x + 4,1$.

b)

Ohjelman tulosteen perusteella korrelaatiokerroin on $r = 0,5986... \approx 0,60$.

Tuloksen perusteella bruttokansantuotteen (BKT) ja hiilidioksidipäästöjen välillä on kohtalaisen ja huomattavan rajalla oleva positiivinen korrelaatio. BKT-arvon noustessa myös hiilidioksidipäästöt keskimäärin nousevat.

c)

BKT on 25 000 USD/hlö eli $x = 25\ 000$ (USD). Lasketaan muuttujan $y = 0,00020x + 4,1$ arvo.

$$y = 0,00020 \cdot 25\ 000 + 4,1 = 9,1$$

Hiilidioksidipäästöt ovat 9,1 tonnia/henkilö.

$$y = 0.0001983372x + 4.1377898118$$

Tarkka arvo: x = y =

Huomautus:

GeoGebran taulukko-sovelluksessa on laskuri, jolla voidaan laskea ennusteita muuttujan y arvolle syöttämällä muuttujan x arvo laskuriin. Laskuri käyttää mallissa oleville kertoimille tarkkoja arvoja, mistä syystä laskurin antama tulos voi poiketa pyöristetyillä arvoilla tehdyn laskun tuloksesta.

Vastaus:

a) $y = 0,00020x + 4,1$

b) $r \approx 0,60$, muuttujien välillä on kohtalainen/huomattava positiivinen korrelaatio

c) 9,1 tonnia/henkilö

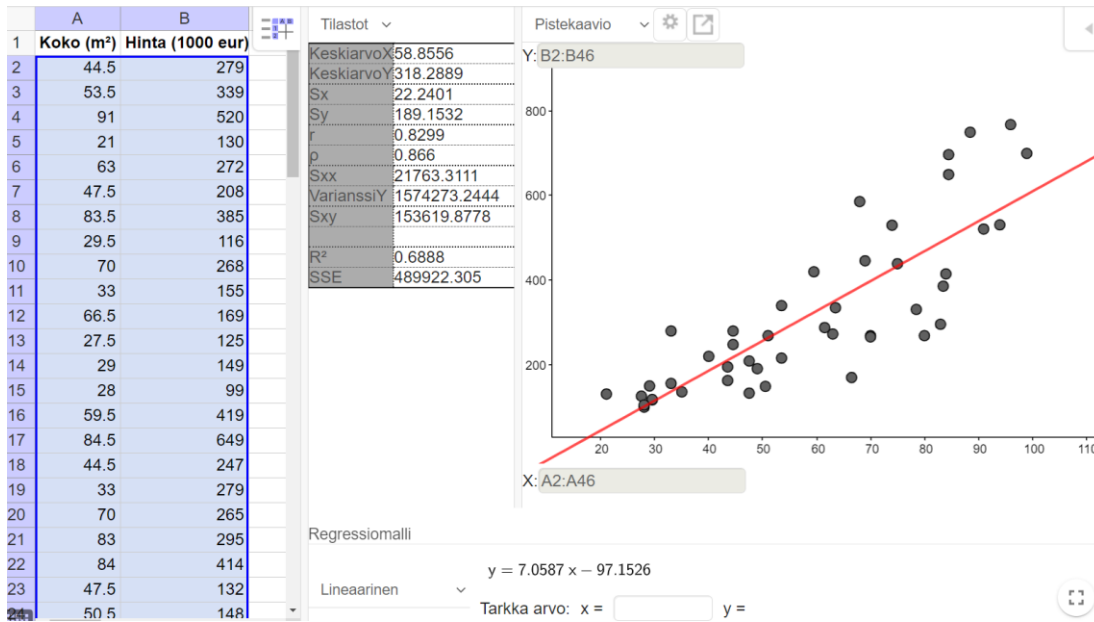
7.13

Ratkaistaan tehtävä GeoGebran taulukko-sovelluksessa.

x = asunnon koko

y = asunnon hinta

Valitaan aineisto ja tehdään kahden muuttujan regressioanalyysi.



a)

Hajontakuvio on näkyvissä ohjelman tulosteessa.

Ohjelman antama regressiosuora on $y = 7,058...x - 97,15...$

Riippuvuutta kuvaa siis yhtälö $y = 7,1x - 97$.

b)

Ohjelman tulosten perusteella korrelaatiokerroin on $r = 0,829... \approx 0,83$.

Tuloksen perusteella kerrostaloasunnon koon (m²) ja myyntihinnan (tuhatta euroa) välillä on voimakas positiivinen korrelaatio. Asunnon koon kasvaessa myös sen hinta nousee.

c)

Asunnon koko on 50 m², joten $x = 50$. Lasketaan muuttujan $y = 7,1x - 97$ arvo.

$$y = 7,1 \cdot 50 - 97 = 258$$

Tämän kokoisen asunnon hinta on mallin mukaan 258 000 euroa.

Huomautus:

GeoGebran taulukko-sovelluksessa on laskuri, jolla voidaan laskea ennusteita muuttujan y arvolle syöttämällä muuttujan x arvo laskuriin. Laskuri käyttää mallissa oleville kertoimille tarkkoja arvoja, mistä syystä laskurin antama tulos voi poiketa pyöristetyillä arvoilla tehdyn laskun tuloksesta.

$$y = 7.0587x - 97.1526$$

Tarkka arvo: x = y =

d)

Hinta on 300 000 eli $y = 300$. Muodostetaan mallin $y = 7,1x - 97$ avulla yhtälö ja ratkaistaan x .

$$7,1x - 97 = 300 \quad || +97$$

$$7,1x = 397 \quad || :7,1$$

$$x = 55,91 \dots$$

$$x \approx 56$$

Mallin mukaan 300 000 eurolla saa 56-neliöisen asunnon.

Vastaus:

a) $y = 7,1x - 97$

b) $r \approx 0,83$. Muuttujien välillä on voimakas positiivinen korrelaatio.

c) 258 000 euroa

d) 56 m²

7.14

Ratkaistaan tehtävä GeoGebran taulukko-sovelluksessa.

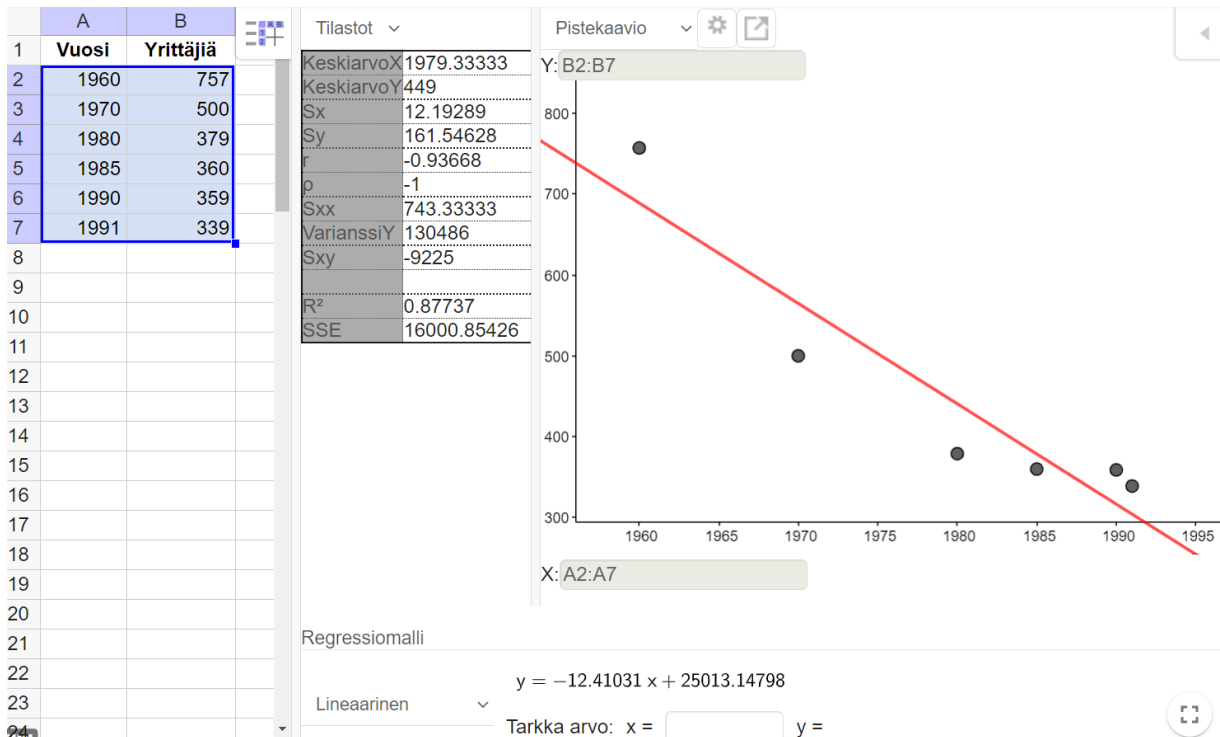
x = vuosiluku

y = yrittäjien lukumäärä

Kirjoitetaan vuodet ja yrittäjien lukumäärät taulukkoon.

Valitaan aineisto ja tehdään kahden muuttujan regressioanalyysi.

Valitaan regressiomalliksi "lineaarinen".



a)

Sirontakuviokuva näkyy ohjelman tulosteissa. Sen perusteella yrittäjien määrä on aikavälillä 1960–1991 ollut laskussa.

b)

Ohjelman antama regressiosuora on $y = -12,41...x + 25013,1...$

Riippuvuutta kuvaa siis yhtälö $y = -12,4x + 25\ 000$.

Ohjelman tulosten perusteella korrelaatiokerroin on $r = -0,9366... \approx -0,937$.

Tulos ilmaisee, että vuosiluvun ja yrittäjien lukumäärän välillä on voimakas negatiivinen korrelaatio, ja regressiosuora sopii hyvin kuvaamaan muuttujien välistä riippuvuutta.

c)

Tarkastellaan vuotta 2019, joten $x = 2019$. Lasketaan muuttujan $y = -12,4x + 25\ 000$ arvo.

$$y = -12,4 \cdot 2019 + 25\ 000 = -35,6$$

Mallin mukaan yrittäjien lukumäärä vuonna 2019 olisi negatiivinen, -35 600 kappaletta.

Huomaa: GeoGebran laskuri laskee tarkoilla arvoilla, joten laskurin tulos poikkeaa pyöristetyillä arvoilla lasketusta tuloksesta.

$$y = -12.41031x + 25013.14798$$

Tarkka arvo: $x = 2019$ $y = -43.27578$

d)

b-kohdan malli antaa järjettömän ennusteen yrittäjien määrälle vuonna 2019. Malli perustuu tietoihin yrittäjien määrästä vuosina 1960–1991, ja kuvaa yrittäjien määrän kehittymistä tällä ajanjaksolla. Malli ei c-kohdan perusteella sovellu yrittäjien määrän ennustamiseen ainakaan kovin pitkälle vuodesta 1991 eteenpäin.

Vastaus:

a) Yrittäjien määrä on aikavälillä 1960–1991 ollut laskussa.

b) $y = -12,4x + 25\,000$ ja $r \approx 0,937$. Muuttujien välillä on voimakas negatiivinen korrelaatio.

c) $-35\,600$ (järjetön ennuste)

d) Malli ei sovellu yrittäjien määrän ennustamiseen ainakaan kovin pitkälle vuodesta 1991 eteenpäin.

7.15

Ratkaistaan tehtävä taulukkolaskentaohjelmalla.

Lasketaan 100 metrin juoksun ja kuulantyyntö tulosten summa sarakkeeseen F.

	A	B	C	D	E	F	
1	Sydneysn olympialaiset, kymmenottelu						
2	10 parhaan urheilijan pistetulokset.						Pistesumma
3							1729
4		100 m juoksu	Kuulantyyntö	Kokonaispistemäärä		Pistesumma	1681
5	1. Erki Nool EST	933	796	8641		=B5+C5	1786
6	2. Roman Sebrle CZE	878	803	8606			1669
7	3. Chris Huffins USA	980	806	8595			1683
8	4. Dean Macey GBR	903	766	8567			1727
9	5. Tom Pappas USA	901	782	8425			1641
10	6. Tomas Dvorak CZE	881	846	8385			1625
11	7. Frank Busemann GER	881	760	8351			1605
12	8. Attila Zsivoczky HUN	838	787	8277			1605
13	9. Stefan Schmid GER	874	731	8206			1685
14	10. Henrik Dagård SWE	897	788	8178			

Piirretään havainnoista korrelaatiodiagrammi eli hajontakuvio.

Määritetään pistesumman ja kokonaispistemäärän korrelaatiokerroin.

Ohje:

Huomaa: tehtävänannon perusteella

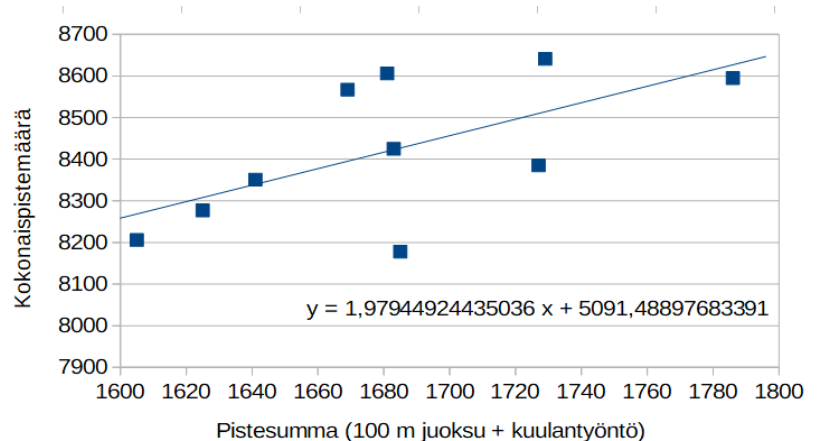
x = pistesumma (selittävä muuttuja)

y = kokonaispistemäärä (selitettävä muuttuja).

- Kopioi kokonaispistemäärät pistesummien viereiseen sarakkeeseen.
- Valitse pistesummat ja kokonaispistemäärät.
- Lisää Kaavio - XY (hajonta). Lisää akselleille otsikot.
- Lisää trendiviiva. Valitse "lineaarinen" ja "Näytä yhtälö".
- Valitse Data - Tilastotiedot - Korrelaatio

Pistesumma	Kokonaispistemäärä
1729	8641
1681	8606
1786	8595
1669	8567
1683	8425
1727	8385
1641	8351
1625	8277
1605	8206
1685	8178

Korrelaatiot	Sarake 1	Sarake 2
Sarake 1	1	
Sarake 2	0,619311861531821	1



Pyöristetään regressiosuoran kertoimet neljän merkitsevän numeron tarkkuuteen.

Regressiosuoran yhtälö on $y = 1,979x + 5091$. Korrelaatiokertoimen arvo on $r \approx 0,62$.

Hämäläisellä 100 m juoksun ja kuulantyyntö pistesumma on $x = 858 + 732 = 1590$.

Lasketaan regressiosuoran mukainen ennuste Hämäläisen loppupistemääräksi.

$$y = 1,979 \cdot 1590 + 5091 = 8237,61 \approx \mathbf{8238}$$

Tämä tulos on suurempi kuin 9. sijoittuneen tulos 8206 ja pienempi kuin 8. sijoittuneen tulos 8277 eli $8206 < \mathbf{8238} < 8277$.

Tällä tuloksella Hämäläinen olisi siis sijoittunut yhdeksänneksi.
(Todellisuudessa hänen sijoituksensa oli 23.)

Huomautus:

Mallin ennuste Hämäläisen pistemäärälle riippuu regressiosuorassa olevien kertoimien tarkkuudesta. Tässä ratkaisussa kertoimille käytettiin neljän merkitsevän numeron tarkkuutta.

Vastaus:

$x = 100$ m juoksun ja kuulantyyönnön pistesumma, $y =$ kokonaispistemäärä

$$y = 1,979x + 5091 \text{ ja } r \approx 0,62$$

Muuttujien välillä on huomattava positiivinen korrelaatio.

Hämäläisen odotettavissa oleva kokonaispistemäärä olisi mallin mukaan 8238.

7.16

a)

- 1) Lapsen kasvaessa sekä pituus että paino nousevat. Yksilöllisiä eroja voi esiintyä, mutta niistä huolimatta yleinen havainto on, että pidempi lapsi on myös painavampi. Muuttujien välillä on siis positiivinen korrelaatio. (Tämä ei kuitenkaan tarkoita *täydellistä* korrelaatiota, eli pidempi lapsi ei aina ole painavampi kuin lyhyempi lapsi.)
- 2) Ihmisen verenpaine riippuu perintötekijöistä ja elintavoista. Verenpaineella ei yleistiedon perusteella ole yhteyttä ihmisen kengännumeroon. Muuttujien välinen korrelaatio on siis lähes nolla.
- 3) Bruttokansantuote asukasta kohti eli BKT/asukas kuvaa kansantalouden yleistä elintaso. Mitä korkeampi BKT, sitä korkeampi elintaso. Korkean elintason maissa käytetään enemmän rahaa lasten terveydenhuoltoon. Tämän vuoksi korkean BKT:n maissa lapsikuolleisuus on tyypillisesti alhaisempi kuin matalan BKT:n maissa. Muuttujien välillä on siis negatiivinen korrelaatio.
- 4) Tikanheittojen pysty- ja vaakakoordinaattinen positiivinen korrelaatio tarkoittaisi, että heitetty tikat osuvat nousevan suoran muotoiseen kuvioon. Vastaavasti negatiivinen korrelaatio tarkoittaisi tikkojen osuvan laskevan suoran muotoiseen kuvioon. Ei ole mitään syytä odottaa, että tikat yleisesti osuisivat jommallakummalla tavalla. Näiden muuttujien välinen korrelaatio on siis lähes nolla.

b)

- 1) Kahden muuttujan välinen korrelaatio tarkoittaa tilastollista yhteyttä näiden muuttujien välillä. Positiivinen korrelaatio tarkoittaa, että toisen kasvaessa myös toinen keskimäärin kasvaa. Negatiivinen korrelaatio tarkoittaa, että toisen kasvaessa toinen keskimäärin pienenee.

Korrelaatio tarkoittaa lineaarista riippuvuutta. Korrelaatio on sitä voimakkaampi, mitä tiiviimmin havaintoarvopisteet asettuvat jonkin nousevan tai laskevan suoran suuntaisesti.

Korrelaatiokerroin mittaa lineaarisen riippuvuuden voimakkuutta. Korrelaatiokertoimen arvo on välillä $[-1, 1]$. Mitä lähempänä korrelaatiokertoimen arvo on lukua 1 tai lukua -1 , sitä voimakkaampi riippuvuus on. Korrelaatiokerroin on lähellä nollaa silloin, kun (lineaarista) riippuvuutta ei ole.

- 2) Esimerkiksi:

muuttuja 1 = päivässä syödyn jäätelön määrä

muuttuja 2 = päivässä myytyjen aurinkolasien lukumäärä

Molempien muuttujien arvo kasvaa aurinkoisina kesäpäivinä ja vähenee sateisina syyspäivinä, joten muuttujien välillä on positiivinen korrelaatio. Jäätelön syöminen ei kuitenkaan sinänsä häikäise ihmisiä ja saa heitä ostamaan aurinkolaseja eikä aurinkolasien ostaminen saa ihmisiä haluamaan jäätelöä. Toinen muuttuja ei siis aiheuta toista eli muuttujien välillä ei ole syy-seuraus-suhdetta.

Vastaus:

- a) 1) positiivinen 2) lähes nolla 3) negatiivinen 4) lähes nolla

7.17

a)

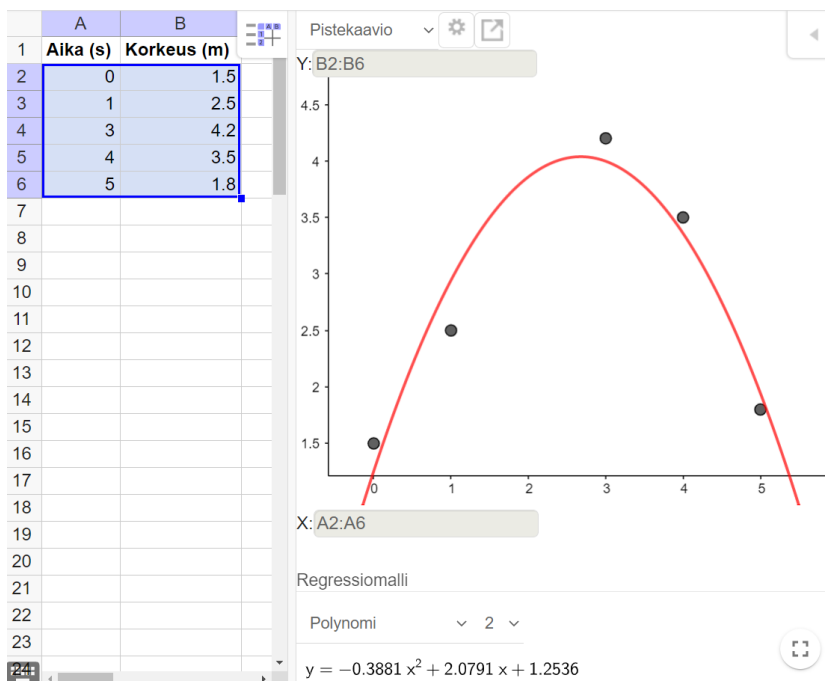
Ratkaistaan tehtävä GeoGebra-ohjelmalla. Pistejoukkoa voidaan tutkia joko piirto- tai taulukko-sovelluksella.

x = aika

y = korkeus

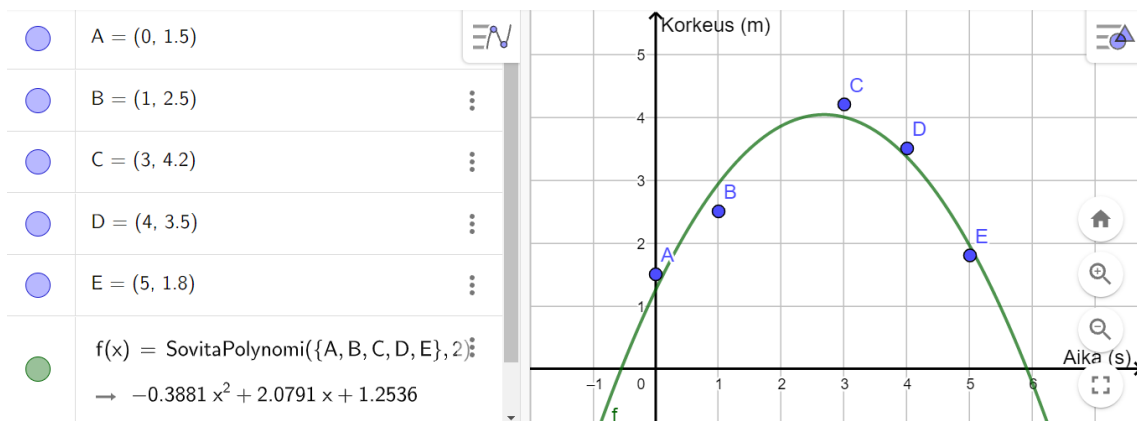
Tapa 1. Sovitetaan malli taulukko-sovelluksella.

Kirjoitetaan tai kopioidaan ajat ja korkeudet taulukkoon. Valitaan aineisto ja tehdään kahden muuttujan regressioanalyysi. Heitetyn pallon lentorata on paraabelin muotoinen, joten sopiva malli kuvaamaan lentorataa on 2. asteen polynomifunktio. Valitaan regressiomalliksi ”polynomi” ja asteluvuksi 2.



Tapa 2. Sovitetaan malli piirtoalueella.

Heitetyn pallon lentorata on paraabelin muotoinen, joten sopiva malli kuvaamaan lentorataa on 2. asteen polynomifunktio. Sovitetaan funktio ohjelmalla.



b)

Mallin eli regressiokäyrän yhtälö on $y = -0,39x^2 + 2,1x + 1,3$.

c)

Aikaa on kulunut 2,0 sekuntia, joten $x = 2,0$.

Lasketaan muuttujan $y = -0,39x^2 + 2,1x + 1,3$ arvo.

$$y = -0,39 \cdot 2,0^2 + 2,1 \cdot 2,0 + 1,3 = 3,94 \approx 3,9$$

Pallon korkeus on 3,9 metriä.

d)

Kun pallo putoaa maahan, sen korkeus on nolla eli $y = 0$.

Muodostetaan mallin $y = -0,39x^2 + 2,1x + 1,3$ avulla yhtälö ja ratkaistaan x .

$$-0,39x^2 + 2,1x + 1,3 = 0$$

$$x = -0,56 \dots \text{ tai } x = 5,94 \dots$$

Pallon lentomatkaan kulunut aika on positiivinen luku, joten $x \approx 5,9$.

Pallo putoaa maahan 5,9 sekunnin kuluttua.

Ratkaistaan yhtälö CAS-laskimessa:

$$\text{Ratkaise}(-0.39 x^2 + 2.1 x + 1.3 = 0)$$

$$\approx \{x = -0.5607, x = 5.9453\}$$

Vastaus:

a) Toisen asteen polynominen malli.

b) $y = -0,39x^2 + 2,1x + 1,3$

c) 3,9 metriä

d) 5,9 sekunnin kuluttua

7.18

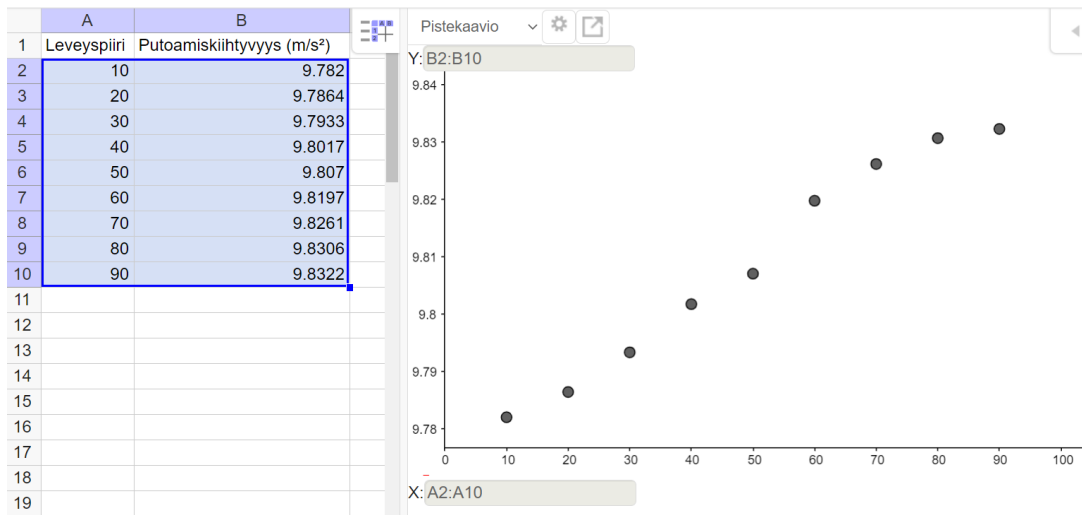
Ratkaistaan tehtävä GeoGebran taulukko-sovelluksella.

x = leveyspiiri

y = putoamiskiikkyvyys

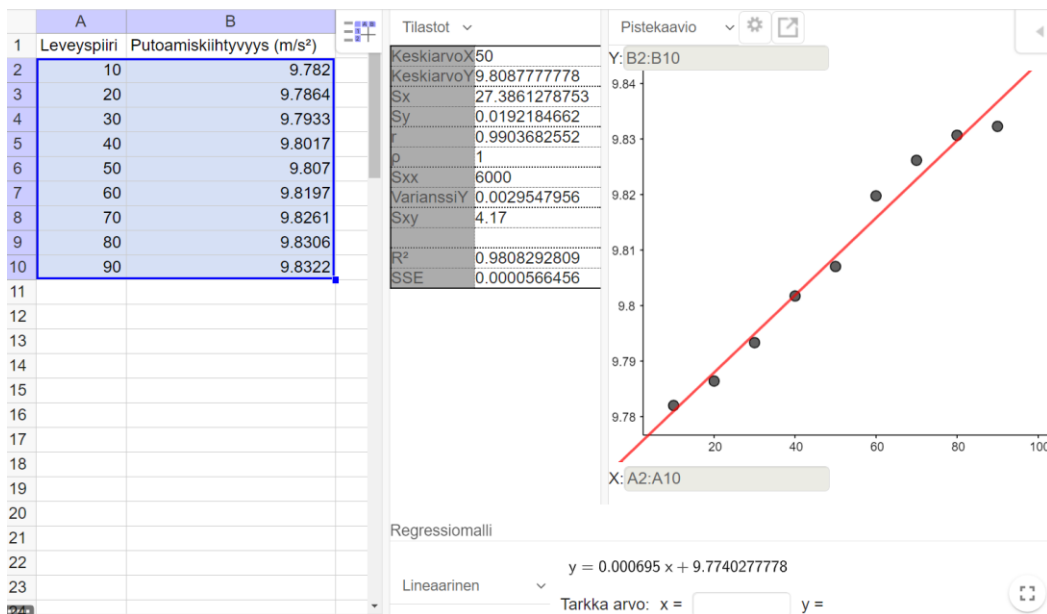
a)

Valitaan aineisto ja tehdään kahden muuttujan regressioanalyysi.



b)

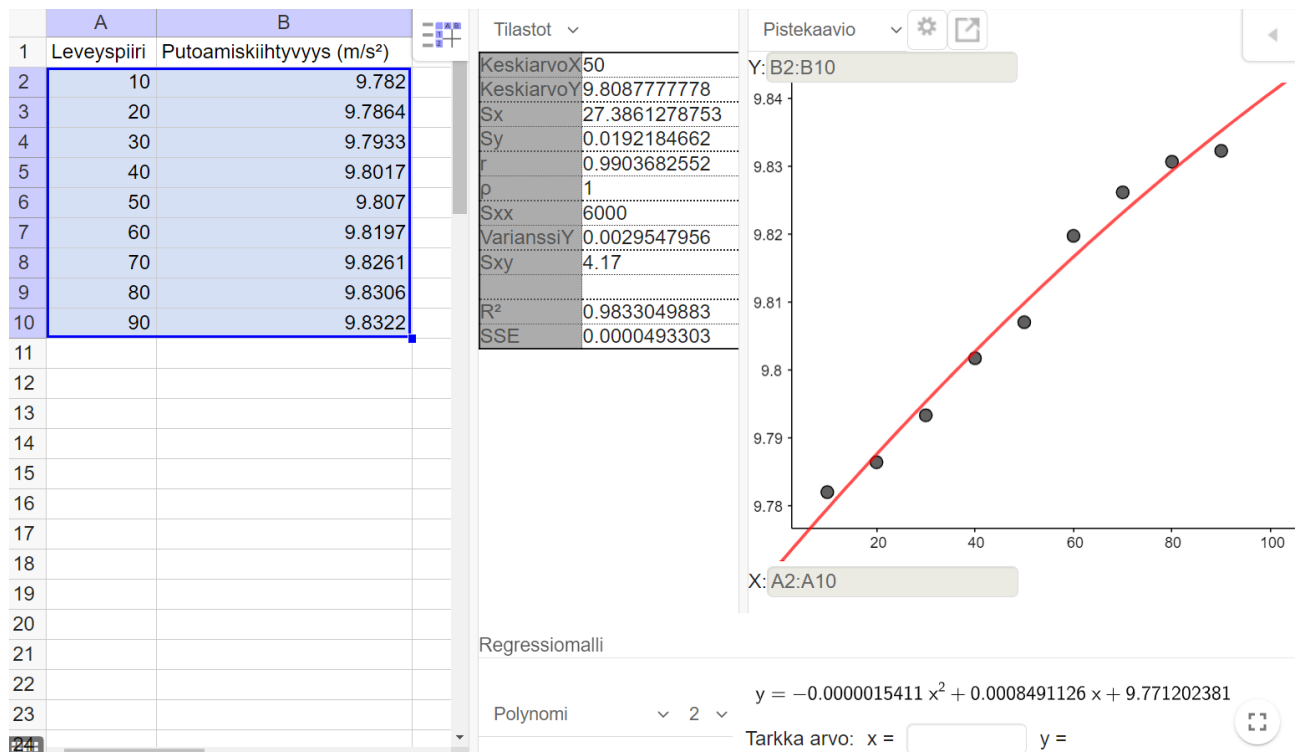
Valitaan regressiomalliksi "lineaarinen".



Lineaarinen malli on $y = 0,00070x + 9,77$.

Mallin selitysaste on $R^2 = 0,9808... \approx 0,981 = 98,1\%$.

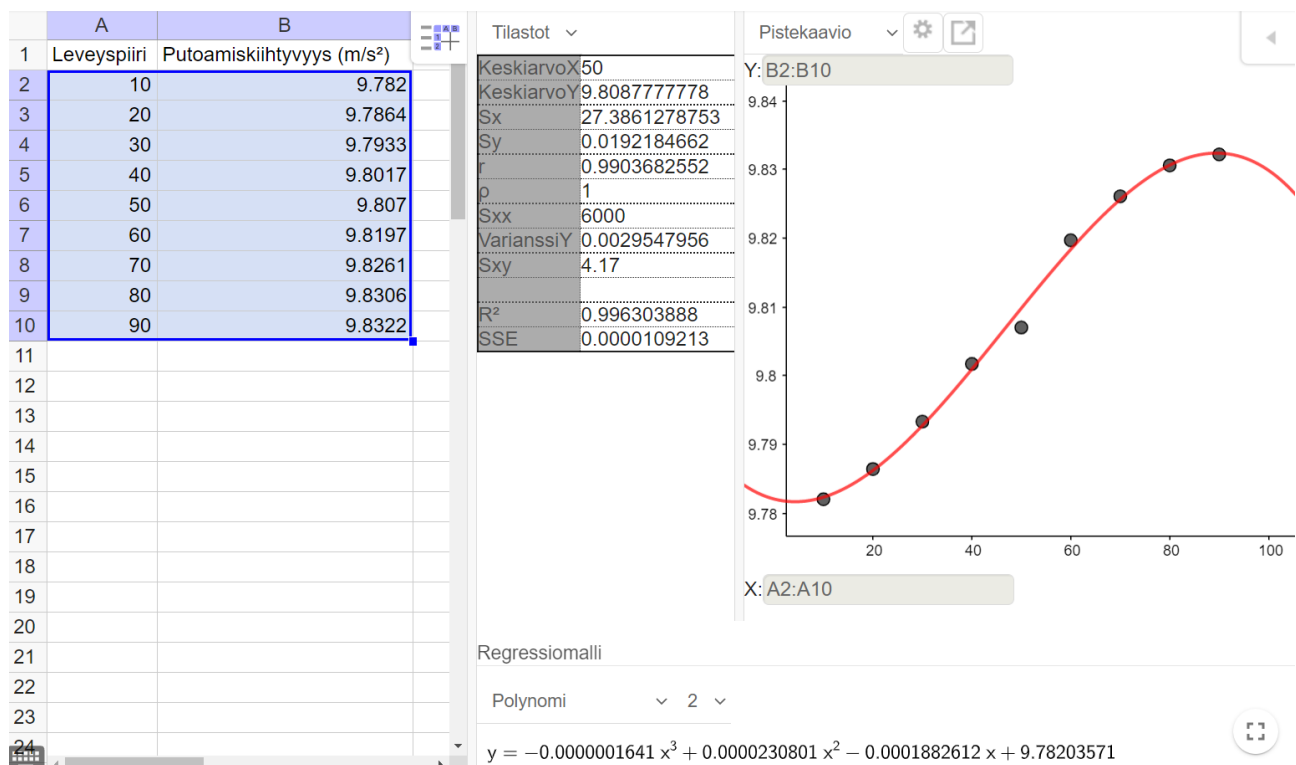
Valitaan regressiomalliksi "polynomi" ja asteluvuksi 2.



Toisen asteen polynominen malli on $y = -0,0000015x^2 + 0,00085x + 9,77$.

Mallin selitysaste on $R^2 = 0,9833... \approx 0,983 = 98,3\%$.

Valitaan regressiomalliksi "polynomi" ja asteluvuksi 3.



Kolmannen asteen polynominen malli on $y = -0,00000016x^3 + 0,000023x^2 - 0,00019x + 9,78$.

Mallin selitysaste on $R^2 = 0,9966... \approx 0,997 = 99,7\%$.

Korkein selitysaste (99,7 %) on 3. asteen polynomisella mallilla, joten tämä malli soveltuu parhaiten kuvaamaan havaintoaineistoa.

c)

Päiväntasaajan leveyspiiri on 0., joten $x = 0$.

Käytetään 3. asteen polynomimallia ja lasketaan putoamiskiihtyvyys.

$$y = -0,00000016 \cdot 2,0^3 + 0,000023 \cdot 0^2 - 0,00019 \cdot 0 + 9,78 = 9,78$$

Päiväntasaajalla putoamiskiihtyvyys on mallin mukaan 9,78 m/s².

Huomautus:

Lineaarisen ja 2. asteen polynomisen mallin mukainen tulos c-kohtaan on 9,77 m/s².

Taulukkokirjan mukainen arvo on 9,7804 m/s², joten 3. asteen polynomisen malli antaa parhaimman arvion.

Vastaus:

b)

Lineaarinen malli: $y = 0,00070x + 9,77$ (selitysaste 98,1 %)

2. asteen malli: $y = -0,0000015x^2 + 0,00085x + 9,77$ (selitysaste 98,3 %)

3. asteen malli: $y = -0,00000016x^3 + 0,000023x^2 - 0,00019x + 9,78$ (selitysaste 99,7 %)

c)

9,78 m/s²

7.19

Ratkaistaan tehtävä GeoGebran taulukko-sovelluksella.

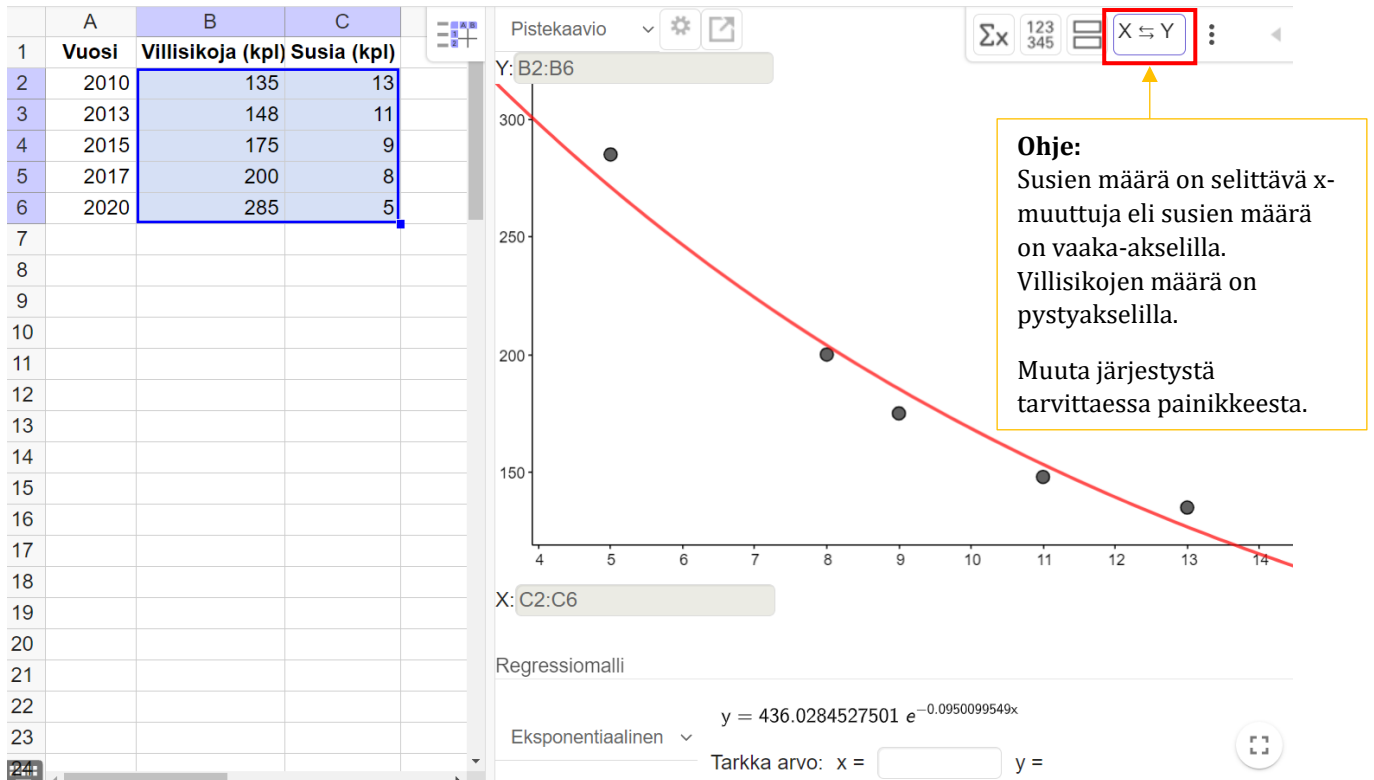
a)

Villisikojen määrä riippuu susien määrästä.

x = susien määrä (vaaka-akselilla)

y = villisikojen määrä (pystyakselilla)

Valitaan villisikojen ja susien määrä ja tehdään kahden muuttujan regressioanalyysi.



b)

Eksponentiaalinen malli on $y = 436 \cdot e^{-0,095x}$.

c)

Kun susia ei enää ole, on $x = 0$.

Lasketaan villisikojen määrä.

$$y = 436 \cdot e^{-0,0950 \cdot 0} = 436 \cdot e^0 = 436 \cdot 1 = 436.$$

Villisikapopulaation koko on 436 yksilöä.

GeoGebra laskurilla:

$$y = 436.0284527501 e^{-0.0950099549x}$$

Tarkka arvo: x = y =

Vastaus:

b) $y = 436 \cdot e^{-0,095x}$

c) 436 villisikaa

7.20

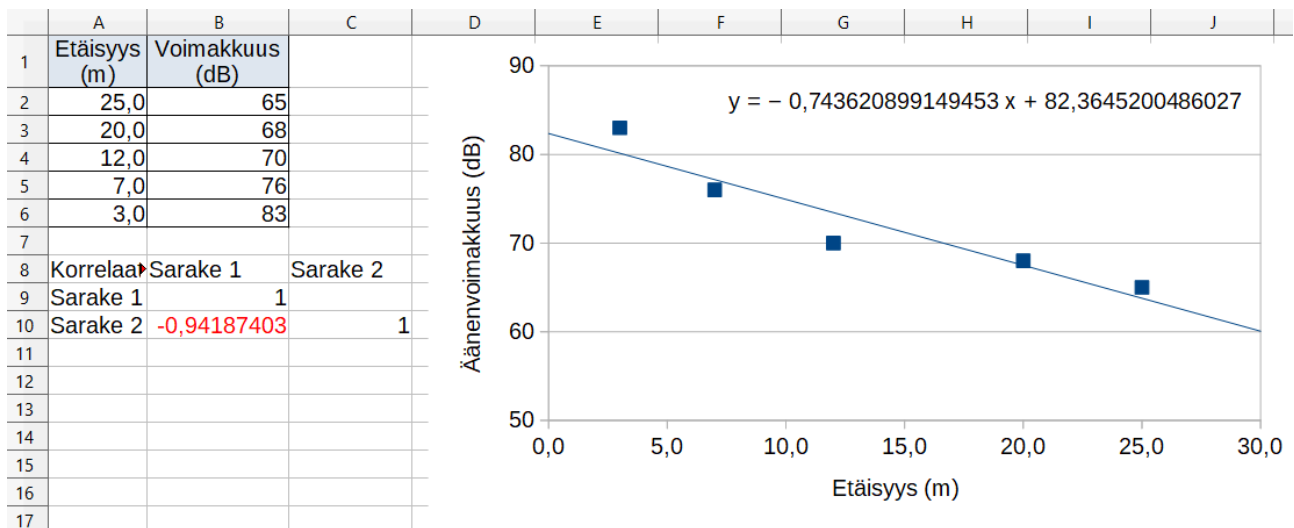
Ratkaistaan tehtävä taulukkolaskentaohjelmalla.

a)

Piirretään hajontakuviota. Määritetään regressiosuoran yhtälö ja korrelaatiokerroin tilastotoiminnoilla.

Ohje:

- Valitse etäisyydet ja voimakkuudet.
- Lisää Kaavio - XY (hajonta). Lisää akselleille otsikot.
- Lisää trendiviiva. Valitse "lineaarinen" ja "Näytä yhtälö".
- Muuta y-akselin asteikkoa asetuksista.
- Valitse Data - Tilastotiedot - Korrelaatio



$x = \text{etäisyys}$

$y = \text{voimakkuus}$

Regressiosuora on $y = -0,74x + 82$.

Korrelaatiokerroin on $r = -0,941\dots \approx -0,94$.

b)

Lasketaan äänivoimakkuuksien logaritmit sarakkeeseen D.

	A	B	C	D
1	Etäisyys (m)	Voimakkuus (dB)		lg(y)
2	25,0	65		=log10(B2)
3	20,0	68		
4	12,0	70		
5	7,0	76		
6	3,0	83		

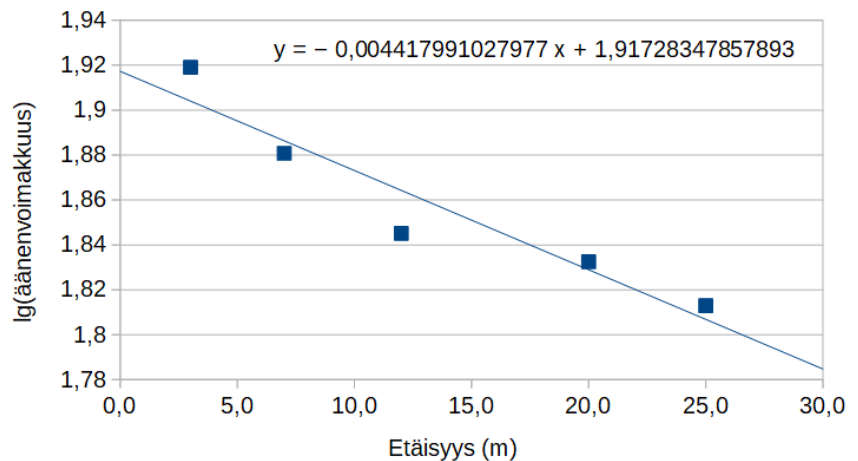
	A	B	C	D
1	Etäisyys (m)	Voimakkuus (dB)		lg(y)
2	25,0	65		1,8129134
3	20,0	68		1,8325089
4	12,0	70		1,845098
5	7,0	76		1,8808136
6	3,0	83		1,9190781

Ohje:

Laske logaritmien arvot esimerkiksi taulukkolaskentaohjelman log10-komennolla.

Piirretään arvoista (x, lg(y)) hajontakuvio. Määritetään regressiosuoran yhtälö ja korrelaatiokerroin tilastotoiminnoilla.

C	D	E	F	G	H	I	J	K	L
x	lg(y)								
25,0	1,8129134								
20,0	1,8325089								
12,0	1,845098								
7,0	1,8808136								
3,0	1,9190781								
Korrelaatiot	Sarake 1	Sarake 2							
Sarake 1		1							
Sarake 2	-0,9513163		1						



x = etäisyys

y = lg(voimakkuus)

Regressiosuora on $y = -0,0044x + 1,9$.

Korrelaatiokerroin on $r = -0,951... \approx -0,95$.

b-kohdan mallin korrelaatiokertoimen itseisarvo on hieman suurempi ($0,95 > 0,94$), joten b-kohdassa muuttujien välinen korrelaatio on hieman voimakkaampi.

b-kohdan malli kuvaa siis tilannetta hieman paremmin.