

Tilastollinen päättely I – kesätentti 15. 6. 2016 (3 h 30 min)

Huom. Kokeessa ei saa käyttää omia taulukoita, kaavakirjoja tai ”lunttilappuja”. Laskinta saa käyttää. Tarvittava taulukko on ohessa.

1. Tarkastellaan seuraavanlaista koeasetelmaa: Tiettyä koetta toistetaan riippumattomasti samanlaisissa olosuhteissa n kertaa. Kukin koe joko *onnistuu* tai *epäonnistuu*, ja onnistumisen todennäköisyys on tuntematon luku θ . Satunnaismuuttuja X kertoo onnistumisten lukumäärän.

a) Minkänimistä jakaumaa X noudattaa, ja mikä on sen pistetodennäköisyysfunktio?

b) Oletetaan, että toistoja tehtiin $n = 7$ ja niistä $X = 2$ onnistui. Muodosta tätä havaintoa vastaava uskottavuusfunktio ja määritä derivaattatarkastelun avulla huolellisesti perustellen suurimman uskottavuuden estimaatti θ :lle.

2. a) Tarkastellaan tilastollista mallia $f(\mathbf{y}; \theta)$, jonka parametri θ on yksiulotteinen eli reaalinen. Oletetaan, että aineiston \mathbf{y} perusteella on θ :lle muodostettu 95 %:n luottamusväli $[L, U]$, jossa L ja U ovat reaalilukuja. Selosta, mitä tämä tarkoittaa ja erityisesti miten luku 95 % eli 0.95 on tulkinnallisesti ymmärrettävä.

b) Olkoon Y_1, \dots, Y_n satunnaisotos normaalijakaumasta $N(\mu, 1)$. Kuinka suuri otoskoon n on oltava, jotta havaintojen y_1, \dots, y_n perusteella muodostettava tavanomainen 95 %:n luottamusväli μ :lle (ns. z -luottamusväli) olisi pituudeltaan (noin) 0.2 eli muotoa $[\bar{y} - 0.1, \bar{y} + 0.1]$, jossa $\bar{y} = (y_1 + \dots + y_n)/n$ on otoskeskiarvo?

3. Puutavaraerän laatuvaatimukseen kuuluu, että puun keskimääräinen kosteusprosentti saa olla korkeintaan 20. Laadun varmistamiseksi tehtiin kuusi riippumatonta kosteusprosentin mittausta, joiden tulokset olivat 21, 18, 24, 23, 20 ja 26. Laske mittaustulosten otoskeskiarvo ja otoskeskihajonta sekä testaa merkitsevyydellä 0.05, voidaanko päätellä, että puutavara ei täytä laatuvaatimusta. Muista lausua testattavat hypoteesit selvästi. Oletetaan, että kosteusprosentin vaihtelut ovat (likimain) normaalisti jakautuneita.

4. a) Bayes-päätelyssä tärkeille funktioille on käytetty kurssilla merkintöjä $p(\theta)$, $p(\theta|\mathbf{y})$ ja $f(\mathbf{y}|\theta)$. Mitä nimityksiä niistä käytetään ja mitä ne kuvaavat? Millainen keskeinen kaava sitoo näitä funktioita toisiinsa?

b) Mainitse pari keskeistä Bayes-päätelyn ja ns. frekventistisen päätelyn välistä eroa.