M… …uanina

MS-A0501 Todennäköisyyslaskennan ja tilastotieteen peruskurssi — Välikoe 2
Matematiikan ja systeemianalyysin laitos — 19.10.2015
Aalto-yliopisto — H Seppälä

Kirjoita selvästi jokaiseen vastauspaperiin:

- Kurssin koodi, päivämäärä, kokeen tyyppi (Tentti)
- Opiskelijanumero + kirjain
- TEKSTATEN sukunimi ja kaikki etunimet
- Koulutusohjelma ja vuosikurssi
- Nimikirjoitus

Vastausohje: Vastaa lyhyesti ja ytimekkäästi, mutta perustele ratkaisusi (paitsi tehtävässä 1). Pelkkä lukuarvo vastauksena ei anna pisteitä. Kokeessa on 4 tehtävää, jokaisesta saa 0–6 pistettä.

Sallitut apuvälineet: Laskin, Mellinin tilastolliset taulukot, ja a4-muistilappu (käsinkirjoitettu, tekstiä vain toisella puolella, oikeassa yläkulmassa opiskelijan nimi)

---

1. Ovatko seuraavat väittämät totta? Vastaa **1** = Totta, **x** = Vastausta ei tunneta annettujen tietojen pohjalta, tai **2** = Tarua. (1 p/kohta)

   (a) Havaitun aineiston $x = (x_1, ..., x_2)$ mediaani on suurempi kuin alakvartiili.

   (b) Tilastokokeen mallinnuksessa käytettävän satunnaismuuttujan X odotusarvon μ ja siihen liittyvän otoskeskiarvon $m(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$ varianssit ovat yhtä suuret.

   (c) Jos pohjaoletukset pitävät paikkansa, niin kiinnostuksen kohteena olevan parametrin luottamusväli on satunnainen väli, joka peittää kyseisen parametrin (tuntemattoman) todellisen arvon valitulla todennäköisyydellä.

   (d) Testin p-arvo tarkoittaa todennäköisyyttä, jolla nollahypoteesi on tosi.

   (e) Oletetaan, että aineistosta laskettu 99%-luottamusväli eräälle parametrille θ on $(3.14, 4.27)$, missä parametrin θ estimaattori on normaalijakautunut. Silloin nollahypoteesi $\theta = 3$ jää voimaan, kun käytetään 1% merkitsevyystasoa.

   (f) Tarkastellaan kaksiulotteista satunnaisilmiötä, jota kuvataan satunnaisvektorilla (X, Y). Oletetaan, että olemme keränneet 100 riippumatonta havaintoa $(x_1, y_1), ..., (x_{100}, y_{100})$. Jos havainnoista koostuvaan aineistoon sovitettu regressiosuora on vaakasuora, niin satunnaisvektorin (X, Y) komponentteja X ja Y voidaan pitää (lähes) riippumattomina.

   Vihje: Ole erityisen huolellinen.

2. Tutkitaan suomalaisten halukkuutta liittyä NATO:on. Poimitaan Suomen väestöstä yksinkertainen satunnaisotos, jonka koko on 3000. Otokseen valituista henkilöista 1273 kannattaa liittymistä ja loput vastustaa.

   (a) Määritä satunnaismuuttuja yksittäisen henkilön mielipiteen mallintamiseksi, kun oletetaan, että kaikki ovat joko puolesta tai vastaan.

   (b) Jos tehdään useampia yhtä laajoja kyselyitä, niin miksi otoksista laskettujen NATO:a kannattavien suhteellisten osuuksien jakaumaa voidaan arvioida normaalijakaumalla? (Lyhyt, 1-2 rivin sanallinen selitys riittää.) (1 p)

   (c) Laske otokseen perustuva 95% luottamusväli NATO:a kannattavien suhteelliselle osuudelle. (3 p)

(b) Testaa hypoteesia H_0: "uusi hoitomuoto on yhtä toimiva kuin vakiintunut hoitomuoto", kun vaihtoehtoisena hypoteesina on H_1: "uuden ja vakiintuneen hoitomuodon toimivuudessa on eroa". Laske p-arvo, ja tee johtopäätökset käyttäen merkitsevyystasoa $\alpha = .01$. (3 p)

3. Kannettavan laseretäisyysmittarin tarkkuutta halutaan arvioida. Tehdään testimittauksia, joissa 500 metrin etäisyydellä olevasta kohteesta saadaan seuraavat mittausvirheet (metreissä)

testimittaus j	1	2	3	4	5	6	7	8
mittausvirhe v_j	-0.10	-0.02	0.10	0.03	0.09	0.01	-0.05	0.05

testimittaus j	9	10	11	12	13	14	15
mittausvirhe v_j	-0.06	0.01	0.03	0.06	0.02	-0.07	0.03

Aputuloksia: $\sum_{j=1}^{15} v_j = 0.13, \quad \sum_{j=1}^{15} v_j^2 = 0.0489.$

(a) Arvioi mittausvirheiden tyypillistä suuruusluokkaa (metreissä) sopivasti valitsemallasi aineiston tunnusluvulla. (2 p)

(b) Mitä pitäisi olettaa, että mittausvirheiden suuruuden arvioimiseksi olisi mielekästä määrittää χ^2-jakaumaan perustuva luottamusväli testimittausten virheiden varianssille? (2 p)

(c) Estimoi aineistosta mittausvirheen suuruutta antamalla 90% luottamusväli varianssille, kun (b)-kohdan oletuksia pidetään perusteltuina. Ilmoita vastaava väli myös keskihajonnalle. (2 p)

4. Monien kemikaalien liukeneminen veteen riippuu veden lämpötilasta. Ilmiötä tutkittiin eräällä kemikaalilla, ja saatiin alla olevat mittaustulokset. Syötemuuttujana on lämpötila (muuttuja x, yksikkönä °C) ja vastemuuttujana liukoisuus (muuttuja y, yksikköinä g/l).

mittaus j	1	2	3	4	5	6
lämpötila x_j	0	10	20	30	40	50
liukoisuus y_j	3.1	4.5	8.2	12.5	15.9	18.8

Aputuloksia: $\sum_{j=1}^{6} x_j = 150, \quad \sum_{j=1}^{6} x_j^2 = 5500, \quad \sum_{j=1}^{6} y_j = 63, \quad \sum_{j=1}^{6} y_j^2 = 859.6, \quad \sum_{j=1}^{6} x_j y_j = 2160.$

(a) Estimoi havaintojen perusteella yhden selittäjän lineaarisen regressiomallin $Y_j = \beta_0 + \beta_1 x_j + \epsilon_j$ regressiokertoimien β_0 ja β_1 pienimmän neliösumman estimaatit. (2 p)

(b) Ennusta (a)-kohdassa estimoidulla mallilla liukoisuus vedessä, jonka lämpötila on 15°C, ja vastaavasti myös vedessä, jonka lämpötila on 60°C. (2 p)

(c) Perustele kummalla kohdan (b) ennusteista olisi kapeampi ennusteväli. (2 p)