

Johdatus tilastolliseen päättelyyn – kurssikoe 5. 5. 2014

Huom. Kokeessa saa käyttää laskinta ja kaksipuolista A4-kokoista käsinkirjoitettua "lunttilappua". Omia taulukoita ja kaavakirjoja ei saa käyttää.

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) Minkä nimistä jakaumaa X noudattaa, ja mikä on sen pistetodennäköisyysfunktio?
 - b) Oletetaan, että toistoja tehtiin $n = 7$ ja niistä $X = 4$ onnistui. Muodosta tätä havaintoa vastaava uskottavuusfunktio ja määritä suurimman uskottavuuden estimaatti θ :lle.
2. Puutavaraerästä tehtiin 9 riippumatonta kosteusmittausta. Mittausten keskiarvoksi saatiin 17.0 ja keskihajonnaksi 3.6 (prosenttia). Oletetaan, että kosteuspitoisuuden vaihtelut ovat likimain normaalisti jakautuneita.
 - a) Muodosta 95 %:n luottamusväli puutavaraerän keskimääräiselle kosteuspitoisuudelle. Tarvittava taulukko on liitteenä.
 - b) Onko seuraava luonnehdinta oikein vai väärin: 95 %:n luottamusväli tarkoittaa sitä, että jos kosteusmittauksia tehtäisiin suuri määrä, niin noin 95 % mittauksista sijoittuisi ko. välille? Perustele lyhyesti.
3. Olkoon Y_1, \dots, Y_n satunnaisotos normaalijakaumasta $N(\mu, \sigma^2)$, jossa sekä odotusarvo μ että varianssi σ^2 ovat tuntemattomia parametreja. Parametria σ^2 on tapana estimoida havainnoista y_1, \dots, y_n laskettavalla otosvarianssilla s^2 .
 - a) Millä kaavalla s^2 lasketaan havainnoista?
 - b) Vastaavalle satunnaismuuttujalle S^2 pätee jakaumatulos $\frac{n-1}{\sigma^2} S^2 \sim \chi_{n-1}^2$ (khii-toiseen-jakauma, $n - 1$ vapausastetta). Johda tämän avulla testi nollahypoteesille $H_0: \sigma = \sigma_0$, kun vastahypoteesi on yksisuuntainen $H_1: \sigma > \sigma_0$. Miten p -arvo lasketaan? Tässä n ja $\sigma_0 > 0$ ovat tunnettuja lukuja.

Tarvittavia taulukoita ei ole liitteenä, vaan sinun tulee kertoa, minkä funktion arvojen avulla p -arvo laskettaisiin (taulukoita tai tietokonetta käyttämällä).
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) Mikä seikka voidaan nähdä ongelmana Bayes-päätelyssä? Mitä etuja siinä toisaalta on ns. frekventistiseen päättelyyn verrattuna? Pari kolme virkettä riittää vastaukseksi.