

2. välikokeen uusintakoe 20.5.2016

Tehtävät ovat samanarvoisia (6 p) paitsi vapaaehtoinen 5. tehtävä (1 p). Kaavojen käyttö ja tarkka esittäminen on välttämätöntä. Perustele huolellisesti kaikki esittämäsi yhtäsuuruudet, laskusi, välituloksesi, käyttämäsi jakaumien oikeutus, päätelmäsi ja niiden perustelut. Pelkkä oikea lopputulos ei ole riittävä vastaus. Palauta koetta varten tekemäsi muistilappu vastauskonseptisi välissä.

Viimeinen esimerkki. Elämänfilosofiaa. Oxfordin yliopisto koostuu 38 college'ista. Somerville College on nimetty skotlantilaisen matemaatikko Mary Somervillen (1780–1872) mukaan. Ensi vuodesta lähtien Royal Bank of Scotlannin 10 punnan setelin kuva-aihe on Mary Somerville. Hän paljasti 91-vuotiaana yhden menestymisensä selityksen:¹

Jollen onnistu tänään, pureudun ongelmaan uudestaan huomenna.

Paina Somervillen ohje mieleesi. Menestystä opinnoillesi! Hyvää kesää! □

1. Apua lapsiperheille (*The Aid to Families with Dependent Children*) -ohjelmalla jaetaan Yhdysvalloissa rahallista tukea apua tarvitseville lapsiperheille. Joskus tuki annetaan erehdyksessä perheille, jotka eivät ole siihen oikeutettuja. Kalifornian osavaltio velvoittaa piirikuntansa valvomaan, että tuki menee vain sitä tarvitseville. Tukea saa ohjautua väärin käsiin 4 prosentissa tukipäätöksistä. Ajatus on, että erehdyksiä tai väärinkäytöksiä tapahtuu väistämättä, mutta ne eivät saa olla liian yleisiä. Osavaltio ei pysty tutkimaan jokaista apupäätöstä, joten se käyttää otantaa 4 prosentin -sääntönsä noudattamisen valvomiseen.

Vuonna 1981 Alamedan piirikunnassa Kaliforniassa tutkittiin 152 tukipäätöstä, joista 9 eli 5.9 prosenttia todettiin perusteettomiksi. Osuus ylitti 4 prosentin rajan, joten osavaltio määräsi piirikunnalle sakon. Piirikunta vei asian oikeuteen argumentoiden, että 9 virheellistä päätöstä 152 päätöksestä ei ole riittävä todiste, että 4 prosentin raja oltiin ylitetty kaikissa päätöksissä. Oikeus päätti, että piirikunta on oikeassa ja kumosi osavaltion määräämän sakon. Oikeus perusti päätöksensä tilastotieteelliseen argumenttiin.²

a) Laske 95 %:n kaksisuuntainen luottamusväli perusteettomien tukipäätösten osuudelle.

b) Laske 95 %:n yksisuuntainen luottamusväli perusteettomien tukipäätösten osuudelle. (Laske suuntaan, joka on mielekäs arvioitaessa, voisiko todellinen osuus olla 4 %.)

c) Testaa kaksisuuntaisella testillä 5 %:n riskitasolla nollahypoteesia, että perusteettomien tukipäätösten osuus on 4 prosenttia.

d) Testaa yksisuuntaisella testillä 5 %:n riskitasolla nollahypoteesia, että perusteettomien tukipäätösten osuus on 4 prosenttia.

e) Summeeraa tulokset ja selitä, mikä oikeuden tilastotieteellinen argumentti saattoi olla ja oliko päätös tilastotieteellisesti perusteltu vai ei.

¹<http://blog.oup.com/2016/03/mary-somerville-royal-bank-of-scotland/> (viitattu 3.5.2016).

²Tehtävän lähde on S. Ross (2010): *Introductory Statistics*, 3. laitos. Academic Press. (S:t 373–374.) Suomessa tilastotiedettä ei hyödynnetä oikeuslaitoksessa Yhdysvaltain tapaan.

2. Tutkitaan, vaikuttaako melu älylliseen suorituskyykyyn.³ Koehenkilöt ratkoivat aritmeettisia laskutoimituksia meluttomassa ja meluisassa ympäristössä. Heidän tekemien virheiden lukumäärien keskiarvot- ja otoskeskihajonnat ovat taulukossa (n on otoskoko).

		virheiden lukumäärän	
	n	keskiarvo	otoskeskihajonta
meluton ympäristö	16	14.4	2.28
meluisa ympäristö	16	16.5	2.45

Oletetaan, että virheiden lukumäärä molemmissa tilanteissa on normaalijakautunut ja että virheiden lukumäärän varianssit voivat poiketa ympäristöissä. Poikkeavatko virheiden lukumäärien odotusarvot ympäristöissä? Testaa odotusarvojen yhtäsuuruutta (Welchin eli Satterthwaiten) t-testillä. Tee testi kaksisuuntaisesti 5 %:n riskitasolla. Testissä tarvittava kaava vapausasteiden laskemiseksi on alla.

$$\nu = \text{int} \left[\frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{\left(\frac{s_1^2}{n_1} \right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2} \right)^2}{n_2 - 1}} \right].$$

3. Gallup-yhtiö, Purduen yliopisto ja Lumina-säätiö tutkivat, mikä yliopisto-opiskelussa (*college*) vaikuttavaa opiskelijoiden hyvinvointiin tutkinnon suorittamisen jälkeen.⁴ Helmi-maaliskuussa 2014 haastateltiin 29 560 alumnia Internetissä. Heidän hyvinvointiaan mitattiin viidellä ulottuvuudella: Tekemisen merkityksellisyys sekä sosiaalinen, taloudellinen, yhteisöllinen ja fyysinen hyvinvointi.

Taulukossa on pyöristetyt osuudet erilaisista yliopistoista valmistuneista alumneista, jotka menestyivät erityisen hyvin (*thrived*) kaikilla viidellä hyvinvoinnin ulottuvuudella. Ryhmän ”100 parasta” yliopistot on valinnut *U.S. News & World Report*. Tämän ryhmän yliopistot kuuluvat myös taulukon muihin ryhmiin (julkinen jne.).⁵

³Tehtävä juontaa esimerkistä opetusmonisteesta I. Mellin (1996): Johdatus tilastotieteen. 2. kirja. Helsingin yliopisto. Tilastotieteen laitos. (S. 156.)

⁴<http://www.gallup.com/poll/168848/life-college-matters-life-college.aspx> (julkaistu 6.5.2014, viitattu 7.5.2016), <http://www.gallup.com/poll/182306/big-six-college-experiences-linked-life-preparedness.aspx> (julkaistu 8.4.2015; viitattu 7.6.2016) ja The 2014 Gallup-Purdue Index Report osoitteessa http://www.gallup.com/file/services/176768/GallupPurdueIndex_Report_2014.pdf (julkaistu 22.9.2014; viitattu 7.6.2016).

⁵Lähteistä selviävät otoksen koko 29 560 sekä taulukon prosentiosuudet. Alumnielämäerät erilaisissa yliopistoissa eivät selviä lähteistä. Ne on keksitty taulukkoon.

yliopisto	lkm	osuus (%)
julkinen	14000	11
yksityinen voittoa tavoittelematon	14000	11
yksityinen voittoa tavoitteleva	1560	4
kaikki	29560	11
100 parasta	500	12

Opiskeluajan ”kuusi suurta kokemusta” olivat:

- Minulla oli ainakin yksi opettaja (*professor*), joka sai minut kiinnostumaan opiskelusta.
- Opettajani yliopistossa välittivät minusta ihmisenä.
- Minulla oli ohjaaja (*mentor*), joka kannusti minua tavoitteisiini ja unelmiini.
- Tein projektin, joka kesti lukukauden tai pidempään.
- Minulla oli työharjoittelupaikka tai työ, jossa pääsin soveltamaan oppimaani.
- Olin erittäin aktiivinen opiskelun ulkopuolisessa elämässä.

Alla on taulukoitu osuuksia alumneista, jotka pärjäisivät erityisen hyvin kaikilla hyvinvoinnin ulottuvuuksilla sen mukaan, kuinka moneen kokemusten lukumäärään (lkm) he vastasivat ”erittäin samaa mieltä”.⁶ Esimerkiksi 3991 alumnia oli ”erittäin samaa mieltä” kolmen kokemuksen kanssa eli oli kokenut kolme ”kuudesta suuresta kokemuksesta”.

kokemusten lkm	erittäin samaa mieltä (%)	erittäin samaa mieltä (lkm)	edellisestä sarakkeesta erityisen hyvin menestyvät (%)
0	24.5	7242	7.000829
1	23.0	6799	9.501397
2	20.0	5912	11.50203
3	13.5	3991	11.50088
4	10.0	2956	12.99053
5	5.5	1626	16.97417
6	3.0	887	17.02368

a) Miten alumnien hyvinvointi vaikuttaa riippuvan yliopistosta ja ”kuudesta suuresta kokemuksesta”?

⁶Taulukon osuudet ja lukumäärät on arvioitu kuviosta sivulla <http://www.gallup.com/opinion/gallup/182312/college-worth-depends.aspx> (viitattu 7.5.2016). Arvioinnin takia lukumäärät eivät aivan summaudu haastateltujen lukumääräksi 29 560.

b) Voitko annetuilla tiedoilla testata nollahypoteesia, että erityisen hyvin menestyvien alumnien osuus on sama ensimmäisen taulukon neljässä ryhmässä ”julkinen”, ”yksityinen voittoa tavoittelematon”, ”yksityinen voittoa tavoitteleva” ja ”100 parasta”. Jos voit, tee testi 0.1 %:n riskitasolla.

c) Voitko annetuilla tiedoilla testata nollahypoteesia, että erityisen hyvin menestyvien alumnien osuus ei riipu siitä, kuinka monen ”kuuden suuren kokeumuksen” kanssa alumni on ”erittäin samaa mieltä”? Jos voit, tee testi 0.1 %:n riskitasolla.

4. Keinänen ja Tukiainen (2010)⁷ estimoivat pienimmän neliösumman (PNS) menetelmällä yhtälön

$$y = 2441.8 + 40.3x + \hat{\varepsilon},$$

$$R^2 = 0.182, n = 258.$$

Yllä y on oikeuden ammattiyhdistykselle määräämän sakon suuruus euroina (vuoden 2009 rahanarvolla mitattuna), x on laittomasta lakosta aiheutunut taloudellinen menetys yritykselle sadoissa tuhansissa euroissa⁸ (vuoden 2009 rahanarvolla mitattuna), $\hat{\varepsilon}$ on mallin residuaali ja n on havaintojen lukumäärä. Jäännösten ε oletetaan noudattavan normaalijakaumaa $N(0, \sigma^2)$ ja olevan keskenään korreloimattomia. Havainnot ja niihin PNS-menetelmällä sovitettu regressiosuora ovat oheisessa kuviossa (mt:sta).

a) Tulkitse malli huolellisesti (vakio, miten x vaikuttaa y :hyn, mikä on R^2).

b) Onko taloudellinen menetys (x) tilastollisesti merkitsevä selittäjä 1 %:n riskitasolla?

c) Kuinka suuri on laittomista lakoista aiheutuneiden taloudellisten menestysten ja sakkojen korrelaatio?

5*. Ylimääräinen pisteen arvoinen lisäkysymys. Yhteiskuntatieteellisen tietokiriston Kvantitatiivisten menetelmien tietovarannossa neuvotaan χ^2 -riippumattomuustestistä:⁹

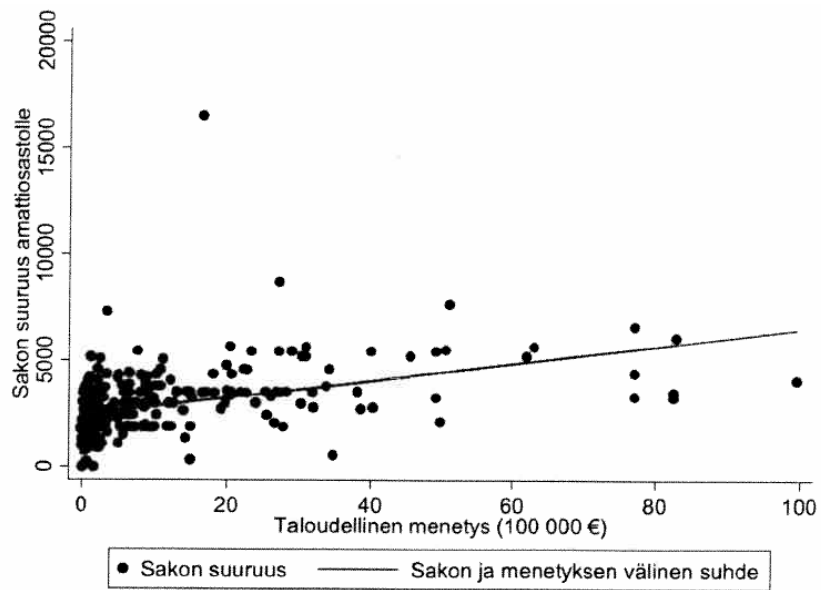
Testi kertoo vain kuinka todennäköistä on, että otoksessa havaitut erot ovat olemassa myös perusjoukossa.

Arvioi opastuksen oikeutta.

⁷A. Keinänen ja T. Tukiainen (2010): Laittomista työtaistelutoimenpiteistä tuomittujen hyvityssakkojen määräytyminen työtuoimioistuinikäytännössä. Edilex 2010/15 (<http://www.edilex.fi/lakikirjasto/asiantuntijakirjoitukset/7008>; viitattu 7.5.2016). Aineisto on työtuoimioistuimen oikeustapauksia 2000–2009. Alkuperäinen aineisto koostui 796 havainnosta. Puuttuvien tietojen ja yhden tuomion poikkeuksellisuuden takia Keinänen ja Tukiainen supistivat aineiston 258 havaintoon. Keinänen ja Tukiainen estimoivat useampia selittäviä muuttujia sisältäviä malleja. He pitävät niitä tehtävän mallia parempina.

⁸Esimerkiksi jos $x = 30$, niin yrityksen taloudellinen menetys on $30 \times 100\,000 = 3\,000\,000$ euroa.

⁹<http://www.fsd.uta.fi/menetelmaopetus/ristiintaulukointi/ristiintaulukointi.html> (viitattu 7.6.2016).



Kuvio 4. Työtaistelutoimenpiteestä aiheutuneet taloudelliset menetykset (100 000 euroa) ja ammattiosastolle tuomittu hyvityssakko⁶⁶.