

MS-A0501 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

2. välikoe 20.10.2014 / Liesiö

Kirjoita selvästi *jokaiseen koepaperiin* seuraavat tiedot:

- MS-A0501 2. vk 20.10.2014
- opiskelijanumero + kirjain
- TEKSTATEN sukunimi ja kaikki etunimet
- koulutusohjelma ja vuosikurssi
- nimikirjoitus

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 *perustelee ratkaisusi*. Pelkkä lukuarvo vastauksena *ei anna* pisteitä.

1. (a) Mitä tarkoitetaan, kun sanotaan, että havainnot

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

muodostavat (yksinkertaisen) **satunnaisotoksen** jakaumasta, jonka pistetodennäköisyys- tai tiheysfunktio on $f(x)$? (2p)

- (b) Oletetaan, että X_1, X_2, \dots, X_n on (yksinkertainen) satunnaisotos $N(\mu, \sigma^2)$ jakaumasta. Mikä on odotusarvoparametrin μ **SU-estimaattori** (estimaattoria ei tarvitse johtaa)? Miten se lasketaan havainnoista X_1, X_2, \dots, X_n ? (2p)
- (c) Oletetaan, että X_1, X_2, \dots, X_n on (yksinkertaisen) satunnaisotos Bernoulli(p) jakaumasta. Osoita että suhteellinen frekvenssi $Y=(X_1+X_2+\dots+X_n)/n$ on parametrin p **harhaton estimaattori**. (2p)

2. Tehdas valmistaa suklaapatukoita. Patukoiden pituus vaihtelee satunnaisesti noudattaen normaalijakaumaa. Patukoiden joukosta poimittiin yksinkertainen satunnaisotos, jonka koko oli 30. Otoskeskiarvoksi saatiin 9.99 cm ja otosvarianssiksi 0.01 cm^2 . Määrää 90 %:n luottamusväli patukoiden pituuden varianssille.

3. Testattaessa erästä verenpaineläkettä samojen potilaiden (8 kpl) verenpaine mitattiin **ennen** ja **jälkeen** lääkkeen nauttimisen. Koetulokset (verenpaineet mm/Hg) on esitetty alla olevassa taulukossa.

Taulukko 1: Havaintoaineisto sekä (*) verenpaineiden erotus ja () erotuksen neliöllinen poikkeama erotusten keskiarvosta.**

	1	2	3	4	5	6	7	8
Jälkeen	128	176	110	149	183	136	118	158
Ennen	134	174	118	152	187	136	125	168
*	6	-2	8	3	4	0	7	10
**	2.25	42.25	12.25	2.25	0.25	20.25	6.25	30.25

- (a) Halutaan testata, alentaako lääke keskimäärin verenpainetta. Selitä lyhyesti miksi tähän ei voida käyttää kahden riippumattoman otoksen t-testiä. (1p)
- (b) Valitse sopiva tilastollinen testi ja testaa nollahypoteesia, että lääke ei keskimäärin muuta verenpainetta, kun vaihtoehtoisena hypoteesina on, että lääke keskimäärin alentaa verenpainetta. Käytä testissä 1 %:n merkitsevyystasoa. (Voit olettaa erotusten (*) olevan riippumattomia ja normaalijakautuneita.) (5p)
4. 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:

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

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

$$s_x = 7.833 \qquad s_y = 1.286$$

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

- (a) Estimoi havaintojen perusteella yhden selittäjän lineaarisen regressiomallin regressiokertoimien β_0 ja β_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)