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 (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, tai **0** = Tarua. Kunkin kohdan oikeasta vastauksesta saa pisteen, väärästä menettää pisteen — kuitenkin niin, että tehtävästä kokonaisuudessaan ei anneta negatiivisia pisteitä. (±1 p/kohta)

   (a) Esimerkiksi gallupissa haastateltujen henkilöiden puoluekannat ovat tilastollisen aineiston määrällisiä (eli kvantitatiivisia) muuttujia.

   (b) Oletetaan, että aineistossa havainnot $x_1, \ldots, x_n$ ovat riippumattomia ja noudattavat normaalijakaumaa tuntemattomin parametrein μ ja σ^2. Silloin $\sum_{j=1}^{n} \frac{(x_j-\mu)^2}{\sigma^2}$ on eräs aineiston tunnusluku.

   (c) Tilastokokeen stokastisen mallin jatkuvan parametrin θ suurimman uskottavuuden estimaatti on se satunnaisen parametrin θ mahdollinen arvo $\hat{\theta}$, jossa θ:n tiheysfunktio saavuttaa maksiminsa.

   (d) Tarkastellaan lineaarista regressiomallia $Y_j = \beta_0 + \beta_1 x_j + \epsilon_j$ $(j = 1, \ldots, n)$, missä x_j ovat syötemuuttujan arvot ja ϵ_j riippumattomat $N(0, \sigma^2)$-jakaumaa noudattavat virhetermit, ja Y_j vastemuuttujan satunnaiset arvot. Mallin parametrit ovat β_0, β_1 ja σ^2. Pienimmän neliösumman menetelmällä saadaan estimaatti parametrille σ^2.

   (e) Tilastollisessa hypoteesintestauksessa p-arvo on todennäköisyys nollahypoteesia vastaavassa stokastisessa mallissa sille, että testisuure saisi vähintään yhtä poikkeuksellisen arvon kuin havaitussa aineistossa.

   (f) Tilastollisessa hypoteesintestauksessa merkitsevyystason valinnalla voidaan vaikuttaa hylkäysvirheiden todennäköisyyteen.

2. Testataan uutta sairauden hoitomuotoa. Vertailukohtana käytettävän vakiintuneen hoitomuodon on todettu toimivan 78% tapauksissa. Kliinisissä testeissä uutta hoitomuotoa kokeillaan 270 potilaaseen, ja hoito osoittautuu toimivaksi 222 potilaan tapauksessa.

   (a) Kuvaile sopiva tilastokokeen stokastinen malli potilaiden hoitojen toimivuudelle kliinisissä testeissä. Mitä tämän stokastisen mallin testisuuretta on perusteltua approksimoida normaalijakaumalla ja miksi? (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)