

Tilastollinen päättely II, syksy 2015 – kevät 2016
Harjoitus 13 (1. ja 3. 3. 2016)

Tehtävät (paitsi viimeinen) liittyvät luottamusväleihin ja -joukkoihin. Monisteen luku 6.

1. (Monisteen teht. 6.1.) Olkoot $Y_1, \dots, Y_{25} \sim N(\mu, \sigma^2) \perp$ ja $S^2 = \sum_{i=1}^{25} (Y_i - \bar{Y})^2 / 24$. Palauta todennäköisyyslaskennan kurssilta mieleen, miten S^2 / σ^2 on jakautunut; erityisesti sen jakauma ei riipu parametreista μ ja σ^2 . Etsi tämän avulla keskihajonnalle σ ylempi 95 %:n luottamusraja eli luottamusväli muotoa $(0, b)$, kun on havaittu $s = 10$.

2. Oletetaan, että eräässä mallissa on löydetty tunnusluku eli aineiston muunnos $t = t(\mathbf{y})$ siten, että vastaavalle satunnaismuuttujalle $T = t(\mathbf{Y})$ pätee $T - \theta \sim \text{Tas}(-1, 1)$ kaikilla θ :n arvoilla. Kyseessä on siis *saranasuure* monisteen kohdan 6.2.2 mukaisesti. Johda tämän avulla jokin 95 %:n luottamusväli θ :lle.

3. (Monisteen teht. 6.2.) Olkoot $Y_1 \perp Y_2$ ja $Y_1 \sim N(\mu_1, 1)$ sekä $Y_2 \sim N(\mu_2, 1)$. Etsi luvut $a, b > 0$ siten, että

$$P\{|Y_1 - \mu_1| \leq a, |Y_2 - \mu_2| \leq a\} = 0.95,$$
$$P\{(Y_1 - \mu_1)^2 + (Y_2 - \mu_2)^2 \leq b^2\} = 0.95.$$

Aineisto on $(y_1, y_2) = (1, 0.5)$. Mitkä kaksi 95 %:n luottamusjoukkoa saadaan yo. yhtälöiden perusteella parametriparille (μ_1, μ_2) ? Piirrä kuva. Kumpi luottamusjoukoista on mielestäsi parempi? [Ohje. Tarvitset jakaumien $N(0, 1)$ ja χ_2^2 taulukoita tai laskimia.]

4. **Vanha koetehtävä.** Oletetaan, että Y_1, \dots, Y_n ovat riippumattomia ja noudattavat kukin jatkuvaa jakaumaa, jonka tiheysfunktio on

$$f(y; \lambda) = 2\lambda y \exp(-\lambda y^2), \quad y > 0,$$

ja jossa λ on positiivinen parametri.

a) Muodosta tilastollinen malli $f_{\mathbf{Y}}$ ja johda parametrin λ suurimman uskottavuuden estimaattori $\hat{\lambda}$ sekä Fisherin informaatio $i(\lambda)$.

b) Mitä normaalijakaumaa $\hat{\lambda}$ approksimatiivisesti noudattaa, kun n on suuri?

c) Muodosta Waldin testiin eli yo. normaaliapproksimaatioon perustuva approksimatiivinen 95 %:n luottamusväli parametrille λ , kun aineisto on $\mathbf{y} = (y_1, \dots, y_n)$. Mikä on $\hat{\lambda}$:n keskivirhe?

5. Osoitteessa <http://www.fsd.uta.fi/menetelmaopetus/paattely/paattely.html#luottamusvali> kuvataan luottamusvälin ja luottamustason käsitteitä seuraavasti:

”Tilastollisen päättelyn kaksi keskeistä käsitettä ovat luottamusväli ja luottamustaso. **Luottamusväli** (*confidence interval*) kertoo millä välillä todellinen perusjoukon tunnusluvun arvo on tietyllä todennäköisyydellä. Käyttäen edelleen Nato-kyselyä esimerkkinä, voidaan kuvitella, että otoksessa 45 % kaikista vastaajista ilmoitti kannattavansa Suomen Nato-jäsenyyttä. Koska tähän lukuun vaikuttavat monet satunnaiset tekijät, emme voi suoraan päätellä, että myös perusjoukossa (kaikki täysi-ikäiset suomalaiset) vastaava osuus on täysin sama. On kuitenkin todennäköistä, että perusjoukon mielipidettä kuvaava arvo on lähellä otoksesta saatua arvoa. Voimme esimerkiksi päätellä, että 95 %:n todennäköisyydellä Nato-jäsenyyttä kannattavien ihmisten osuus perusjoukossa on välillä 40–50 %. Tätä väliä kutsutaan luottamusväliksi.

Luottamustaso (*confidence level*) kertoo, millä todennäköisyydellä perusjoukkoa kuvaava tunnusluku on jollain tietyllä luottamusvälillä. Esimerkiksi 95 %:n todennäköisyydellä 40–50 % suomalaisista haluaa Suomen liittyvän Natoon. Luottamustaso on tällöin 95 %:n todennäköisyys.

Luottamustaso ja luottamusväli ovat siis täysin toisiinsa sitoutuneita käsitteitä. Tieto luottamusvälistä ei ole mielekäs, jos ei ole tietoa luottamustasosta ja päinvastoin. Olennaista on, että luottamustason kasvaessa laajenee myös luottamusväli. Toisin sanoen tämä tarkoittaa siis sitä, että mitä suuremmalla varmuudella haluamme tietää, millä välillä jokin perusjoukon tunnusluku sijaitsee, sitä suurempi on luottamusväli. Jos esimerkiksi haluaisimme tietää, millä välillä suomalaisten Nato-jäsenyyden kannatus on 99 %:n luottamustasolla, luottamusväli olisi suurempi kuin 95 prosentin luottamustasolla (esimerkiksi 30–60 %). Jos olisimme valmiita tyytymään esimerkiksi 90 %:n luottamustasoon, väli voisi olla 43–47 %.”

Mikä tässä kuvauksessa on hyvää ja mikä antaa kritiikin aihetta?

6. Keskusteltavaksi:

- a) Mainitse pari kolme mielestäsi tärkeintä tai mielenkiintoisinta tilastollisen päättelyn piiriin kuuluvaa käsitettä tai menetelmää, jotka olet tällä kurssilla oppinut. Perustele!
- b) Miten käsityksesi tilastollisesta päättelystä on muuttunut tällä kurssilla verrattuna *Johdatus tilastolliseen päättelyyn* -kurssiin (tai vastaavaan)?

Huom. Toinen kurssikoe pidetään pe 11. 3. klo 9.15–11.45 Exactumin auditoriossa CK112. Kokeessa kuulustellaan monisteen jaksot 3.6–6.4 ja harjoitukset 7–13. ”Lunttilappu” sallitaan kokeessa; tarkempaa tietoa siitä kurssin kotisivulla. Kokeeseen liittyviä kysymyksiä voi esittää viimeisellä luennolla to 3. 3. Silloin luvassa myös pientä tarjoilua.