Tilastollinen päättely I, kevät 2017
Harjoitus 4 – palautuspäivämäärä 11. 4.

1. Kertaus uskottavuusfunktiosta. Arpaliput on numeroitu luvuin $1, 2, \ldots, \theta$, jossa $\theta \geq 1$ on kokonaisluku. Poimitaan yksi lippu umpimähkään. Olkoon Y siinä oleva luku.

a) Mikä on satunnaismuuttujan Y pistetodennäköisyysfunktio $f(y; \theta)$?

b) Oletetaan, että tuloksena saatiin $Y = 7$. Ilmoita tätä vastaava uskottavuusfunktio. Mikä on θ:n suurimman uskottavuuden estimaatti?

2. Kertaus normaalijakauman parametrien estimoinnista. Tutkimusryhmät A ja B ovat toisistaan riippumattomasti tutkineet ilmiötä, jota voidaan kuvata satunnaisotoksella normaalijakaumasta $N(\mu, \sigma^2)$, jossa molemmat parametrit ovat tuntemattomia. Kiinnostuksen kohteena on populaation keskiarvo μ. Tutkimusryhmien tulokset (otoskoko, otoskeskiarvo ja otosvarianssi) ovat vastaavasti

$$n_A = 10, \quad \bar{y}_A = 3.952, \quad s_A^2 = 1.981,$$
$$n_B = 15, \quad \bar{y}_B = 3.411, \quad s_B^2 = 1.321.$$

Tehtäväsi on harjoittaa meta-analyysiä eli yhdistää näiden kahden tutkimuksen tulokset. Laske parametrien μ ja σ^2 tavanomaiset piste-estimaatit käyttämällä hyväksi sekä ryhmän A että ryhmän B havainnot yhdessä.

Tehtävät 3–5 liittyvät luottamusväleihin (monisteen jaksot 5.1–5.6).

3. Erääseen ammattiryhmään kuuluvien miesten päivittäistä energiansaantia selvitettiin pienellä otantatutkimuksella, jossa seitsemän miehen päivittäiset energiansaannit (seurantajakson aikana) olivat 9730, 8240, 8060, 9100, 8880, 10070 ja 9810 kilojoulea päivässä.

a) Laske energiansaantimittausten keskiarvo $\bar{y}$ ja keskihajonta s.

b) Muodosta aineiston perusteella 95 %:n luottamusväli ammattiryhmään kuuluvien miesten keskimääräiselle päivittäiselle energiansaannille μ. Oletamme, että energiansaanneissa esiintyvä (satunnais)vaihtelu on normaalisti jakautunutta.

Vihje: Jos lasket keskihajonnan "käsin" laskimella, kaava $\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$ voi olla avuksi. Se seuraa harjoituksen 2 tehtävästä 5.

Tarvittava t-jakauman taulukko jaettiin harjoituksen 3 yhteydessä. Vaihtoehtoisesti käytä online-laskinta http://surfstat.anu.edu.au/surfstat-home/tables/t.php.

4. Mallimme on yksi havainto Y_1 jakaumasta $\text{Tas}(0, \theta)$, jossa $\theta > 0$ on tuntematon parametri. Käytännön esimerkkinä voisi olla harjoituksen 1 tehtävän 4a asetelma.

a) Päättele, että satunnaismuuttujan Y_1/θ jakauma ei riipu θ:n arvosta. Kyseessä on siis *saranasuure* parametrille θ (ks. monisteen jaksot 5.3 ja 5.5). Totea, että itse asiassa pätee

$$P_\theta(Y_1/\theta \leq q) = q, \quad \text{kun } 0 < q < 1,$$

eli Y_1/θ noudattaa tasajakaumaa $\text{Tas}(0, 1)$. Tässä P:n alaindeksi korostaa sitä, että ajattelemme Y_1:n noudattavan jakaumaa $\text{Tas}(0, \theta)$ parametriarvolla θ ja samaa lukua θ käytetään osamäärän Y_1/θ nimittäjässä.
b) Totea, että edellisen perusteella pätee $P_\theta(0.05 \leq Y_1/\theta \leq 1) = 0.95$ ja päättele tästä 95 %:n luottamusvälin lauseke θ:lle, kun on havaittu $Y_1 = y_1$. Minkä luottamusvälin saat harjoituksen 1 tehtävän 4a tilanteessa, jossa $y_1 = 4.5$?

Lisätehtävä innokkaille: Yleistä tarkastelu malliin, joka koostuu n riippumattomasta havainnosta jakaumasta $\text{Tas}(0, \theta)$ (harjoituksen 2 tehtävän 2 malli).

5. Opiskele, miten monisteen jaksossa 5.6.4 johdetaan mallissa $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$ $\perp\!\!\!\perp$ parametrille σ^2 luottamusväli muotoa $[c_1(\alpha, n)s^2,\ c_2(\alpha, n)s^2]$ (ks. kaava (5.18)).

Muokkaa tätä tarkastelua ja esitä, miten johdetaan σ^2:lle luottamusväli muotoa $]0,\ b(n, \alpha)s^2]$, kun luottamustaso on $1 - \alpha$. Miten kerroin $b(n, \alpha)$ on valittava? Oletamme, että meillä on käytössä χ^2-jakaumien taulukot tai tietokoneohjelma niiden kvantiilien laskemiseen.

Lukua $b(n, \alpha)s^2$ kutsutaan σ^2:n *ylemmäksi luottamusrajaksi*. Se on monissa sovelluksissa luontevampi tieto varianssista σ^2 kuin "kaksisuuntainen" luottamusväli: Kuvaahan σ^2 nimittäin havaintoihin liittyvän satunnaisvaihtelun suuruutta ja usein on tarpeen esittää arvioita siitä, kuinka suurta tämä satunnaisvaihtelu "pahimmillaan" voi olla. Sen pienuus sen sijaan harvoin on ongelma.

Vastaavalla tavalla voitaisiin määritellä *alempi luottamusraja*.

6. Uskottavuusvälit: vaihtoehtoinen menetelmä väliestimointiin. Olkoon $L(\theta; \mathbf{y})$ johonkin tilastolliseen malliin ja aineistoon $\mathbf{y}$ liittyvä uskottavuusfunktio sekä $\hat{\theta}$ parametrin θ suurimman uskottavuuden estimaatti. Jos $0 < c < 1$, sanotaan, että joukko

$$\{\theta : L(\theta; \mathbf{y}) \geq c \cdot L(\hat{\theta}; \mathbf{y})\} = \left\{\theta : \frac{L(\theta; \mathbf{y})}{L(\hat{\theta}; \mathbf{y})} \geq c\right\}$$

on $100c$ %:n *uskottavuusjoukko* parametrille θ. Jos θ on yksiulotteinen, kyseinen joukko on tavallisesti väli ja sitä kutsutaan *uskottavuusväliksi*. Esimerkiksi $c = 0.1$ vastaa 10 %:n uskottavuusjoukkoa.

Tarkastellaan binomikoetta (eli toistokoetta), jossa onnistumistodennäköisyys on tuntematon parametri $0 \leq \theta \leq 1$, toistojen lukumäärä on $n = 10$ ja on havaittu 4 onnistumista. Palauta mieleen tätä vastaava uskottavuusfunktio ja suurimman uskottavuuden estimaatti $\hat{\theta}$. Etsi sitten 50 %:n ja 10 %:n uskottavuusvälit θ:lle. Riittää antaa välien päätepisteet yhden desimaalin tarkkuudella. Halutessasi tarkastele asiaa graafisesti tai käytä apuna esim. R:ää.

Opetus. Uskottavuusvälit ja -joukot ovat hyvin luonteva "uskottavuuspohjainen" tapa suorittaa väliestimointia ja samalla täydentää suurimman uskottavuuden menetelmän antamaa piste-estimaattia. Tulkinnallisestikin ne ovat kenties helpommin ymmärrettäviä kuin luottamusvälit. Niitä kuitenkin käytetään (frekventistisessä) tilastotieteessä hyvin vähän luottamusväleihin verrattuna.