

Kirjoita selvästi jokaiseen vastauspaperiin:

- Kurssin koodi, päivämäärä, kokeen tyyppi (välikoe 2)
- Opiskelijanumero + kirjain
- TEKSTATEN sukunimi ja kaikki etunimet
- Koulutusohjelma ja vuosikurssi
- Nimikirjoitus

Vastausohje: Vastaa lyhyesti ja ytimekkäästi, mutta perustele ratkaisusi. 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. Selitä muutamalla rivillä, mitä seuraavat tilastotieteen käsitteet tarkoittavat:

- (a) Pearsonin otoskorrelaatio. (2 p)
- (b) Tilastollisen hypoteesitestin p-arvo. (2 p)
- (c) Parametrin estimointiin liittyvä havaitusta datajoukosta laskettu 90 % luottamustason luottamusväli. (2 p)

2. Kimblen ranking-tilastoissa Adele on luokiteltu Britneytä korkeammalle. Rankingien luotettavuutta epäilevä teekkari haluaa selvittää, onko Adele todellakin parempi pelaaja kuin Britney. Yleisenä hypoteesinä voidaan olettaa, että yksittäisessä pelissä Adele voittaa todennäköisyydellä $\theta \in [0, 1]$, ja että eri pelikertojen tulokset ovat keskenään riippumattomat. Teekkari järjestää turnauksen, jossa Adele ja Britney pelaavat keskenään viiden pelin sarjan, ja hän muotoilee hypoteesit

$$H_0 : \text{Pelaajat ovat yhtä hyviä, joten } \theta = 0.5$$

$$H_1 : \text{Adele on parempi, joten } \theta > 0.5.$$

Teekkari valitsee testin merkitsevyytasoksi 5 % ja testisuureksi $V =$ "Adelen voittojen lukumäärä viiden pelin sarjassa". Jos yleinen hypoteesi pätee, niin testisuure noudattaa $\text{Bin}(5, \theta)$ -jakaumaa.

- (a) Jos nollihypoteesi pätee, mikä on Adelen voittojen lukumäärän odotusarvo? (1 p)
- (b) Oletetaan, että Adele voittaa turnauksessa kaikki viisi peliä. Testaa nollihypoteesiä ja kerro, mikä on tekkarin johtopäätös testistä. (3 p)
- (c) Oletetaan sitten, että pelaajilla tosiaan on tasoero siten, että Adelen voiton todennäköisyys on $\theta = 0.7$. Mikä on todennäköisyys, että tasoerosta huolimatta teekkari hyväksyy nollihypoteesin? (2 p)

3. Suuren elektroniikka-alan sopimusvalmistajan tehtailla otettiin käyttöön uusi työsuoje luohjelma, jonka tavoitteena on tehostaa yrityksen tuottavuutta parantamalla työturvallisuutta. Alla on listattu työtaturmien takia menetetyt viikottaiset henkilötyötunnit (keskiarvoistettuna 1 kk ajalta) kymmenessä samankaltaisessa tehtaassa ennen ja jälkeen työsuoje luohjelman käyttöönoton.

Tehdas	1	2	3	4	5	6	7	8	9	10
Ennen	30.5	18.5	24.5	32	16	15	23.5	25.5	28	18
Jälkeen	23	21	22	28.5	14.5	15.5	24.5	21	23.5	16.5
Erotus	7.5	-2.5	2.5	3.5	1.5	-0.5	-1	4.5	4.5	1.5

Yllä listattujen lukujen keskiarvot ovat 23.15 (ennen), 21.00 (jälkeen) ja 2.15 (erotus). Vastaavat otosvarianssit ovat 36.56 (ennen), 19.17 (jälkeen) ja 9.00 (erotus).

Testaa 5 % merkitsevyytasolla, onko työsuoje luohjelmasta ollut hyötyä. Käytä nollahypoteesina oletusta, että työsuoje luohjelmalla ei ole vaikutusta työtaturmien takia menetettyjen työtuntien lukumäärään. (6 p)

4. Urnassa on 16 kuulaa, joista osa on mustia ja loput valkeita. Teekkari kahmaisee urnasta käteensä umpimähkään kolme kuulaa, havaitsee niiden joukossa kaksi mustaa ja yhden valkean. Tämän jälkeen teekkari palauttaa kuulat urnaan ja sekoittaa kuulat.
- (a) Merkitään symbolilla θ mustien kuulien (tuntematonta) lukumäärää urnassa. Olkoon $X \in \{0, 1, 2, 3\}$ mustien kuulien havaittu lukumäärä, joka havaittaisiin, mikäli teekkari toistaisi yllä kuvatun otoskokeen. Perustelee, miksi

$$\mathbb{P}(X = 2 | \theta = k) = \begin{cases} \frac{k(k-1)(16-k)}{m}, & k = 2, 3, \dots, 15, \\ 0, & \text{muuten,} \end{cases}$$

ja määritä vakion m arvo. (2 p)

- (b) Määritä tekkarin suorittaman havainnon pohjalta suurimman uskottavuuden estimaatti mustien kuulien lukumäärälle urnassa. (4 p)