

Johdatus tilastolliseen päättelyyn, kevät 2014
Harjoitus 5 (14.–25. 4.)

1. a) Näytä, että otoksista $\mathbf{x} = (x_1, \dots, x_n)$ ja $\mathbf{y} = (y_1, \dots, y_n)$ laskettu otoskovarianssi

$$s(\mathbf{x}, \mathbf{y}) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

voidaan lausua myös seuraavissa muodoissa:

$$s(\mathbf{x}, \mathbf{y}) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})y_i = \frac{1}{n-1} \left(\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y} \right). \quad (*)$$

- b) Näytä, että pistepareihin (x_i, y_i) sovitetun PNS-suoran yhtälö $y = a + bx$ (ks. monisteen sivu 107) voidaan esittää myös muodossa

$$\frac{y - \bar{y}}{s_y} = r(\mathbf{x}, \mathbf{y}) \frac{x - \bar{x}}{s_x},$$

jossa $s_x = \sqrt{s(\mathbf{x}, \mathbf{x})}$ ja $s_y = \sqrt{s(\mathbf{y}, \mathbf{y})}$ ovat otoskeskihajonnat ja $r(\mathbf{x}, \mathbf{y})$ on eräs luvuista x_i ja y_i laskettu kerroin. Mikä on $r(\mathbf{x}, \mathbf{y})$:n lauseke? Entä mitä nimitystä siitä käytetään? (Jos et ole sitä aikaisemmin nähnyt, selvitä asia kirjallisuudesta!)

Huom. Kohdan (b) esityksessä on suoritettu eräänlainen standardointi sekä x - että y -havaintoihin: keskiarvon vähentäminen merkitsee keskistämistä ja keskihajonnalla jakaminen skaalausta siten, että muunnettujen havaintojen keskihajonnaksi tulee 1.

2. JTP-kurssilla 2012 tehdyssä kyselytutkimuksessa (otoskoko $n = 37$) saatiin miesopiskelijoiden pituudelle x (cm) ja painolle y (kg) seuraavat yhteenvedot:

$$\bar{x} = 180.1, \quad \bar{y} = 77.07$$

$$s(\mathbf{x}, \mathbf{x}) = 56.52, \quad s(\mathbf{x}, \mathbf{y}) = 46.37, \quad s(\mathbf{y}, \mathbf{y}) = 184.56.$$

Mallinnetaan aineisto lineaarisen regressiomallin avulla.

- a) Laske mallin kertomien SU-estimaatit (eli PNS-estimaatit) ja ilmoita PNS-suoran yhtälö.
 b) Laske virhevarianssin estimaatti s^2 .
3. Jatkoa edelliseen tehtävään. Muodosta mallin parametrille β eli PNS-suoran kulmakertoimelle 95 %:n luottamusväli. Jääkö nollahypoteesi $H_0: \beta = 0$ (ts. pituudella ja painolla ei riippuvuutta) voimaan, jos sitä testataan merkitsevyystasolla 0.05 hypoteesia $H_1: \beta \neq 0$ vastaan?

Tarvittava t -jakauman taulukko on esimerkiksi 3. harjoitusten tehtäväpaperissa. Voit myös katsoa kurssin kotisivulta löytyviä linkkejä.

4. Monisteen kaavassa (9.14) sivulla 111 todetaan lineaarisen regressiomallin kulmakertoimen PNS-estimaattorille jakaumatulos

$$b(\mathbf{Y}) \sim N(\beta, \sigma^2/q_{xx}),$$

jossa $q_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$. Perustele tämä tulos laskemalla $b(\mathbf{Y})$:n odotusarvo ja varianssi; normaalisuus seuraa siitä yleisestä tuloksesta, että lineaarikombinaatiot riippumattomista normaalisti jakautuneista satunnaismuuttujista ovat normaalisti jakautuneita.

Ohje. Otoskovarianssille $b(\mathbf{Y})$:n kaavassa voit käyttää tehtävän 1 kaavan (*) keskimmäistä lauseketta. Muista korvata havainnot y_i vastaavilla satunnaismuuttujilla Y_i .

KÄÄNNÄ!

5. (Bayesin kaavan mieliinpalautus viimeistä viikkoa varten.) Vuonna 1975 uutisoitiin tutkimuksesta, jonka mukaan 50 % kanadalaisista miehistä käytti värillisiä alushousuja kun taas amerikkalaisista miehistä sellaisia käytti vain 20 %. Bermudalaisen hotellin asiakaskunta koostui yksinomaan amerikkalaisista ja kanadalaisista siten, että miesasiakkaista 80 % oli amerikkalaisia ja 20 % kanadalaisia. Todennäköisyyslaskentaa opiskellut siivooja huomasi miesasiakkaan huoneessa punaiset alushousut. Millä todennäköisyydellä hän päätteli asukkaan olevan kanadalaisen? Muotoile sopivat tapahtumat ja sovelle niihin Bayesin kaavaa.

Pääsiäisloma on 17.–23.4. Näitä harjoituksia käsitellään ennen (ma, ti) ja jälkeen (to, pe) pääsiäisen. Viimeiset luennot pidetään to 24.4. ja pe 25.4. ja viimeiset (kuudennet) harjoitukset 28.4.–2.5. (ei kuitenkaan to 1.5.).