# MS-A0501 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

Tentti ja välikokeet 11.11.2014 / Liesiö

Kirjoita selvästi *jokaiseen koepaperiin* seuraavat tiedot:

- MS-A0501 + Tentti/1.vk/2.vk **(merkitse oikea vaihtoehto)** 11.11.2014
- opiskelijanumero + kirjain
- TEKSTATEN sukunimi ja kaikki etunimet
- koulutusohjelma ja vuosikurssi
- nimikirjoitus

**Lue *tarkasti* seuraavat ohjeet:**

(1) **Jos olet *suorittamassa kurssia tentillä*, vastaa kysymyksiin 2, 4, 5, 7 ja 8 (merkitty *).**

(2) **Jos olet *suorittamassa* 1. *välikoetta ensimmäisen kerran* ja *saanut siihen erikoisluvan*, vastaa kysymyksiin 1-4.**

(3) **Jos olet *suorittamassa* 2. *välikoetta ensimmäisen kerran* ja *saanut siihen erikoisluvan*, vastaa kysymyksiin 5-8.**

***Merkitse vastauspapereihisi selvästi mitä vaihtoehtoa olet suorittamassa!***

**Sallitut apuvälineet:** **Laskin**, **Tilastolliset taulukot** (Ilkka Mellinin) ja **muistiinpanolappu** (käsinkirjoitettu A4 kokoinen paperi, tekstiä vain toisella puolella ja oikeassa yläkulmassa opiskelijan nimi)

**Vastausohje:** **Vastaa *lyhyesti* ja *ytimekkäästi*, mutta *perustele ratkaisusi*. Esimerkiksi pelkkä lukuarvo tai numeerinen laskutoimitus ilman kaavoja *ei anna* pisteitä.**

**1.** Eräässä maassa 10 % sen asukkaista kantaa tietämättään erästä bakteeria, joka tietyissä olosuhteissa saattaa aiheuttaa erään taudin puhkeamisen. Pikatesti paljastaa bakteerin kantajista 95 %, mutta toisaalta se antaa virheellisen positiivisen tuloksen 5 %:lle niistä, jotka eivät kanna ko. bakteeria.

(a) Mikä on todennäköisyys että satunnaisesti valitulle henkilölle testi antaa positiivisen tuloksen?

(b) Mikä on todennäköisyys, että positiivisen testituloksen saanut henkilö, todella myös kantaa ko. bakteeria?

**2.*** Uurnassa on 5 punaista ja 7 sinistä palloa.

(a) Poimit uurnasta satunnaisesti kolme palloa *takaisinpanolla*. Mikä on todennäköisyys, että saat kolme punaista palloa?

(b) Poimit uurnasta satunnaisesti kolme palloa *ilman takaisinpanoa*. Mikä on todennäköisyys, että saat kolme punaista palloa?

(c) Poimit uurnasta satunnaisesti kolme palloa *ilman takaisinpanoa*. Mikä on todennäköisyys, että viimeisenä poimittava pallo on punainen, jos kaksi edellistä ovat olleet sinisiä?

(d) Poimit uurnasta satunnaisesti palloja *takaisinpanolla* kunnes tulee ensimmäinen punainen pallo. Mikä on poimittujen pallojen lukumäärän odotusarvo?

**3.** Satunnaismuuttujan $X$ tiheysfunktio on muotoa

$$f(x)=\begin{cases} x+b\,, \text{ kun } 0\le x\le 1 \\ 0\,, \text{ muulloin}\end{cases}$$

(a) Määrää vakion $b$ arvo.

(b) Määrää tapahtuman $0 \le X \le 0.5$ todennäköisyys.

(c) Määrää satunnaismuuttujan $X$ kertymäfunktio.

**4.*** Alla oleva taulukko esittää diskreettien satunnaismuuttujien $X$ ja $Y$ yhteisjakauman pistetodennäköisyysfunktiota, $f_{XY}(x, y) = \Pr(X = x \text{ ja } Y = y)$.

(a) Määrää $X$:n ja $Y$:n reunajakaumien pistetodennäköisyysfunktiot.

(b) Laske $X$:n ja $Y$:n kovarianssi.

(c) Määrää ehdollinen jakauma $f_{X|Y}(x|1)$ ja ehdollinen odotusarvo $E(X|Y=-1)$.

(d) Ovatko $X$ ja $Y$ riippumattomia?

| $f_{XY}(x,y)$ | | $x$ | | |
|---|---|---|---|---|
| | | −1 | 0 | +1 |
| $y$ | +1 | 0 | 3/8 | 0 |
| | 0 | 3/8 | 0 | 1/8 |
| | −1 | 0 | 1/8 | 0 |

**5.*** (a) Mitä tarkoitetaan, kun sanotaan, että havainnot

$$X_1, X_2, \dots, X_n$$

muodostavat (yksinkertaisen) **satunnaisotoksen** jakaumasta, jonka pistetodennäköisyys- tai tiheysfunktio on $f(x)$?

(b) Oletetaan, että $X_1, X_2, \dots, X_n$ on (yksinkertainen) satunnaisotos Bernoulli(p) jakaumasta. Mikä on parametrin $p$ **SU-estimaattori** (estimaattoria ei tarvitse johtaa)? Miten se lasketaan havainnoista $X_1, X_2, \dots, X_n$?

(c) Oletetaan, että $X_1, X_2, \dots, X_n$ on (yksinkertaisen) satunnaisotos $N(\mu,\sigma^2)$ jakaumasta. Osoita, että aritmeettinen keskiarvo $Y=(X_1+X_{2+} \dots + X_n)/n$ on parametrin $\mu$ **harhaton estimaattori**.

**6.** Kaverisi on kehittänyt algoritmin, joka generoi satunnaislukuja normaalijakaumasta $N(\mu,\sigma^2)$. Hän toimittaa sinulle sähköpostitse 30 lukua sisältävän satunnaisotoksen. Lukujen otoskeskiarvo on 10.09 ja otosvarianssi 0.0108.

(a) Määrää 95 %:n luottamusväli odotusarvolle $\mu$.

(b) Kaverisi kertoo, että hän käytti varianssia $\sigma^2$=0.01 lukujen generointiin. Määrää Määrää 95 %:n luottamusväli odotusarvolle $\mu$.

**7*.** Eräässä kokeessa verrattiin kahta sademäärän mittaukseen käytettävää laitetta. Kummallakin laitteella mitattiin sademäärät *samalla paikalla* 6 sadepäivän aikana. Mittaustulokset (sademäärät mm:nä) on annettu alla olevassa taulukossa.

(a) Halutaan testata, tuottavatko mittarit keskimäärin samoja mittaustuloksia. Selitä lyhyesti miksi tähän ei voida käyttää kahden riippumattoman otoksen t-testiä. (1p)

(b) Valitse sopiva tilastollinen testi ja testaa hypoteesia, että mittarit tuottavat keskimäärin samoja mittaustuloksia, kun vaihtoehtoisena hypoteesina on, että mittarit tuottavat keskimäärin eri mittaustuloksia. Käytä testissä 1 %:n merkitsevyystasoa.(5p)

| **Laite** | **1** | **2** | **3** | **4** | **5** | **6** |
|---|---|---|---|---|---|---|
| **A** | 1.3 | 9.6 | 0.3 | 1.4 | 5.9 | 0.5 |
| **B** | 1.4 | 10.3 | 0.3 | 1.5 | 6.1 | 0.6 |

**8.*** Kokeessa tutkittiin seitsemän kuorma-auton polttoainetaloudellisuus (muuttuja $y$, yksikkönä *mi/gal*, mailia per gallona) riippuvuutta ajoneuvon painosta (muuttuja $x$, yksikkönä *ton*, tonni). Kokeesta saadut tiedot on annettu alla olevassa taulukossa.

| $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $x_i$ | 8.00 | 24.50 | 27.00 | 14.50 | 28.50 | 12.75 | 21.25 |
| $y_i$ | 7.69 | 4.97 | 4.56 | 6.49 | 4.34 | 6.24 | 4.45 |

Aineistoa kuvaavat otostunnusluvut ovat:

$\bar{x} = 19.5$ $\bar{y} = 5.5343$

$s_x^2 = 61.354$ $s_y^2 = 1.655$

$s_x = 7.833$ $s_y = 1.286$

$s_{xy} = -9.518$ $r_{xy} = -0.945$

(a) Estimoi havaintojen perusteella yhden selittäjän lineaarisen regressiomallin regressiokertoimien $\beta_0$ ja $\beta_1$ pienimmän neliösumman (PNS-) estimaatit. (2p)

(b) Ennusta estimoidulla mallilla 40 tonnia painavan kuorma-auton polttoainetaloudellisuus. (1p)

(c) Ennusta estimoidulla mallilla 10 tonnia painavan kuorma-auton polttoainetaloudellisuus. (1p)

(d) Perustele kummalla kohtien (b) ja (c) ennusteista olisi kapeampi luottamusväli. (2p)