

7.1

Suurin osa suorituksista on välillä 6–10 pistettä. Kaksi 1 pisteen suoritusta poikkeavat näistä selvästi.

Ratkaistaan tehtävä GeoGebran taulukkolaskennalla.

1) Koko jakauma

Syötetään pisteet ja frekvenssit taulukkolaskentaan ja määritetään tilastolliset tunnusluvut.

| | A | B |
|---|----------------|----------|
| 1 | Pisteet | f |
| 2 | 1 | 2 |
| 3 | 6 | 1 |
| 4 | 7 | 2 |
| 5 | 8 | 4 |
| 6 | 9 | 5 |
| 7 | 10 | 3 |

| | |
|--------------|--------|
| n | 17 |
| Keskiarvo | 7.5882 |
| σ | 2.6359 |
| s | 2.717 |
| Σx | 129 |
| Σx^2 | 1097 |
| Min | 1 |
| Q1 | 7 |
| Mediaani | 8 |
| Q3 | 9 |
| Max | 10 |

Moodi on 9 pistettä.

Mediaani on 8 pistettä.

Keskiarvo on 7,6 pistettä.

Vaihteluvälin pituus on $10 - 1 = 9$ pistettä.

Keskihajonta on 2,7 pistettä.

2) Jakauma ilman 1 pisteen suorituksia

Poistetaan 1 pisteen suoritukset ja määritetään tilastolliset tunnusluvut.

| | A | B |
|---|---------|----------|
| 1 | Pisteet | <i>f</i> |
| 2 | 6 | 1 |
| 3 | 7 | 2 |
| 4 | 8 | 4 |
| 5 | 9 | 5 |
| 6 | 10 | 3 |

| | |
|--------------|--------|
| n | 15 |
| Keskiarvo | 8.4667 |
| σ | 1.147 |
| s | 1.1872 |
| Σx | 127 |
| Σx^2 | 1095 |
| Min | 6 |
| Q1 | 8 |
| Mediaani | 9 |
| Q3 | 9 |
| Max | 10 |

Moodi on 9 pistettä.

Mediaani on 9 pistettä.

Keskiarvo on 8,5 pistettä.

Vaihteluvälin pituus on $10 - 6 = 4$ pistettä.

Keskihajonta on 1,2 pistettä.

Vastaus

Koko jakauma: moodi 9 p, mediaani 8 p, keskiarvo 7,6 p, vaihteluvälin pituus 9 p ja keskihajonta 2,7 p.

Ilman yhden pisteen suorituksia: moodi 9 p, mediaani 9 p, keskiarvo 8,5 p, vaihteluvälin pituus 4 p ja keskihajonta 1,2 p.

Ilman poikkeavia havaintoja keskiarvo on selkeästi korkeampi ja vaihteluvälin pituus ja keskihajonta pienempiä.

7.2

- a) Kaikki havaintoarvot ovat välillä 7–10. Jakaumaan ei sisälly poikkeavia arvoja.
- b) Suurin osa havaintoarvoista on välillä 1–3. Kaksi havaintoarvoa 10 poikkeavat näistä selvästi.

Tutkitaan tunnuslukuja Geogebbran taulukkolaskennalla.

1) Koko jakauma

Syötetään havaintoarvot ja frekvenssit taulukkolaskentaan ja määritetään tilastolliset tunnusluvut.

| | A | B |
|---|---------------------|----------|
| 1 | Havaintoarvo | f |
| 2 | 1 | 3 |
| 3 | 2 | 2 |
| 4 | 3 | 1 |
| 5 | 10 | 2 |

| | |
|--------------|--------|
| n | 8 |
| Keskiarvo | 3.75 |
| σ | 3.6657 |
| s | 3.9188 |
| Σx | 30 |
| Σx^2 | 220 |
| Min | 1 |
| Q1 | 1 |
| Mediaani | 2 |
| Q3 | 6.5 |
| Max | 10 |

Moodi on 1, mediaani 2 ja keskiarvo 3,75.

2) Jakauma ilman 1 pisteen suorituksia

Poistetaan havaintoarvot 10 ja määritetään tilastolliset tunnusluvut.

| | A | B |
|---|--------------|----------|
| 1 | Havaintoarvo | <i>f</i> |
| 2 | 1 | 3 |
| 3 | 2 | 2 |
| 4 | 3 | 1 |

| | |
|--------------|--------|
| n | 6 |
| Keskiarvo | 1.6667 |
| σ | 0.7454 |
| s | 0.8165 |
| Σx | 10 |
| Σx^2 | 20 |
| Min | 1 |
| Q1 | 1 |
| Mediaani | 1.5 |
| Q3 | 2 |
| Max | 3 |

Moodi on 1, mediaani 1,5 ja keskiarvo 1,67.

Poikkeavat havaintoarvot eivät vaikuta moodiin, mutta koska poikkeavat havaintoarvot ovat selkeästi suurempia kuin muut havaintoarvot, ne kasvattavat keskiarvoa ja mediaanin arvoa. Ilman poikkeavia arvoja mediaani ja keskiarvo ovat pienempiä.

Vastaus

- a) Jakaumaan ei sisälly poikkeavia arvoja.
- b) Havaintoarvo 10 poikkeaa selvästi muista. Ilman poikkeavia arvoja mediaani ja keskiarvo ovat pienempiä.
(Mediaani 2 ja keskiarvo 3,75. Ilman havaintoarvoa 10 mediaani 1,5 ja keskiarvo 1,67.)

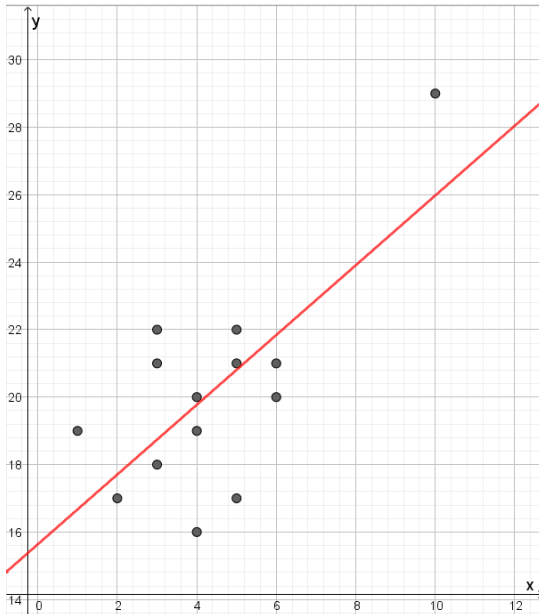
7.3

Ratkaistaan tehtävä GeoGebran taulukkolaskennalla.

Valitaan selittäväksi muuttujaksi x tuloluokka ja selitettäväksi muuttujaksi y onnellisuus. Taulukoidaan havaintoarvot.

| | A | B |
|----|----------------|-----------------|
| 1 | Tuloluokka x | Onnellisuus y |
| 2 | 1 | 19 |
| 3 | 3 | 22 |
| 4 | 3 | 21 |
| 5 | 4 | 20 |
| 6 | 5 | 21 |
| 7 | 5 | 17 |
| 8 | 6 | 21 |
| 9 | 2 | 17 |
| 10 | 3 | 18 |
| 11 | 4 | 19 |
| 12 | 4 | 16 |
| 13 | 5 | 22 |
| 14 | 6 | 20 |
| 15 | 10 | 29 |

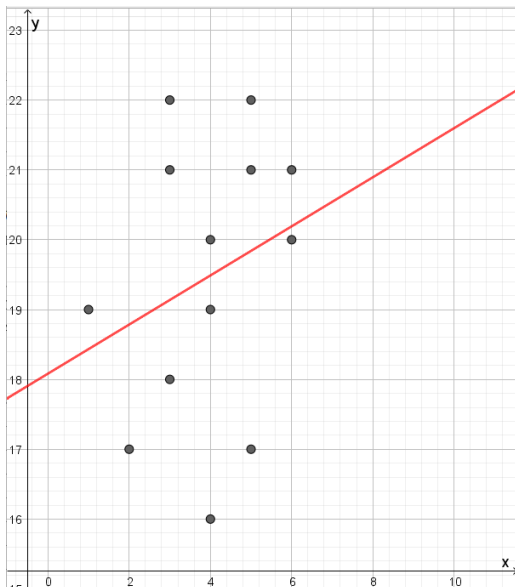
Piirretään hajontakuvio ja määritetään korrelaatiokerroin.



| | |
|----------------|----------|
| KeskiarvoX | 4.3571 |
| KeskiarvoY | 20.1429 |
| Sx | 2.17 |
| Sy | 3.1831 |
| r | 0.7048 |
| p | 0.4128 |
| Sxx | 61.2143 |
| VarianssiY | 131.7143 |
| Sxy | 63.2857 |
| | |
| R ² | 0.4967 |
| SSE | 66.287 |

Korrelaatiokertoimen $r = 0,70$ perusteella tuloluokan ja onnellisuuden välillä näyttää olevan huomattava positiivinen lineaarinen riippuvuus.

Hajontakuvioista nähdään, että aineistoon sisältyy yksi muista selvästi poikkeava havainto. Poistetaan taulukosta rivillä 15 oleva havainto, piirretään uusi hajontakuvio ja määritetään korrelaatiokerroin.



| | |
|------------|---------|
| KeskiarvoX | 3.9231 |
| KeskiarvoY | 19.4615 |
| Sx | 1.4979 |
| Sy | 1.9839 |
| r | 0.2653 |
| ρ | 0.2617 |
| Sxx | 26.9231 |
| VarianssiY | 47.2308 |
| Sxy | 9.4615 |
| | |
| R^2 | 0.0704 |
| SSE | 43.9057 |

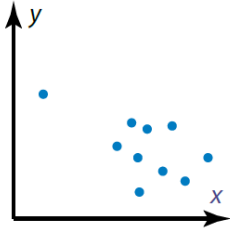
Nyt korrelaatiokertoimen $r = 0,27$ perusteella tuloluokan ja onnellisuuden välinen lineaarinen riippuvuus onkin voimakkuudeltaan merkityksetön.

Vastaus

Tuloksiin sisältyy yksi poikkeava havainto. Kun se poistetaan, lineaarisen riippuvuuden voimakkuus muuttuu huomattavasta merkityksettömäksi.

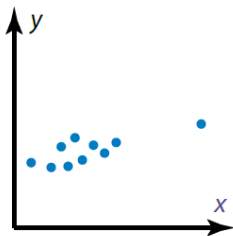
7.4

- a) Tuloksiin sisältyy yksi poikkeava havaintoarvo.



Poikkeava havaintoarvo on koordinaatistossa ylävasemmalla. Kun se poistetaan, lineaarinen riippuvuus heikkenee, sillä havaintoarvot sopivat huonommin suoralle.

- b) Tuloksiin sisältyy yksi poikkeava havaintoarvo.



Poikkeava havaintoarvo on koordinaatistossa yläoikealla. Kun se poistetaan, lineaarinen riippuvuus säilyy edelleen voimakkaana, sillä havaintoarvot sopivat edelleen hyvin suoralle.

Vastaus

- a) Tuloksiin sisältyy yksi poikkeava havainto. Kun se poistetaan, lineaarinen riippuvuus heikkenee.
- b) Tuloksiin sisältyy yksi poikkeava havainto. Kun se poistetaan, lineaarinen riippuvuus säilyy voimakkaana.

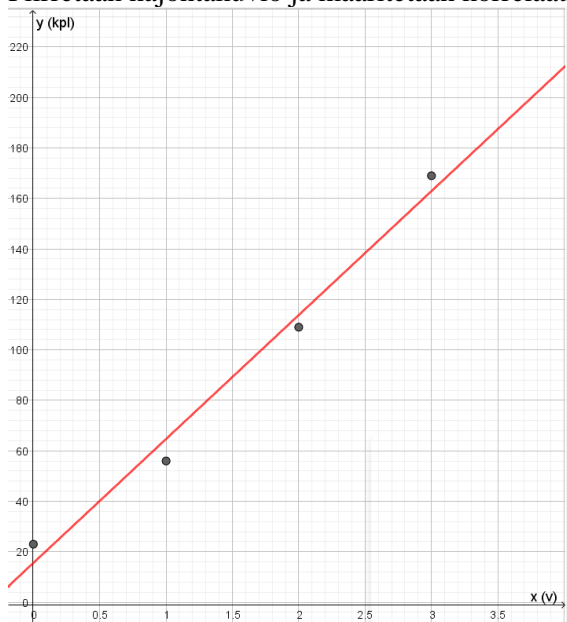
7.5

- a) Ratkaistaan tehtävä GeoGebran taulukkolaskennalla.

Muuttuja x on vuodesta 2010 kulunut aika vuosina ja muuttuja y täyssähköautojen määrä. Taulukoidaan havaintoarvot.

| | A | B |
|---|--------------|----------------------------------|
| 1 | Aika x (v) | Täyssähköautojen määrä y (kpl) |
| 2 | 0 | 23 |
| 3 | 1 | 56 |
| 4 | 2 | 109 |
| 5 | 3 | 169 |

Piirretään hajontakuviot ja määritetään korrelaatiokerroin.

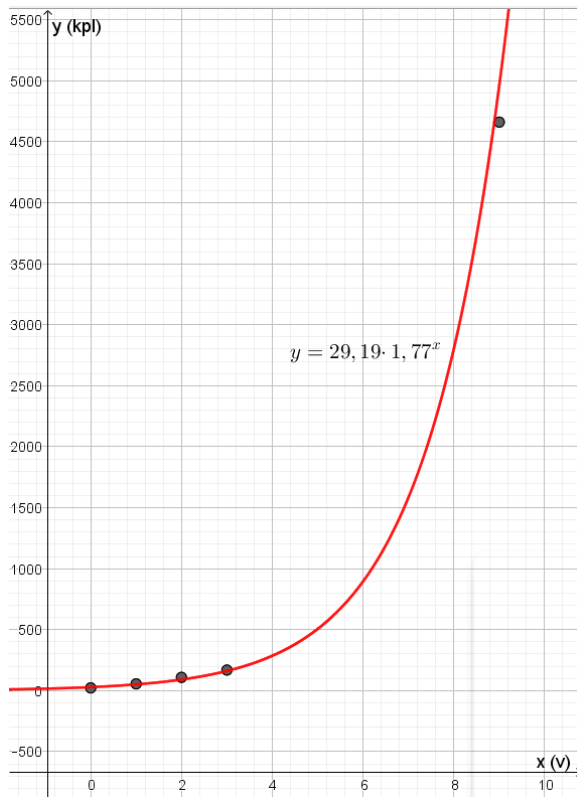


| | |
|----------------|----------|
| KeskiarvoX | 1.5 |
| KeskiarvoY | 89.25 |
| Sx | 1.291 |
| Sy | 63.8873 |
| r | 0.9922 |
| ρ | 1 |
| Sxx | 5 |
| VarianssiY | 12244.75 |
| Sxy | 245.5 |
| | |
| R ² | 0.9844 |
| SSE | 190.7 |

Hajontakuvion ja korrelaatiokertoimen $r \approx 0,99$ perusteella muuttujien välillä näyttää olevan voimakas positiivinen lineaarinen riippuvuus.

Regressiosuoran yhtälö $y = 49,1x + 15,6$ antaa vuoden 2025 ennusteeksi $y = 49,1 \cdot 15 + 15,6 \approx 752$ täyssähköautoa.

- b)** Lisätään taulukkoon havaintoarvo $x = 9$ ja $y = 4661$, piirretään uusi hajontakuviokuva ja kokeillaan, mikä regressiomalli näyttää sopivan parhaiten havaintoarvoihin.



| | |
|----------------|------------|
| KeskiarvoX | 3 |
| KeskiarvoY | 1003.6 |
| Sx | 3.5355 |
| Sy | 2045.2972 |
| r | 0.9568 |
| p | 1 |
| Sxx | 50 |
| VarianssiY | 16732963.2 |
| Sxy | 27676 |
| | |
| R ² | 0.9945 |
| SSE | 91790.0655 |

Regressiomalliksi näyttää sopivan eksponentiaalinen malli (Kasvu).

Regressiokäyrän yhtälö $y = 29,19 \cdot 1,77^x$ antaa vuoden 2025 ennusteeksi $y = 29,19 \cdot 1,77^{15} \approx 153000$ täyssähköautoa.

- c) Molemmat ennusteet sisältävät epävarmuutta. Ne myös poikkeavat toisistaan merkittävästi. Korrelaatiokertoimen perusteella a-kohdassa lineaarinen riippuvuus on voimakas, mutta harha johtuu havaintoarvojen vähyydestä. Varmemman ennusteen tekemiseen tarvitaan enemmän havaintoarvoja.

Vastaus

- a) lineaarisen mallin mukaan 752
- b) eksponentiaalisen mallin mukaan 153 000
- c) Molemmat ennusteet sisältävät epävarmuutta. Varmemman ennusteen tekemiseen tarvitaan enemmän havaintoarvoja.

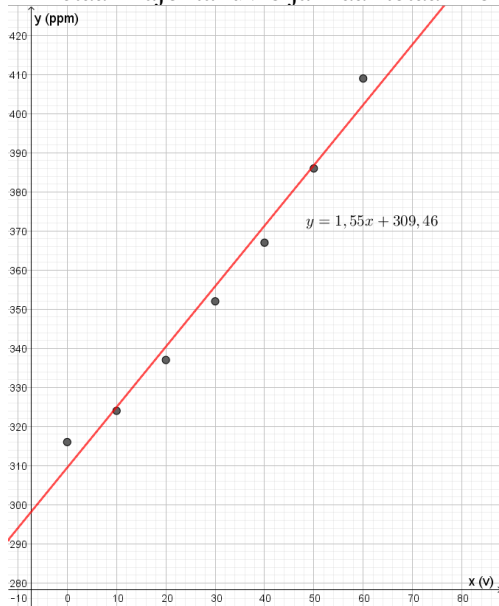
7.6

a) Ratkaistaan tehtävä GeoGebran taulukkolaskennalla.

Muuttuja x on vuodesta 2058 kulunut aika vuosina ja muuttuja y hiilidioksidipitoisuuden keskiarvo maaliskuussa. Taulukoidaan havaintoarvot.

| | A | B |
|---|--------------|------------------------------|
| 1 | Aika x (v) | Hiilidioksidipitoisuus (ppm) |
| 2 | 0 | 316 |
| 3 | 10 | 324 |
| 4 | 20 | 337 |
| 5 | 30 | 352 |
| 6 | 40 | 367 |
| 7 | 50 | 386 |
| 8 | 60 | 409 |

Piirretään hajontakuviio ja määritetään korrelaatiokerroin.



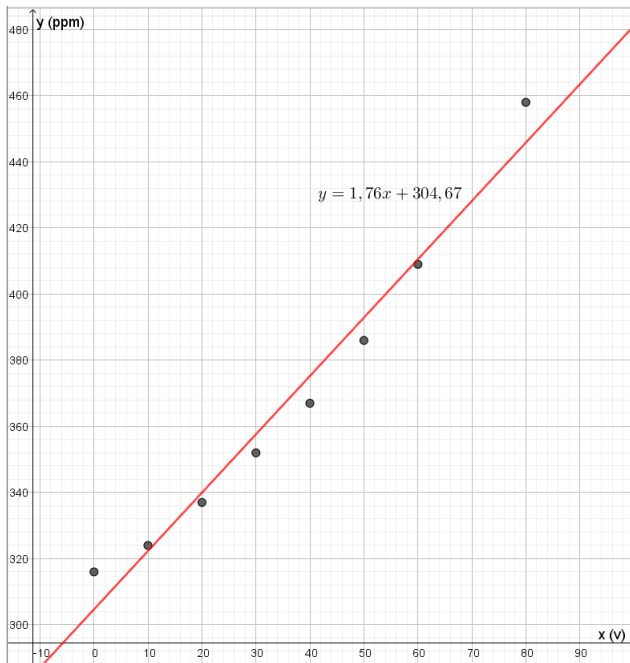
| | |
|------------|-----------|
| KeskiarvoX | 30 |
| KeskiarvoY | 355.8571 |
| Sx | 21.6025 |
| Sy | 33.7413 |
| r | 0.9901 |
| ρ | 1 |
| Sxx | 2800 |
| VarianssiY | 6830.8571 |
| Sxy | 4330 |
| | |
| R^2 | 0.9803 |
| SSE | 134.8214 |

Hajontakuvion ja korrelaatiokertoimen $r \approx 0,99$ perusteella muuttujien välillä näyttää olevan voimakas positiivinen lineaarinen riippuvuus.

Regressiosuoran yhtälö $y = 1,55x + 309,46$ antaa vuoden 2058 ennusteeksi $y = 1,55 \cdot 100 + 309,46 \approx 464$ ppm.

Havaintoarvoja, joiden perusteella ennuste on tehty, on kohtuullinen määrä ja niiden perusteella tehdyn hajontakuvion ja korrelaatiokertoimen perusteella muuttujien välillä on voimakas positiivinen lineaarinen riippuvuus. Ennusteen tekemiseen käytetty aineisto on kerätty 60 vuoden aikana, mutta ennuste tehdään 40 vuoden päähän viimeisimmästä havaintoarvosta, joten tämä tuo ennusteeseen lisää epävarmuutta.

- b)** Lisätään taulukkoon havaintoarvo $x = 80$ ja $y = 458$ piirretään uusi hajontakuvi ja määritetään korrelaatiokerroin



| | |
|------------|-----------|
| KeskiarvoX | 36.25 |
| KeskiarvoY | 368.625 |
| Sx | 26.6927 |
| Sy | 47.7492 |
| r | 0.9862 |
| ρ | 1 |
| Sxx | 4987.5 |
| VarianssiY | 15959.875 |
| Sxy | 8798.75 |
| | |
| R^2 | 0.9726 |
| SSE | 437.4687 |

Hajontakuvion ja korrelaatiokertoimen $r \approx 0,99$ perusteella muuttujien välillä näyttää olevan edelleen voimakas positiivinen lineaarinen riippuvuus.

Regressiosuoran yhtälö $y = 1,76x + 304,67$ antaa vuoden 2058 ennusteeksi $y = 1,76 \cdot 100 + 304,67 \approx 481$ ppm.

Vastaus

a) 464 ppm

b) 481 ppm

7.7

- a) Asetetaan appletilla havaintoarvojen lukumääräksi $n = 3$. Appletista saadaan regressiosuoran yhtälöksi $y = 73x + 233,33$. Lasketaan y:n arvo, kun $x = 15$.
$$y = 73 \cdot 15 + 233,33 \approx 1328$$
- b) Asetetaan appletilla havaintoarvojen lukumääräksi $n = 5$. Appletista saadaan regressiosuoran yhtälöksi $y = 125,8x + 138,2$. Lasketaan y:n arvo, kun $x = 15$.
$$y = 125,8 \cdot 15 + 138,2 \approx 2025$$
- c) Asetetaan appletilla havaintoarvojen lukumääräksi $n = 7$. Huomataan, että sovitus muuttuu eksponentiaaliseksi. Appletista saadaan regressiokäyrän yhtälöksi $y = 231,32 \cdot 1,28^x$. Lasketaan y:n arvo, kun $x = 15$.
$$y = 231,32 \cdot 1,28^{15} \approx 9383$$
- d) Asetetaan appletilla havaintoarvojen lukumääräksi $n = 9$. Appletista saadaan regressiokäyrän yhtälöksi $y = 229,99 \cdot 1,28^x$. Lasketaan y:n arvo, kun $x = 15$.
$$y = 229,99 \cdot 1,28^{15} \approx 9330$$
- e) Asetetaan appletilla havaintoarvojen lukumääräksi $n = 11$. Appletista saadaan regressiokäyrän yhtälöksi $y = 229,85 \cdot 1,28^x$. Lasketaan y:n arvo, kun $x = 15$.
$$y = 229,85 \cdot 1,28^{15} \approx 9324$$

Vastaus

- a) 1328
- b) 2025
- c) 9383
- d) 9330
- e) 9324

7.8

- a) Luetaan tilastokuvioista työllisyysasteen trendiluvut (sininen käyrä) toukokuulta 2012 ja toukokuulta 2020. Toukokuussa 2012 työllisyysasteen trendiluku on ollut 68 % ja toukokuussa 2020 trendiluku on ollut 70 %.

Verrataan trendilukujen erotusta toukokuun 2012 trendilukuun.

$$\frac{70\% - 68\%}{68\%} \approx 0,029 = 2,9\%$$

Trendiluku nousi 2,9 %.

- b) Luetaan tilastokuvioista työllisyysasteen suurimmat arvot (vihreä käyrä) vuosilta 2012 ja 2020. Toukokuussa 2012 työllisyysasteen suurin arvo on ollut 72 % ja toukokuussa 2020 suurin arvo on ollut 73 %.

Verrataan suurimpien arvojen erotusta vuoden 2012 suurimpaan arvoon.

$$\frac{73\% - 72\%}{72\%} \approx 0,014 = 1,4\%$$

Suurin arvo nousi 1,4 %.

- c) Esimerkiksi kesätyöt ja kausiluontoiset työt esimerkiksi maatalous-, matkailu- ja rakennusalalla aiheuttavat kausivaihteluita työllisyysasteeseen.

Osa töistä on sellaisia, että niitä voi tehdä ainoastaan tiettyyn vuodenaikaan (esim. kesällä mansikan poiminta ja teiden päällystäminen, talvella laskettelukeskusten lipunmyynti).

Vastaus

- a) 68 % ja 70 %, nousi 2,9 %
b) 72 % ja 73 %, nousi 1,4 %
c) kesätyöt, kausiluontoiset työt esimerkiksi maatalous-, matkailu- ja rakennusalalla

7.10

Suurin osa suorituksista on välillä 7–12 pistettä. Kaksi 22 pisteen suoritusta ja yksi 21 pisteen suoritus poikkeavat näistä selvästi.

Ratkaistaan tehtävä GeoGebran taulukkolaskennalla.

1) Koko jakauma

Syötetään pisteet ja frekvenssit taulukkolaskentaan ja määritetään tilastolliset tunnusluvut.

| | A | B |
|---|----------------|----------|
| 1 | Pisteet | f |
| 2 | 7 | 8 |
| 3 | 9 | 9 |
| 4 | 10 | 6 |
| 5 | 11 | 5 |
| 6 | 12 | 3 |
| 7 | 22 | 2 |
| 8 | 24 | 1 |

| | |
|--------------|---------|
| n | 34 |
| Keskiarvo | 10.4706 |
| σ | 4.1105 |
| s | 4.1723 |
| Σx | 356 |
| Σx^2 | 4302 |
| Min | 7 |
| Q1 | 9 |
| Mediaani | 9.5 |
| Q3 | 11 |
| Max | 24 |

Moodi on 9 pistettä.

Mediaani on 9,5 pistettä.

Keskiarvo on 10,5 pistettä.

Vaihteluvälin pituus on $24 - 7 = 17$ pistettä.

Keskihajonta on 4,2 pistettä.

2) Jakauma ilman 1 pisteen suorituksia

Poistetaan 22 ja 24 pisteen suoritukset ja määritetään tilastolliset tunnusluvut.

| | A | B |
|---|----------------|----------|
| 1 | Pisteet | f |
| 2 | 7 | 8 |
| 3 | 9 | 9 |
| 4 | 10 | 6 |
| 5 | 11 | 5 |
| 6 | 12 | 3 |

| | |
|--------------|--------|
| n | 31 |
| Keskiarvo | 9.2903 |
| σ | 1.6302 |
| s | 1.6572 |
| Σx | 288 |
| Σx^2 | 2758 |
| Min | 7 |
| Q1 | 7 |
| Mediaani | 9 |
| Q3 | 11 |
| Max | 12 |

Moodi on 9 pistettä.

Mediaani on 9 pistettä.

Keskiarvo on 9,3 pistettä.

Vaihteluvälin pituus on $12 - 7 = 5$ pistettä.

Keskihajonta on 1,7 pistettä.

Vastaus

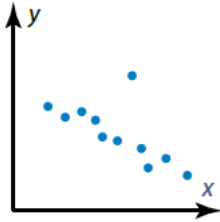
Koko jakauma: moodi 9 p, mediaani 9,5 p, keskiarvo 10,5 p, vaihteluvälin pituus 17 p ja keskihajonta 4,2 p.

Ilman 22 ja 24 pisteen suorituksia: moodi 9 p, mediaani 9 p, keskiarvo 9,3 p, vaihteluvälin pituus 5 p ja keskihajonta 1,7 p.

Ilman poikkeavia havaintoja keskiarvo on alempi ja vaihteluvälin pituus ja keskihajonta pienempiä.

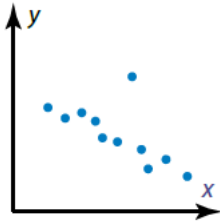
7.11

- a) Poikkeavan havaintoarvon x -koordinaatti ei poikkea merkittävästi muiden havaintoarvojen x -koordinaatista.



Havaintoarvo ei vaikuta merkittävästi muuttujan x keskiarvoon eikä keskihajontaa.

- b) Poikkeavan havaintoarvon y -koordinaatti on selkeästi suurempi kuin muiden havaintoarvojen y -koordinaatit.



Havaintoarvo siis suurentaa muuttujan y keskiarvoa. Tällöin muut arvot poikkeavat enemmän keskiarvosta, jolloin myös keskihajonta suurenee.

Vastaus

- a) ei vaikuta merkittävästi
b) suurentaa keskiarvoa ja keskihajontaa

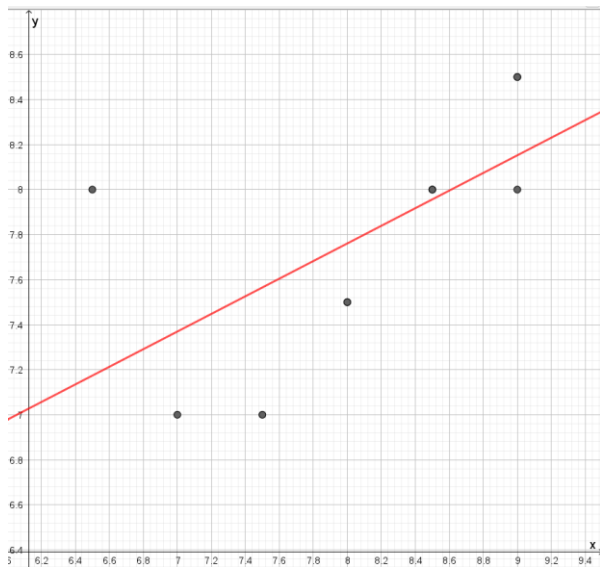
7.12

Ratkaistaan tehtävä GeoGebran taulukkolaskennalla.

Valitaan selittäväksi muuttujaksi x Tuomarin X antamat pisteet ja selitettäväksi muuttujaksi y Tuomarin Y antamat pisteet. Taulukoidaan havaintoarvot.

| | A | B |
|----|-----------|-----------|
| 1 | Tuomari X | Tuomari Y |
| 2 | 8 | 7.5 |
| 3 | 6.5 | 8 |
| 4 | 8.5 | 8 |
| 5 | 7 | 7 |
| 6 | 8 | 7.5 |
| 7 | 8.5 | 8 |
| 8 | 9 | 8 |
| 9 | 9 | 8.5 |
| 10 | 7.5 | 7 |
| 11 | 9 | 8.5 |

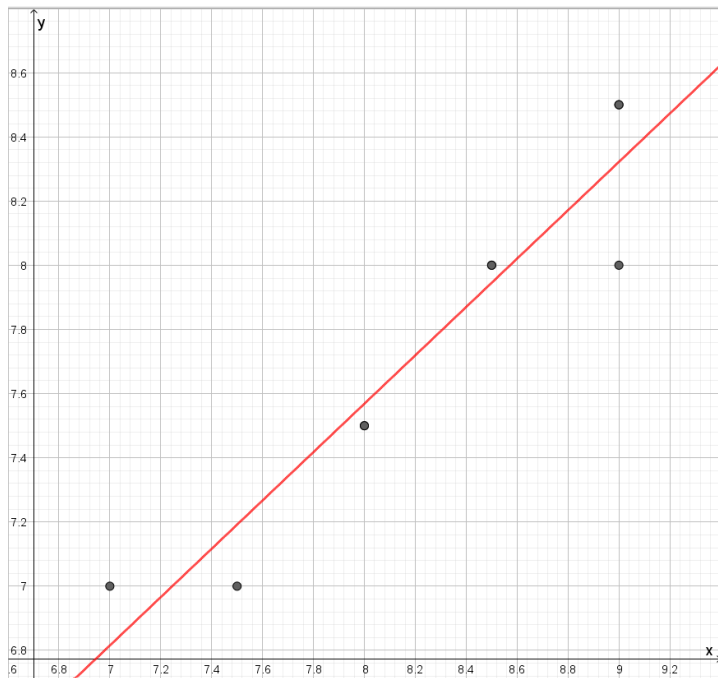
Piirretään hajontakuviota ja määritetään korrelaatiokerroin.



| | |
|----------------|--------|
| KeskiarvoX | 8.1 |
| KeskiarvoY | 7.8 |
| Sx | 0.8756 |
| Sy | 0.5375 |
| r | 0.6375 |
| ρ | 0.7333 |
| Sxx | 6.9 |
| VarianssiY | 2.6 |
| Sxy | 2.7 |
| | |
| R ² | 0.4064 |
| SSE | 1.5435 |

Korrelaatiokertoimen $r = 0,64$ perusteella tuomarien pisteiden välillä näyttää olevan huomattava positiivinen lineaarinen riippuvuus.

Hajontakuviosta nähdään, että aineistoon sisältyy yksi muista selvästi poikkeava havainto. Poistetaan taulukosta rivillä 3 oleva havainto, piirretään uusi hajontakuviio ja määritetään korrelaatiokerroin.



| | |
|------------|--------|
| KeskiarvoX | 8.2778 |
| KeskiarvoY | 7.7778 |
| Sx | 0.712 |
| Sy | 0.5652 |
| r | 0.9491 |
| ρ | 0.9516 |
| Sxx | 4.0556 |
| VarianssiY | 2.5556 |
| Sxy | 3.0556 |
| | |
| R^2 | 0.9008 |
| SSE | 0.2534 |

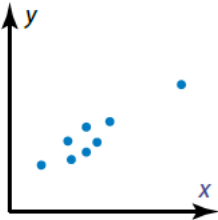
Nyt korrelaatiokertoimen $r = 0,95$ perusteella tuomarien pisteiden välinen lineaarinen riippuvuus onkin voimakkuudeltaan voimakas.

Vastaus

Tuloksiin sisältyy yksi poikkeava havainto. Kun se poistetaan, lineaarisen riippuvuuden voimakkuus muuttuu huomattavasta voimakkaaksi.

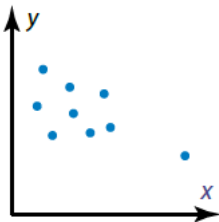
7.13

- a) Tuloksiin sisältyy yksi poikkeava havaintoarvo.



Poikkeava havaintoarvo on koordinaatistossa ylävasemmalla. Kun se poistetaan, korrelaatio säilyy edelleen voimakkaana, sillä havaintoarvot sopivat edelleen hyvin suoralle.

- b) Tuloksiin sisältyy yksi poikkeava havaintoarvo.



Poikkeava havaintoarvo on koordinaatistossa yläoikealla. Kun se poistetaan, korrelaatio heikkenee, sillä havaintoarvot sopivat huonommin suoralle.

Vastaus

- a) Tuloksiin sisältyy yksi poikkeava havainto. Kun se poistetaan, korrelaatio säilyy voimakkaana.
- b) Tuloksiin sisältyy yksi poikkeava havainto. Kun se poistetaan, korrelaatio heikkenee.

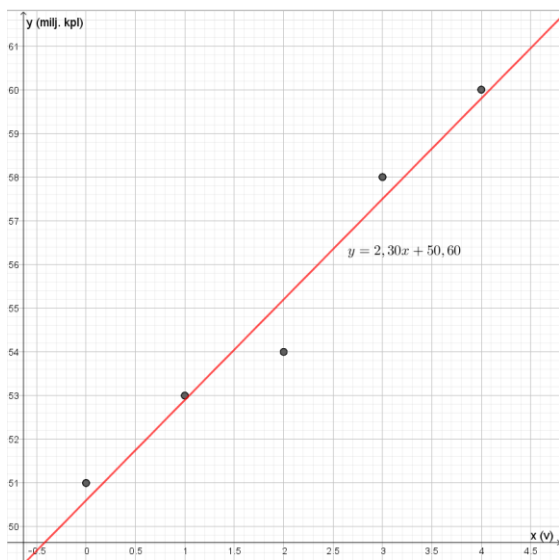
7.14

a) Ratkaistaan tehtävä GeoGebran taulukkolaskennalla.

Muuttuja x on vuodesta 2010 kulunut aika vuosina ja muuttuja y pakettien määrä. Taulukoidaan havaintoarvot.

| | A | B |
|---|--------------|-----------------------------|
| 1 | Aika x (v) | Pakettien määrä (milj. kpl) |
| 2 | 0 | 51 |
| 3 | 1 | 53 |
| 4 | 2 | 54 |
| 5 | 3 | 58 |
| 6 | 4 | 60 |

Piirretään hajontakuvio, tehdään lineaarinen sovitus ja määritetään korrelaatiokerroin.

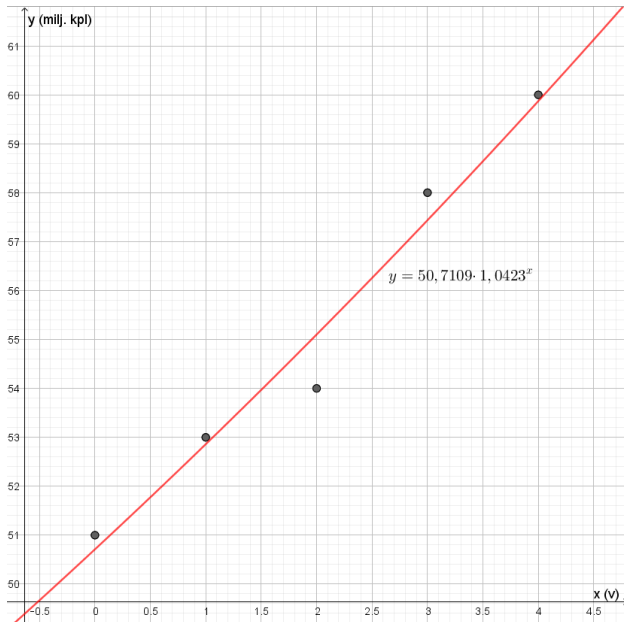


| | |
|------------|--------|
| KeskiarvoX | 2 |
| KeskiarvoY | 55.2 |
| Sx | 1.5811 |
| Sy | 3.7014 |
| r | 0.9825 |
| ρ | 1 |
| Sxx | 10 |
| VarianssiY | 54.8 |
| Sxy | 23 |
| | |
| R^2 | 0.9653 |
| SSE | 1.9 |

Hajontakuvion ja selityksasteen $r^2 \approx 0,97$ perusteella muuttujien välillä näyttää olevan voimakas positiivinen lineaarinen riippuvuus.

Regressiosuoran yhtälö $y = 2,30x + 50,60$ antaa vuoden 2023 ennusteeksi $y = 2,30 \cdot 13 + 50,60 \approx 81$ pakettia.

Kokeillaan seuraavaksi eksponentiaalista mallia.



| | |
|------------|--------|
| KeskiarvoX | 2 |
| KeskiarvoY | 55.2 |
| Sx | 1.5811 |
| Sy | 3.7014 |
| r | 0.9825 |
| ρ | 1 |
| Sxx | 10 |
| VarianssiY | 54.8 |
| Sxy | 23 |
| | |
| R^2 | 0.9699 |
| SSE | 1.6492 |

Hajontakuvion ja selitystasteen $r^2 \approx 0,97$ perusteella muuttujien välistä riippuvuutta kuvaa hyvin myös eksponentiaalinen riippuvuus (Kasvu).

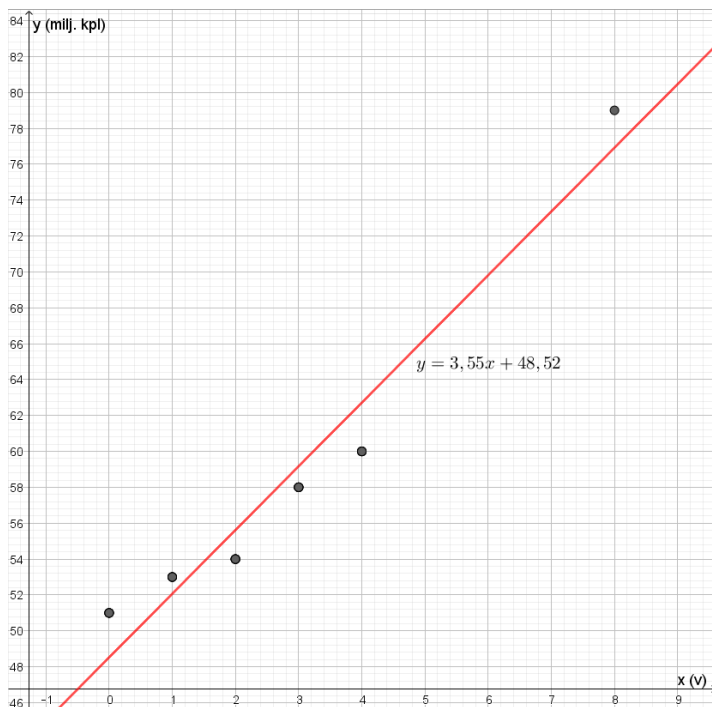
Neljän desimaalin tarkkuudella ilmoitettu regressiokäyrän yhtälö

$y = 50,7109 \cdot 1,0423^x$ antaa vuoden 2023 ennusteeksi

$y = 50,7109 \cdot 1,0423^{13} \approx 87$ miljoonaa pakettia.

Myös epätarkempaa regressiokäyrän yhtälöä voi käyttää. Silloin ennuste saattaa hieman muuttua.

- b)** Lisätään taulukkoon havaintoarvo $x = 8$ ja $y = 79$, piirretään uusi hajontakuvio ja tehdään ensiksi lineaarinen sovitus.

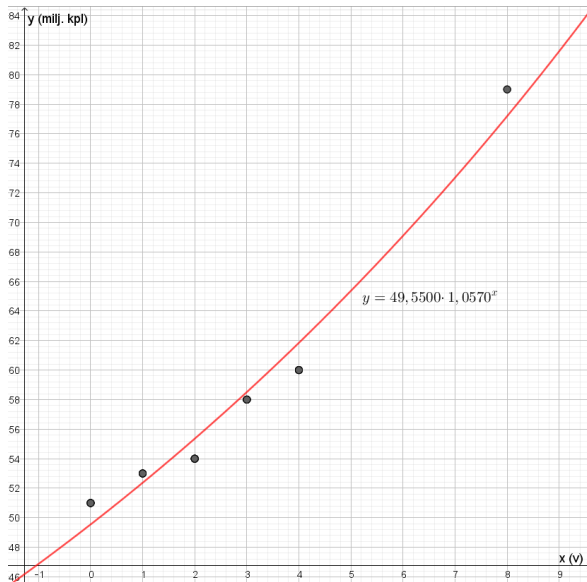


| | |
|----------------|----------|
| KeskiarvoX | 3 |
| KeskiarvoY | 59.1667 |
| Sx | 2.8284 |
| Sy | 10.2648 |
| r | 0.9782 |
| ρ | 1 |
| Sxx | 40 |
| VarianssiY | 526.8333 |
| Sxy | 142 |
| | |
| R ² | 0.9568 |
| SSE | 22.7333 |

Hajontakuvion ja selityssasteen $r^2 \approx 0,96$ perusteella muuttujien välillä näyttää olevan voimakas positiivinen lineaarinen riippuvuus.

Regressiosuoran yhtälö $y = 3,55x + 48,52$ antaa vuoden 2023 ennusteeksi $y = 3,55 \cdot 13 + 48,52 \approx 95$ pakettia.

Kokeillaan seuraavaksi eksponentiaalista mallia.



| | |
|----------------|----------|
| KeskiarvoX | 3 |
| KeskiarvoY | 59.1667 |
| Sx | 2.8284 |
| Sy | 10.2648 |
| r | 0.9782 |
| p | 1 |
| Sxx | 40 |
| VarianssiY | 526.8333 |
| Sxy | 142 |
| | |
| R ² | 0.9786 |
| SSE | 11.2749 |

Hajontakuvion ja selityssasteen $r^2 \approx 0,98$ perusteella muuttujien välistä riippuvuutta kuvaa hyvin myös eksponentiaalinen riippuvuus (Kasvu).

Neljän desimaalin tarkkuudella ilmoitettu regressiokäyrän yhtälö

$$y = 49,5500 \cdot 1,0570^x \text{ antaa vuoden 2023 ennusteeksi}$$

$$y = 49,5500 \cdot 1,0570^{13} \approx 102 \text{ miljoonaa pakettia.}$$

- c) Kaikki ennusteet sisältävät epävarmuutta. Ne myös poikkeavat toisistaan jonkin verran. Varmemman ennusteen tekemiseen tarvitaan enemmän havaintoarvoja.

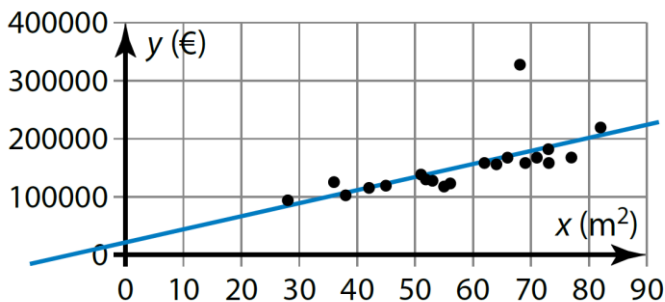
Vastaus

- a) Lineaarisen mallin ($y = 2,30x + 50,60$ $r^2 \approx 0,97$) mukaan
81 miljoonaa kappaletta.
Eksponentiaalisen mallin ($y = 50,7109 \cdot 1,0423^x$; $r^2 \approx 0,97$) mukaan
87 miljoonaa kappaletta.
- b) Lineaarisen mallin ($y = 3,55x + 48,52$ $r^2 \approx 0,96$) mukaan
81 miljoonaa kappaletta.
Eksponentiaalisen mallin ($y = 49,5500 \cdot 1,0570^x$ $r^2 \approx 0,98$) mukaan
102 miljoonaa kappaletta.
- c) Molemmat ennusteet sisältävät epävarmuutta. Varmemman ennusteen tekemiseen tarvitaan enemmän havaintoarvoja.

7.15

Ratkaistaan tehtävä GeoGebran taulukkolaskennalla. Avataan aineisto GeoGebralla.

Piirretään hajontakuvio.



Kuviossa on poikkeava havaintoarvo (68, 329 000). Aineistossa arvo sijaitsee rivillä 21.

- a) Lasketaan pinta-alojen keskiarvo ja keskihajonta koko aineistosta.

Keskiarvo on $58,1 \text{ m}^2$ ja keskihajonta $14,9 \text{ m}^2$.

| | |
|--------------|---------|
| n | 20 |
| Keskiarvo | 58.05 |
| σ | 14.5344 |
| s | 14.9119 |
| Σx | 1161 |
| Σx^2 | 71621 |
| Min | 28 |
| Q1 | 48 |
| Mediaani | 59 |
| Q3 | 70 |
| Max | 82 |

Poistetaan aineistosta poikkeava arvo ja lasketaan keskiarvo ja keskihajonta uudestaan.

Ilman poikkeavaa arvoa keskiarvo on $57,5 \text{ m}^2$ ja keskihajonta $15,1 \text{ m}^2$. Poikkeava arvo ei vaikuta merkittävästi pinta-alan keskiarvoon eikä keskihajontaan.

| | |
|--------------|---------|
| n | 19 |
| Keskiarvo | 57.5263 |
| σ | 14.7269 |
| s | 15.1304 |
| Σx | 1093 |
| Σx^2 | 66997 |
| Min | 28 |
| Q1 | 45 |
| Mediaani | 56 |
| Q3 | 71 |
| Max | 82 |

- b) Lasketaan hintojen keskiarvo ja keskihajonta koko aineistosta.

Keskiarvo on 153 925 € ja keskihajonta 50 988 €.

| | |
|--------------|--------------|
| n | 20 |
| Keskiarvo | 153925 |
| σ | 49696.6485 |
| s | 50987.6856 |
| Σx | 3078500 |
| Σx^2 | 523253250000 |
| Min | 95000 |
| Q1 | 122000 |
| Mediaani | 148000 |
| Q3 | 168500 |
| Max | 329000 |

Poistetaan aineistosta poikkeava arvo ja lasketaan keskiarvo ja keskihajonta uudestaan.

Ilman poikkeavaa arvoa keskiarvo on 144 711 € ja keskihajonta 30 850 €. Poikkeava arvo nostaa sekä keskiarvoa että keskihajontaa.

| | |
|--------------|--------------|
| n | 19 |
| Keskiarvo | 144710.5263 |
| σ | 30026.8808 |
| s | 30849.6875 |
| Σx | 2749500 |
| Σx^2 | 415012250000 |
| Min | 95000 |
| Q1 | 120000 |
| Mediaani | 139000 |
| Q3 | 168000 |
| Max | 220000 |

- c) Määritetään korrelaatiokerroin, kun käytetään koko aineistoa.

Koko aineiston korrelaatiokertoimen $r = 0,66$ perusteella pinta-alan ja hinnan välillä näyttää olevan kohtalainen positiivinen lineaarinen riippuvuus.

| | |
|----------------|------------------|
| KeskiarvoX | 58.05 |
| KeskiarvoY | 153925 |
| Sx | 14.9119 |
| Sy | 50987.6856 |
| r | 0.657 |
| p | 0.8731 |
| Sxx | 4224.95 |
| VarianssiY | 49395137500 |
| Sxy | 9491575 |
| | |
| R ² | 0.4317 |
| SSE | 28071809181.1737 |

Poistetaan aineistosta poikkeava arvo ja määritetään korrelaatiokerroin uudestaan.

Ilman poikkeavaa arvoa korrelaatiokerroin on $r = 0,91$ eli pinta-alan ja hinnan välillä näyttää olevan voimakas positiivinen lineaarinen riippuvuus.

| | |
|----------------|------------------|
| KeskiarvoX | 57.5263 |
| KeskiarvoY | 144710.5263 |
| Sx | 15.1304 |
| Sy | 30849.6875 |
| r | 0.9115 |
| p | 0.8941 |
| Sxx | 4120.7368 |
| VarianssiY | 17130657894.7368 |
| Sxy | 7657894.7368 |
| | |
| R ² | 0.8307 |
| SSE | 2899379837.5354 |

- d) Määritetään regressiosuoran yhtälö, kun käytetään aineistoa, josta on poistettu poikkeava arvo.

Regressiosuoran yhtälö on $y = 1858,38x + 37804,77$. Sijoitetaan yhtälöön poistetun asunnon pinta-ala $x = 68$ (m²).

$$y = 1858,38 \cdot 68 + 37804,77 \approx 164000$$

Poistetun asunnon hinta olisi regressiosuoran yhtälön perusteella 164 000 €.

Vastaus

- a) Koko aineisto: keskiarvo $58,1 \text{ m}^2$ ja keskihajonta $14,9 \text{ m}^2$.
Ilman poikkeavaa arvoa: keskiarvo $57,5 \text{ m}^2$ ja keskihajonta $15,1 \text{ m}^2$.
Poikkeava arvo ei vaikuta merkittävästi.
- b) Koko aineisto: keskiarvo $153\,925 \text{ €}$ ja keskihajonta $50\,988 \text{ €}$.
Ilman poikkeavaa arvoa: keskiarvo $144\,711 \text{ €}$ ja keskihajonta $30\,850 \text{ €}$.
Poikkeava arvo nostaa keskiarvoa ja keskihajontaa.
- c) Koko aineisto: $r = 0,66$; kohtalainen lineaarinen riippuvuus.
Ilman poikkeavaa arvoa: $r = 0,91$; voimakas lineaarinen riippuvuus.
- d) $164\,000 \text{ €}$

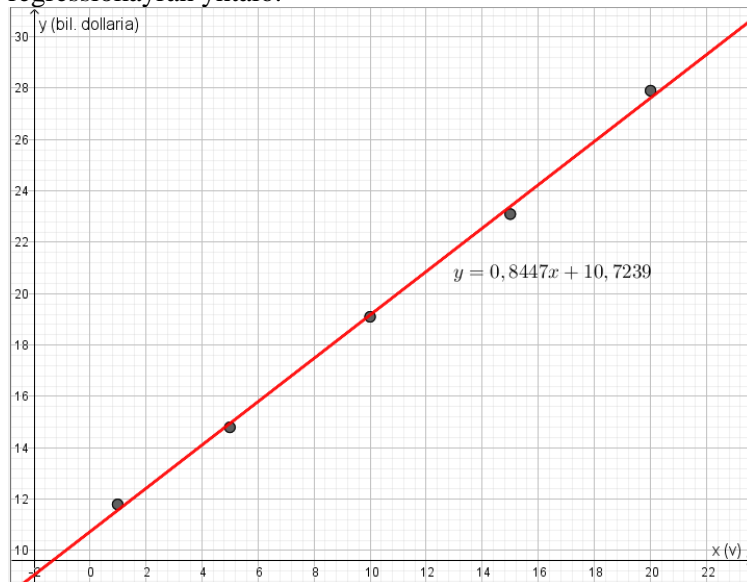
7.16

- a) Ratkaistaan tehtävä GeoGebran taulukkolaskennalla. Avataan aineisto GeoGebralla.

Muuttuja x on vuodesta 1960 kulunut aika vuosina ja muuttuja y bruttokansantuote. Taulukoidaan kulunut aika.

| | A | B | C |
|---|-------|--------------|----------------------|
| 1 | Vuosi | Aika x (v) | GWP y |
| 2 | | | (biljoonaa dollaria) |
| 3 | 1961 | 1 | 11.8 |
| 4 | 1965 | 5 | 14.8 |
| 5 | 1970 | 10 | 19.1 |
| 6 | 1975 | 15 | 23.1 |
| 7 | 1980 | 20 | 27.9 |

Huomioidaan ainoastaan vuodet 1961–1980. Piirretään hajontakuviota, kokeillaan, mikä sovitus sopii aineistoon parhaiten ja määritetään regressiokäyrän yhtälö.



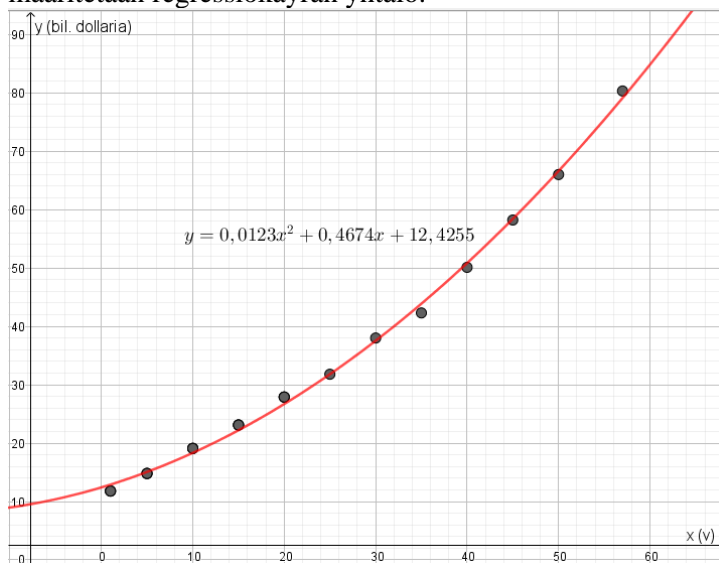
Lineaarinen malli sopii aineistoon parhaiten.

| | |
|----------------|---------|
| KeskiarvoX | 10.2 |
| KeskiarvoY | 19.34 |
| Sx | 7.5961 |
| Sy | 6.4213 |
| r | 0.9993 |
| ρ | 1 |
| Sxx | 230.8 |
| VarianssiY | 164.932 |
| Sxy | 194.96 |
| | |
| R ² | 0.9985 |
| SSE | 0.2466 |

Regressiosuoran yhtälö $y = 0,8447x + 10,7239$ antaa vuoden 2050 ennusteeksi $y = 0,8447 \cdot 90 + 10,7239 = 86,7$ biljoonaa dollaria.

Regressiosuoran yhtälön voi määrittää myös epätarkemmaksi, jolloin ennuste saattaa poiketa hieman tästä.

- b) Huomioidaan seuraavaksi kaikki aineiston vuodet. Piirretään hajontakuvi, kokeillaan, mikä sovitus sopii aineistoon parhaiten ja määritetään regressiokäyrän yhtälö.



2. asteen polynominen malli sopii aineistoon parhaiten.

| | |
|----------------|-----------|
| KeskiarvoX | 27.75 |
| KeskiarvoY | 38.6167 |
| Sx | 18.1765 |
| Sy | 21.5027 |
| r | 0.9856 |
| p | 1 |
| Sxx | 3634.25 |
| VarianssiY | 5086.0167 |
| Sxy | 4237.55 |
| | |
| R ² | 0.9982 |
| SSE | 9.2049 |

Regressiokäyrän yhtälö $y = 0,0123x^2 + 0,4674x + 12,4255$ antaa vuoden 2050 ennusteeksi

$$y = 0,0123 \cdot 90^2 + 0,4674 \cdot 90 + 12,4255 \approx 154 \text{ biljoonaa dollaria.}$$

- c) Vuoden 1961 bruttokansantuote oli 11,8 biljoonaa dollaria.

Kymmenkertainen bruttokansantuote on 118 biljoonaa dollaria.

Sijoitetaan a-kohdan regressiokäyrän yhtälöön $y = 0,8447x + 10,7239$

bruttokansantuotteeksi $y = 118$ ja ratkaistaan aika x CAS-laskimella.

$$118 = 0,8447x + 10,7239$$

$$x \approx 127$$

Lineaarisen mallin mukaan bruttokansantuote on kasvanut kymmenkertaiseksi vuonna $1960 + 127 = 2087$.

Sijoitetaan b-kohdan regressiokäyrän yhtälöön

$$y = 0,0123x^2 + 0,4674x + 12,4255 \text{ bruttokansantuotteeksi } y = 118$$

ja ratkaistaan aika x CAS-laskimella.

$$118 = 0,0123x^2 + 0,4674x + 12,4255$$

$$x \approx 76 \text{ (tai } x \approx -114)$$

Mallin perusteella negatiivinen ratkaisu voidaan hylätä.

2 asteen polynomisen mallin mukaan bruttokansantuote on kasvanut kymmenkertaiseksi vuonna $1960 + 76 = 2036$.

Vastaus

- a) lineaarisen mallin mukaan 86,7 biljoonaa dollaria
- b) 2. asteen polynomisen mallin mukaan 154 biljoonaa dollaria
- c) lineaarisen mallin mukaan vuonna 2087,
2. asteen polynomisen mallin mukaan vuonna 2036

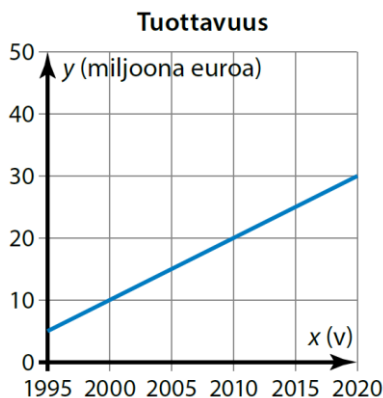
7.17

Aika-akseli ei ole tasavälinen, vaan yksittäinen väli kuvaa eri aikaa. Akselin alkupäässä yksi väli on 10 vuotta ja loppupäässä yksi vuosi.

Piirretään korjattu kuvio LibreOffice Calc -ohjelmistolla. Syötetään tehtävän kuvioista saadut tiedot taulukkolaskennan sarakkeisiin.

| | A | B |
|---|--------------|----------------------------|
| 1 | x (v) | y (miljoonaa euroa) |
| 2 | 1995 | 6 |
| 3 | 2005 | 15 |
| 4 | 2010 | 20 |
| 5 | 2013 | 23 |
| 6 | 2016 | 26 |
| 7 | 2017 | 27 |
| 8 | 2019 | 29 |
| 9 | 2020 | 30 |

Aika-akselin saa tasaväliiseksi valitsemalla **Ohjattu kaavion luonti** -valikosta **XY (hajonta)**.



Tuottavuus on todellisuudessa kasvanut tasaisesti.