PIENIMMÄN NELIÖSUMMAN MENETELMÄ

Tekijät: Ida Arhosalo, Otso Huuska, Leena Kalliovirta, Mikko Kulmunki, Jouni Marjakangas, Hamidreza Maleki, Tatu Mustonen

Laadittu keväällä 2023.

Kohderyhmä: Yläkoulun tai lukion lyhyen matematiikan tilasto-opinnot

Esitiedot: Oppilas tuntee suoran yhtälön ja tietää, että kahden pisteen avulla suora voidaan yksikäsitteisesti määrittää. Lisäksi hän tietää itseisarvon käsitteen, osaa laskea itseisarvoja ja lukujen toisia potensseja sekä käyttää Geogebraa tai muuta taulukkolaskentaohjelmaa.

Oppitunnin kesto: n. 60 min

Oppimistavoitteet: Oppitunnin jälkeen oppilas ymmärtää, millä mekanismilla ohjelmat tekevät suoran sovituksen pistejoukkoon ja hahmottaa, että sovite on yksikäsitteinen.

Muut tavoitteet: Oppilas osaa tehdä sovitteen jollakin ohjelmistolla. Oppilas osaa hyödyntää saatua suoran yhtälöä ennustamiseen.

Sisältö ja toteutus:

- Datan keruu esimerkkiin 5 (5min)
- Johdattelu aiheeseen sisältäen tutkimus simulaatiolla (10 min)
- Pienimmän neliösumma menetelmän esittely (10 min)
- Geogebraharjoitus pienimmän neliösumman menetelmään (15 min)
- Esimerkkejä (n. 20 min)
- Kotitehtävän tekoa jos aikaa on jäljellä

Oppitunnissa on asiaa aiheen pidempäänkin kässittelyyn kuin 60 minuuttiin. Opettaja voi halutessaan painottaa simulaatiojohdattelua enemmän ja jättää teoriaosan vähemmälle, tai käydä teoriaa huolellisemmin ja jättää esimerkkejä kotitehtäviksi, tai soveltaa aikataulua muilla vastaavilla tavoilla.

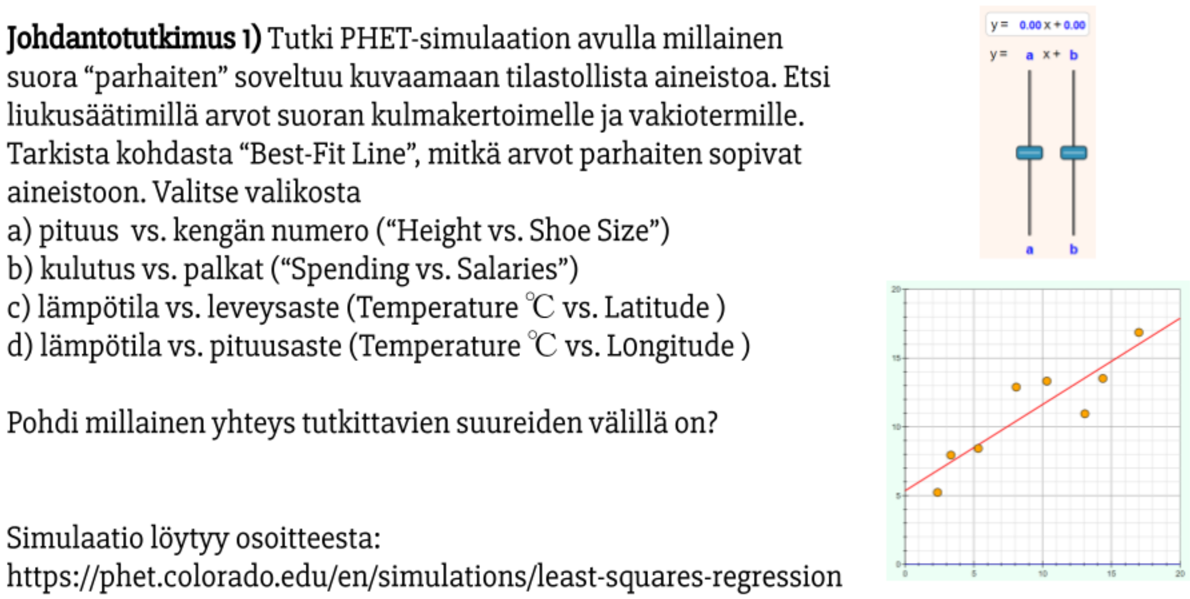
Datan keruun voi suorittaa oppitunnin aluksi tai valmiiksi etukäteen. Katso kohta 5.

1. Motivaatio

Todellisen maailman ilmiöihin liittyy aina satunnaisuutta ja epävarmuutta. Satunnaisuus voi liittyä esimerkiksi tutkimusasetelmaan, mittausvälineiden tarkkuuteen ja muihin tutkittavan ilmiöiden ominaisuuksiin. *Pienimmän neliösumman suora* on eräs työkalu tutkia kahden tilastollisen muuttujan välistä selityssuhdetta. Asetelma on epäsymmetrinen, jossa toinen muuttuja asetetaan selittäväksi muuttujaksi ja toinen selitettäväksi. Pienimmän neliösumman suoran menetelmässä ideana on sovittaa havantoparijoukkoon suora, joka edustaa aineistoa ”parhaiten”. Menetelmä on tärkeä tilastollisessa päättelyssä, ja sen avulla voidaan mahdollisesti tehdä ennusteita tutkittavasta ilmiöstä.

Eräs yksinkertainen esimerkki kahden tilastollisen muuttujan tutkimuksesta voi olla vaikkapa, kuinka hyvin henkilön kengännumero selittää henkilön pituutta.

Osoitteesta PHET simulaatio voi tutustua simulaation avulla pienimmän neliösumman suoran sovittamiseen helposti lähestyttävässä visuaalisessa muodossa. Simulaatioon liittyvän johdantotehtävän löydät kuvasta 1.

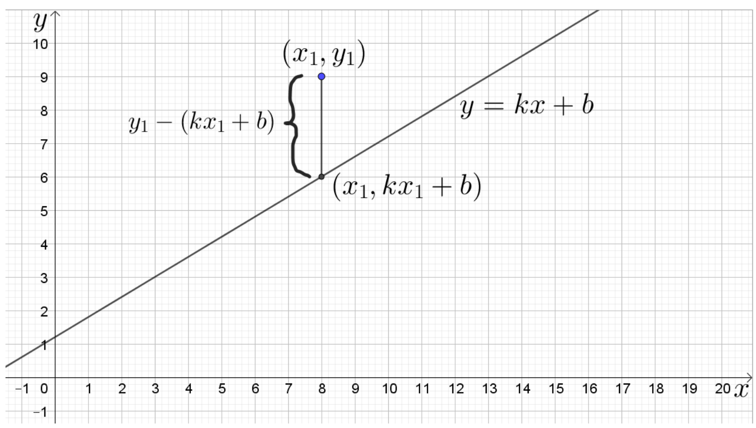


Kuva 1. Johdantotutkimus PNS-suoraan, linkki simulaatioon PHET simulaatio, josta kuva on peräisin.

2. Pienimmän neliösumman menetelmä

Pyrkimyksenä on etsiä eli sovittaa koordinaatistoon sellainen suora, joka kuvaa pisteiden sijaintia mahdollisimman hyvin. Yksi mahdollisuus on laskea jokaisen pisteen (x, y) etäisyys sovitettavasta suorasta y-suunnassa ja etsiä suoraa, joissa nämä yhteenlasketut etäisyydet ovat mahdollisimman pieniä.

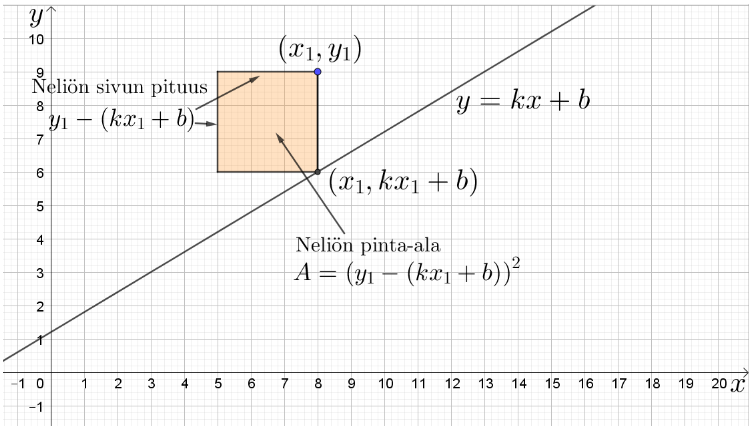
Merkitään parasta sovitettua suoraa $y = kx + b$. Tällöin pisteen (x_1, y_1) etäisyys suorasta y-suunnassa on $|y_1 - (kx_1 + b)|$.



Kuva 2. Periaatekuva havaintopisteen (x_1, y_1) etäisyys PNS-suorasta.

Osoittautuu, että on parempi tutkia etäisyyksien neliöiden summaa kuin etäisyyksien summaa[1]. Neliö on myös aina positiivinen, jolloin itseisarvoja ei tarvita.

Pisteen etäisyyden neliö on $(y_1 - (kx_1 + b))^2$. Paras suora on sellainen, jolle näiden neliöiden summa on pienin. Kuvana tämä tarkoittaa etäisyyksistä muodostettujen neliöiden pinta-alojen summaa.



Kuva 3. Periaatekuva havaintopistettä (x_1, y_1) vastaava etäisyyden neliö.

On olemassa kaava[1], joka antaa parhaiten sovitetun suoran kulmakertoimen ja vakiotermin suoraan pisteiden koordinaateista. Kaavan johtamisessa vaaditaan derivointia, joka sisältyy lukion pitkään matematiikkaan ja perusteita käydään myös lyhyen matematiikan valinnaisissa opinnoissa. Jotta kaava ei näytä kovin pitkältä ja työläältä, esitetään se palasina ja määritellään ensin muutama apumuuttuja.

Pisteet, joihin suora sovitetaan, ovat $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ eli n on niiden pisteiden lukumäärä, joihin suora sovitetaan.

Merkitään kaikkien pisteiden x-koordinaattien summaa symbolilla D:

$$D = x_1 + x_2 + \cdots + x_n.$$

Merkitään kaikkien pisteiden y-koordinaattien summaa symbolilla E:

$$E = y_1 + y_2 + \cdots + y_n.$$

Merkitään kaikkien pisteiden x-koordinaattien neliöiden summaa symbolilla F:

$$F = x_1^2 + x_2^2 + \cdots + x_n^2.$$

Merkitään kaikkien pisteiden x- ja y-koordinaattien tulojen summaa symbolilla G:

$$G = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Näiden summien avulla saadaan laskettua suoran kulmakerroin ja vakiotermi:

$$k = \frac{nG - DE}{nF - D^2},$$

$$b = \frac{FE - DG}{nF - D^2}.$$

Tätä suoraa kutsutaan *pienimmän neliösumman suoraksi* eli *PNS-suoraksi*.

Kaavan käyttö käsin muuttuu todella työlääksi, jos havaintoaineisto on mielekkään kokoinen (eli pisteitä on ainakin noin 10). Summien laskeminen on kuitenkin täysin mekaanista ja tietokone tekee sen helposti ja nopeasti, vaikka havaintoaineiston koko olisi tuhansia pisteitä.

[1]Katso diat "9.1 Johdanto lukuun 9: lineaarinen regressio" tai Johdatus tilastolliseen päättelyyn.pdf, luku 9, jotka yliopistonlehtori Petri Koistinen, Helsingin yliopisto on laatinut vuonna 2013 Matematiikan ja Tilastotieteen laitoksen kurssia Tilastollinen päättely I varten.

3. Laskuesimerkki ja tilanne kolmella pisteellä

Tarkastellaan tilannetta, jossa suora sovitetaan kolmeen pisteeseen $(x_1, y_1), (x_2, y_2), (x_3, y_3)$. Pienimmän neliösumman menetelmä GeoGebralla. Avaa Geogebra appletti sovelluksessa ja lisää kaavat joilla lasket pienimmän neliösumman suoran kulmakertoimen k ja vakiotermin b. Tarkista ratkaisusi vertaamalla Geogebran SovitaSuora-komenolla piirretyn suoran yhtälöön.

4. Esimerkki: jousivakio

Lineaarinen sovitus on keskeinen työkalu kaikissa kokeellisissa luonnontieteissä. Monissa tilanteissa kaksi suuretta riippuvat toisistaan lineaarisesti, ja sovitetta voidaan käyttää ennustamiseen. Toisaalta se, kuinka tarkkaan havaintopisteet sopivat suoralle, kertoo siitä, kuinka hyvin tarkasteltu ilmiö noudattaa lineaarista mallia.

Jousi on eräs tilanne, jota voidaan tarkastella helposti kokeellisesti ja jossa lineaarinen malli on hyödyllinen. Havaintojen mukaan tyypillisessä jousessa venyttävä tai puristava voima riippuu lineaarisesti jousen pituudesta. Tällaista voimaa kutsutaan yleisesti *harmoniseksi.*

Ohessa on aineisto Jousidata.xlsx, jossa on tutkittu tiettyä jousta. Jousi asetettiin roikkumaan telineestä alaspäin. Tutkimuksessa jousta venytettiin eri pituuksiin ja venyttämiseen vaadittu voima mitattiin voima-anturilla. Syntynyt aineisto on liitteenä.

(1) Esitä mittausaineisto valitsemallasi ohjelmalla koordinaatistossa siten, että jousen pituudet ovat x-akselilla ja niitä vastaavat mitatut venyttävät voimat ovat y-akselilla.
(2) Sovita ohjelmalla mittausaineistoon suora pienimmän neliösumman menetelmää käyttäen.

Voidaan todeta, että tutkittu jousi käyttäytyy varsin tarkasti odotusten mukaan ja suora sopii mittauspisteisiin varsin hyvin.

Suoran kulmakerroin $k = \dfrac{\Delta y}{\Delta x}$. Siten kulmakertoimen yksiköksi tulee nyt N/m, koska y-akselilla on newtoneita ja vaaka-akselilla metrejä. Kulmakerrointa kutsutaan jousen *jousivakioksi* ja se kuvaa, montako newtonia vaaditaan, jotta jousi venyy metrin lisää.

(3) Poimi jousivakio sovittamastasi suorasta.
(4) Poimi myös vakiotermi ja päättele sen yksikkö.

Sovitesuoraa voidaan käyttää ennustamaan jousta tiettyyn pituuteen venyttävää voimaa tai päinvastoin. Suoraan $F = kx + b$ sijoitetaan tällöin tunnettu pituus ja ratkaistaan voima, tai päinvastoin.

(5) Ennusta, millainen voima vastaa jousen pituutta 32 cm.
(6) Ennusta, kuinka pitkä jousi on venyttävän voiman ollessa 140 N.
(7) Jousen pituus venyttämättömänä oli 25,0 cm. Laske, millaista venyttävää voimaa tämä vastaa mallin mukaan. Pohdi, mistä kertoo, että mallin mukaan voima ei ole venyttämättömänä nolla.

5. Esimerkki: kengänkoot ja pituudet

Kysytään opiskelijoita (voi kysyä jo edellisen oppitunnin lopussa tai tämän tunnin alkupuolella, jotta opettaja ehtii valmistella datan taulukkoon) lapulla tai sähköisellä kyselykaavakkeella kengänkoko ja henkilön pituus. Taulukko jaetaan opiskelijoille.
Kysymyksiä:

(1) Mikä on pituus kun kengänkoko on 38?
(2) Mikä on pituus kun kengänkoko on 42?
(3) Millainen kengänkoko on 170 cm pituisella ihmisellä?

6. Kotitehtävä

Liittenä on aineisto Data.txt, joka simuloitu käyttäen todellisen havaintoaineiston ominaisuksia siten, että simuloitua aineiston voidaan olettaa kuvaavan 78 satunnaisesti valitun miehen pituutta tuumissa ja kengännumeroa. *Ohje opettajalle: Tehtävän ideana on pohjustaa seuraavaa oppituntia, jossa pohditaan mallin selitysasteen merkitystä ja poikkeavan havainnon määritelmää.*

(1) Lataa aineisto taulukkolaskentaohjelmaan.
(2) Piirrä aineistosta hajontakuvio (eli kengännumerot ja pituudet pistepareina) siten, että x-akselilla on kengännumero ja y-akselilla on pituus.
(3) Onko henkilön kengännumeron ja pituuden välillä riippuvuutta tässä aineistossa? Vastaa tähän pelkän hajontakuvion perusteella.
(4) Sovita aineistoon PNS-suora taulukkolaskentaohjelman avulla. Tarkastele saatuja tuloksia.
(5) Onko aineistossa selvästi muista poikkeavia havaintoja?
(6) Jos poikkeavat havainnot poistetaan, kuinka PNS-suora muuttuu?

7. Tehtävien vastaukset

3. Laskuesimerkki:

$$k = \frac{3 \cdot 42 - 9 \cdot 10}{3 \cdot 35 - 9^2} = 1{,}5, \quad b = \frac{35 \cdot 10 - 9 \cdot 42}{3 \cdot 35 - 9^2} = -\frac{7}{6} \approx -1{,}17$$

4. Jousivakio:
Kohta 3: n. 830 N/m
Kohta 4: n. -200 N
Kohta 5: n. 63 N
Kohta 6: n. 41 cm
Kohta 7: n. 5,1 N. Kyseessä on voima-anturin systemaattinen virhe (anturi nollaamatta)

5. Kengänkoot ja pituudet: Vastaus riippuu aineistosta.

6. Kotitehtävä:
Kohta 3: Kyllä, suurempi kengännumero näyttäisi tarkoittavan myös pidempää ihmistä.
Kohta 5: PNS-suoran sovituksen perusteella henkilö, jonka kengännumero on 7,5 ja pituus 96 tuumaa on poikkeava havainto. Kyseessä lienee mittausvirhe tai tallennusvirhe.
Kohta 6: Kulmakerroin kasvaa ja termi b pienenee. Tämän lisäksi havaitaan, että selitysaste kasvaa arvosta 0,39 arvoon 0,88 eli selittämätön osa pituuksien vaihtelusta kengänkoon funktiona pienenee. Tietenkään havaintoja ei voi lähteä poistamaan analyysistä mielivaltaisesti, vaan poistoon on oltava aina painavat perustelut. Tässä tapauksessa 96 tuumaa eli 254 cm on niin poikkeava pituus, että kyseistä havaintoa voidaan pitää epäluotettavana ja poistaa sen vaikutus mallista.