

**Kirjoita ratkaisusi tietokoneella tai selkeällä käsialalla paperille.** Mikäli käytät tietokonetta apuvälineenä laskemiseen, kerro selkeästi mitä tuloksia olet laskenut tietokoneella, millä ohjelmalla, ja mitä komentoja olet käyttänyt.

**Palautusohje:** Palauta ratkaisusi MyCourses-sivuston kautta yhtenä pdf-tiedostona sisältäen nimesi ja opiskelijanumerosi selkeästi esillä ensimmäisellä sivulla. Palauta ratkaisusi **5.6.2021 klo 12.00 mennessä**. Pidä tarkasti huolta, että palautat ratkaisusi ajoissa.

**Vastausohje:** Vastaa lyhyesti ja ytimekkäästi, mutta **perustele ratkaisusi huolella**. Arvostelussa kiinnitetään erityistä huomiota asioiden ymmärtämiseen. Siten on erittäin tärkeää, että osoitat osaamisesi vastauksissasi. Vastaus ilman perustelua ei anna pisteitä. Mikäli vastauksestasi ei ilmene selvästi miten olet asian sisäistänyt, voi myös se aiheuttaa vähennyksiä pisteisiin. Kokeessa on 5 tehtävää, joista jokaisesta saa 0–6 pistettä.

**Sallitut apuvälineet:** laskin, kurssimateriaali, excel, tai mitä koetkaan tarpeelliseksi. Kotitentissä saat vapaasti käyttää saatavilla olevia materiaaleja.

---

1. Vastaa lyhyesti.

- (a) Selitä käsite ehdollinen todennäköisyys. (1p)
- (b) Selitä käsite estimaatti. (1p)
- (c) Selitä käsite satunnaismuuttuja. (1p)
- (d) Selitä käsite keskeinen raja-arvolause. (1p)
- (e) Esitä näkemyksesi siitä, mitä hyötyä keskeisestä raja-arvolauseesta voisi olla käytännön tilastollisissa ongelmissa? (2p)

2. Selosta omin sanoin tilastollisen testaamisen idean pääpiirteet. Selostuksestasi on löydyttävä (ainakin) käsitteet: (6p)

- yleinen hypoteesi,
- nollahypoteesi ja vastahypoteesi,
- testisuure ja testisuureen jakauma,
- p-arvo ja merkitsevyystaso,
- hylkäys- ja hyväksymisvirhe,
- johtopäätökset.

3. (a) Selosta lyhyesti, mitä tarkoittavat estimaattorin harhattomuus sekä tarkentuvuus. (2p)
- (b) Mittauksista saatiin havainnot  $(-2, 0, 1, 3, -2, 4)$ . Laske estimaatit odotusarvolle  $\mu$  ja hajonnalle  $\sigma$ . **Huomautus:** Raportoi myös käyttämäsi kaavat - pelkkä Excelin vastaus ei riitä. (2p)
- (c) Olkoon  $Z \sim N(0, 1)$ . Määrää todennäköisyydet  $\mathbb{P}(Z < 1)$ ,  $\mathbb{P}(Z < -1)$ , ja  $\mathbb{P}(|Z| > 1)$  käyttämällä hyväksesi todennäköisyyden laskusääntöjä, normaalijakauman ominaisuuksia, sekä tietoa  $\mathbb{P}(-1 \leq Z \leq 1) = 0.6827$  (et siis voi poimia todennäköisyyksiä excelistä, vaan sinun on johdettava vaaditut todennäköisyydet annetuista tiedoista). (2p)

4. (a) Yliopisto järjestää suuntauksen X pääaineen syventäviä kursseja yhteensä 12 kappaletta, jotka on jaettu ryhmiin A (pakolliset) ja B (valinnaiset). Ryhmässä A on kursseja kuusi, ja niistä on suoritettava vähintään kolme. Loput kurssit voi valita vapaasti ryhmästä B. Erkkä Ekonomi suorittaa opintoihinsa seitsemän kurssia. Montako erilaista tapaa Erkalla on valita kurssinsa? (2p)
- (b) Selosta omin sanoin, minkälaisissa tilanteissa satunnaismuuttuja noudattaa Hypergeometrista jakaumaa  $Hypergeometric(N, K, n)$ ? Tulkitse myös parametrit  $N, K, n$ , ja perustele (omin sanoin) miksi pistetodennäköisyysfunktio on

$$\Pr(X = k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}, \quad k \leq K \text{ ja } n - k \leq N - K.$$

(2p)

- (c) Seuraavan sivun alussa on pätkä maaseudun tulevaisuus lehden gallup-kyselyitä koskevista artikkelista (<https://www.maaseuduntulevaisuus.fi/politiikka/artikkeli-1.948693>). Poimi tekstistä kolme virhettä, ja selosta miten kyseinen kohta on virheellinen. (2p)

*Mielipidetutkimuksiin liittyy aina virhemarginaali.*

*Tarkalleen virhemarginaali tarkoittaa, että todellinen tulos on 95 prosentin todennäköisyydellä virhemarginaalin sisällä. Tuhannen ihmisen haastattelussa virhemarginaali on 1,5 - 3 prosenttiyksikköä. Virhemarginaali vaihtelee tuloksen mukaan niin, että lähellä 50 prosenttia olevissa tuloksissa virhemarginaali on suurin. MT kertoo tulosten yhteydessä aina suurimman mahdollisen virhemarginaalin.*

*Ajatellaan esimerkkinä, että puolueen kannatus on mittauksessa kymmenen prosenttia ja virhemarginaali kaksi prosenttiyksikköä suuntaansa. Tällöin on viiden prosentin mahdollisuus, että oikea kannatus onkin seitsemän tai 13 prosenttia.*

*Gallupissa kerrotaan kaikkein todennäköisin tulos. On siis todennäköisempää, että oikea tulos on prosenttiyksikön sisällä kerrotusta, kuin kolmen prosenttiyksikön vaikka virhemarginaali olisikin kolme prosenttiyksikköä.*

5. (a) Markettiin toimitettavista ruoka-annoksista 30% tulee tehtaalta X, 40% tehtaalta Y, ja 30% tehtaalta Z. X:n valmistamista annoksista pilaantuneita on 0.04%. Vastaavasti 0.03% Y:n valmistamista annoksista ja 0.03% Z:n valmistamista annoksista on pilaantuneita. Marketin johtajana haluat vähentää pilaantuneiden annosten osuutta (välittämättä siitä, minkä verran tuotettu kokonaismäärä vähenee). Jos voit katkaista toimitus-sopimuksen vain yhden tehtaan kanssa, minkä kolmikosta X, Y, tai Z valitsisit? Perustele vastauksesi huolella! (2p)
- (b) Laadunvalvonnassa verrattiin kahta suklaalevyjä tuottavaa tuotantolinjaa keskenään poimimalla molemmista  $n = 200$  suuruinen otos. Linjan 1 otoksesta mitattiin keskiarvoksi 195 grammaa sekä hajonnaksi 5 grammaa, ja linjan 2 otoksesta mitattiin keskiarvoksi 205 grammaa sekä hajonnaksi 8 grammaa. Testaa 5% merkitsevyydellä, eroavatko linjojen tuottamien levyjen painot toisistaan. (2p)
- (c) Yrityksen johtaja väittää, että vain yhdessä prosentissa tuotteista esiintyy vikoja. Perusteluiksi poimittiin satunnaisesti 50 tuotetta suhteellisen osuuden testiä varten. Nol-

lahypoteesina oli, että viallisten osuus on 1%, ja testissä käytettiin kaksisuuntaista vaihtoehtoista hypoteesia. Testisuureen arvoksi saatiin 3.03, josta edelleen normaalijakau-  
masta p-arvoksi noin 0.2%. Yritysjohdaja hehkuttaa olleensa oikeassa. Esitä ja perustele  
huolellisesti analyysisi testaus-asetelmasta ja sen tuloksista. Ovatko tulokset mielestäsi  
luotettavia? Löydätkö mahdollisia virhelähteitä? (2p)