

TILASTOTIETEEN JATKOKURSSI. 5+5 OP. 21.1.–6.5.2016. Luennoi yliopistonlehtori Pekka Pere.

2. välikoe 6.5.2016

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

Toivottavasti annat kurssipalautetta täällä:

<https://elomake.helsinki.fi/lomakkeet/70467/lomake.html>

Lomake on auki 9.5. (ma) klo 23:59 asti. Palautteesi on hyvin tärkeää. Ota koepaperi mukaasi, niin sinun ei tarvitse kopioida linkkiä yllä kurssipalautteen antamista varten.

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! □

1. Nisén ym. (2014) tutkivat syntyvyyttä (lasten lukumäärää) eri koulutustasoilla vuosina 1940–1950 syntyneiden naisten joukossa (taulukko).² Laske 99 %:n luottamusväli lasten lukumäärän odotusarvojen erotukselle alimman ja ylimmän koulutustason naisten välillä. Oleta laskussasi, että lasten lukumäärä on Poisson-jakautunut ja että otokset eri koulutustasoa olevista naisista ovat riippumattomia. Et tarvitse kaikkea taulukon tietoa luottamusvälin laskussa.

koulutustaso	lasten lukumäärän		
	keskiarvo	otosvarianssi	n
I	1.94	1.36	16216
II	1.83	1.30	9429
III	1.73	1.26	5231
IV	1.73	1.43	4336
I–IV (kaikki)	1.85	1.38	35212

Selityksiä: I = alin koulutustaso, ..., IV = ylin koulutustaso ja n on otoskoko. Alimman rivin tiedot ovat koko otoksesta.

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

²J. Nisén, M. Myrskylä, K. Silventoinen ja P. Martikainen (2014): Effect of Family Background on the Educational Gradient in Lifetime Fertility of Finnish Women Born 1940–50. *Population Studies: A Journal of Demography*, 68, 321–337.

2–3. Suomen kuvalehti 23.5.2014:

Kotikenttäedun arvoitus ratkeaa. Yleisön huuto vaikuttaa tuomariin. Vaikutus hävisi, kun tuomaria videoitiin. – – Sports Illustrated -lehden urheilutoimittaja Jon Wertheim ja Chicagon yliopiston taloustieteen professori Tobias Moskowitz käsittelevät kaikenlaisia urheiluun liittyviä myyttejä tilastojen valossa kirjassaan *Scorecasting: The Hidden Influences Behind How Sports Are Played and Games Are Won* (2011).

Wertheim ja Moskowitz louhivat suuria määriä urheiludataa erottaakseen yleisön vaikutuksen pelaajiin yleisön vaikutuksesta tuomareihin. He huomasivat, että jalkapallossa vieraille tuomitaan useammin keltainen tai punainen kortti kuin kotijoukkueelle. – –

Wertheimin ja Moskowitzin mukaan on hämmästyttävää, kuinka universaali ilmiö kotikenttätuet on. Se on säilynyt vuosikymmenestä toiseen merkittävän suuruisena eri lajeissa, eri kulttuureissa, niin miehillä kuin naisilla ja kaikilla sarjatasoilla. Suurin kotikenttäedun vaikutus on jalkapallossa – lajissa, jossa yksikin rangaistuspotku voi hyvin ratkaista koko ottelun.

Englannin ja Ranskan maajoukkueiden otteluiden tulokset syyskuusta 1996 lokakuun puoliväliin 2011 ovat luokiteltuna pelipaikan eli mahdollisen kotikenttäedun mukaan taulukossa³ alla.

	voitto	tasapeli	häviö	Σ	voitto-osuus
Englanti					
kotiottelut	51	17	13	81	0.629
vierasottelut	33	16	11	60	0.550
neutraalit ottelut	15	10	8	33	0.455
kaikki ottelut	99	43	32	174	0.569
Ranska					
kotiottelut	67	24	11	102	0.657
vierasottelut	40	17	9	66	0.606
neutraalit ottelut	16	7	9	32	0.500
kaikki ottelut	123	48	29	200	0.615

a) Testaa 5 %:n riskitasolla nollahypoteesia, että Englannin joukkue voittaa ottelun (koosteryhmä ”kaikki ottelut”) todennäköisyydellä 0.5 (kaksisuuntainen testi). Testaa 1 %:n riskitasolla nollahypoteesia, että Ranskan joukkue voittaa ottelun todennäköisyydellä 0.5 (kaksisuuntainen testi).

b) Selitä, mistä johtuu, että edellisissä testeissä toisessa nollahypoteesi hylätään mutta toisessa ei.

c) Testaa 2.5 %:n riskitasolla nollahypoteesia, että Ranskan joukkueen todennäköisyys voittaa koti- ja vieraskentällä on yhtäsuuri. Tee testi yksisuuntaisena siten, että vastahypoteesi on, että todennäköisyys voittaa kotikentällä on suurempi.

d) Oletetaan, että tehtävänä olisi testata nollahypoteesia, että Englannin ja Ranskan joukkueet voittavat ottelun (koosteryhmät ”kaikki ottelut”) samalla todennäköisyydellä. Päätisivätkö suhteellisten osuuksien vertailutestin taustaolelutukset? Pohdi ja perustele huolella.

³R. Bachan, B. Reilly ja R. Witt (2014): Team Performance and Race: Evidence from the English and French National Soccer Teams. *Applied Economics*, 46, 1535–1546.

4. Chicagon yliopiston Kansallinen mielipidetutkimuskeskus (*National Research Center*) kysyi yhdysvaltalaiselta aikuisväestöotokselta vuonna 2006 onnellisuudesta ja seksikumppaneiden lukumäärästä menneen vuoden aikana.⁴ Tulokset ovat taulukoissa alla. Oleta, että kysely tehtiin yksinkertaisella satunnaisotannalla.

a) Miten onnellisuus ja seksikumppanien lukumäärä vaikuttavat riippuvan toisistaan? (2 p)

b) Testaa 1 %:n riskitasolla, ovatko onnellisuus ja seksikumppanien lukumäärä riippumattomia muuttujia. Mitä päättelet? (4 p)

seksikumppanien lkm	onnellisuus			Σ (lkm)
	ei kovin	melko	hyvin	
0	112	329	154	595
1	118	832	535	1485
≥ 2	57	198	57	312
Σ (lkm)	287	1359	746	2392

seksikumppanien lkm	onnellisuus			Σ (%)
	ei kovin	melko	hyvin	
0	18.82	55.29	25.88	100
1	7.95	56.03	36.03	100
≥ 2	18.27	63.46	18.27	100
Σ (%)	12.00	56.81	31.19	100

⁴Tehtävä on A. Agrestin (2010) kirjasta *Analysis of Ordinal Categorical Data*. 2. laitos. Wiley. (S. 202.)