30A0200 Tilastotieteen perusteet
Mallivastaukset

Date of exam: 7.3.2022 **Solved by**: mathwiz.ai

Disclaimer:
These problems are solved by mathwiz.ai, a service leveraging the latest generative AI technologies. This means the solutions come with inherent limitations:

1. **Understanding:** The AI doesn't understand mathematics like humans do; it predicts answers based on historical data patterns.

2. **Reliability and Context:** The AI's responses can be contextually inappropriate or inconsistent, and thus contain errors.

3. **Creativity:** AI applies known patterns, it does not invent new mathematical methods.

Use these solutions as guides, not definitive answers. Despite occasional errors, the solutions can be useful for providing context around the problems. For verifiable accurate solutions, seek guidance from a qualified educator. At the time of solving these problems, the mathematical capabilities of AIs were at the level of an advanced university mathematics student. If you find errors in the solutions, feel free to inform us. Contact information: mathwizcontact@gmail.com.

Problem 1

Selitä lyhyesti seuraavat käsitteet omin sanoin.

(a) Ehdollinen todennäköisyys.
(b) Keskeinen raja-arvolause.
(c) Satunnaismuuttuja ja sen jakauma.

Problem 1 solution

(a) Ehdollinen todennäköisyys

Ehdollinen todennäköisyys on todennäköisyys, että jokin tapahtuma tapahtuu, kun tiedetään, että jokin toinen tapahtuma on jo tapahtunut. Se lasketaan jakamalla tapahtumien A ja B yhteisen todennäköisyyden tapahtuman B todennäköisyydellä. Matemaattisesti tämä voidaan esittää seuraavasti:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

(b) Keskeinen raja-arvolause

Keskeinen raja-arvolause on tilastotieteen perusperiaate, joka kertoo, että suuresta määrästä satunnaisia, samanlaisia muuttujia muodostettu summa tai keskiarvo noudattaa normaalijakaumaa, riippumatta alkuperäisten muuttujien jakaumasta.

(c) Satunnaismuuttuja ja sen jakauma

Satunnaismuuttuja on muuttuja, jonka arvo riippuu satunnaisesta ilmiöstä. Sen jakauma kertoo, kuinka todennäköistä on, että satunnaismuuttuja saa tietyn arvon. Esimerkiksi, jos X on satunnaismuuttuja, joka kuvaa kolikonheiton tulosta (1 jos tulee kruuna, 0 jos tulee klaava), sen jakauma voisi olla seuraava:

$$P(X = 0) = 0.5, P(X = 1) = 0.5$$

Problem 2

(a) Selosta lyhyesti käsitteet estimaattori ja estimaatti, sekä selitä niiden ero.
(b) Selosta lyhyesti, mitä tarkoittavat estimaattorin harhattomuus sekä tarkentuvuus. (2p)
(c) Mittauksista saatiin havainnot $(-2, 0, 1, 3, -2, 4)$. Laske estimaatit odotusarvolle μ ja hajonnalle σ käyttäen jotakin harhatonta estimaattoria. Huomautus: Raportoi myös käyttämäsi kaavat - pelkkä Excelin vastaus ei riitä.
(2p)
Problem 2 solution

(a) Estimaattori ja estimaatti

Estimaattori on tilastollinen menetelmä tai kaava, jota käytetään tuntemattoman parametrin, kuten populaation keskiarvon tai hajonnan, arvioimiseen. Estimaatti on sen sijaan estimaattorin tuottama arvo eli konkreettinen arvio tuntemattomasta parametrista.

Ero näiden kahden välillä on, että estimaattori on menetelmä tai kaava, kun taas estimaatti on tämän menetelmän tai kaavan tuottama luku.

(b) Estimaattorin harhattomuus ja tarkentuvuus

Estimaattorin harhattomuus tarkoittaa, että estimaattorin odotusarvo on yhtä suuri kuin estimoitavan parametrin todellinen arvo. Toisin sanoen, jos estimaattoria käytetään monta kertaa, keskimääräinen estimaatti on lähellä todellista parametrin arvoa.

Tarkentuvuus tarkoittaa, että estimaattorin varianssi pienenee, kun otoskoko kasvaa. Toisin sanoen, mitä enemmän dataa on käytettävissä, sitä tarkempi estimaatti on.

(c) Estimaatit odotusarvolle μ ja hajonnalle σ

Odotusarvon estimaatti on otoksen keskiarvo. Se lasketaan seuraavasti:

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Hajonnan estimaatti on otoksen varianssi. Se lasketaan seuraavasti:

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2$$

Kun otos on $(-2, 0, 1, 3, -2, 4)$, saadaan

$$\mu = \frac{1}{6}(-2 + 0 + 1 + 3 - 2 + 4) = 0.67$$

ja

$$\sigma^2 = \frac{1}{6-1}((-2-0.67)^2+(0-0.67)^2+(1-0.67)^2+(3-0.67)^2+(-2-0.67)^2+(4-0.67)^2) = 4.67$$

Joten odotusarvon estimaatti on 0.67 ja hajonnan estimaatti on 4.67.

Problem 3

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.
(a) Millä todennäköisyydellä satunnaisesti valittu annos on pilaantunut?
(b) Valittu annos osoittautuu pilaantuneeksi. Millä todennäköisyydellä sen on valmistanut tehdas X ?
(2p) (c) 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! (3p)

Problem 3 solution

(a) Todennäköisyys, että satunnaisesti valittu annos on pilaantunut

Tämä voidaan laskea käyttämällä ehdollisen todennäköisyyden kaavaa ja summaamalla kaikki mahdolliset tapaukset. Tämä voidaan esittää seuraavasti:

$$P(\text{Pilaantunut}) = P(\text{Pilaantunut}|\text{X})P(\text{X})+P(\text{Pilaantunut}|\text{Y})P(\text{Y})+P(\text{Pilaantunut}|\text{Z})P(\text{Z})$$

Kun arvot korvataan, saadaan

$$P(\text{Pilaantunut}) = 0.0004 \cdot 0.3 + 0.0003 \cdot 0.4 + 0.0003 \cdot 0.3 = 0.00034$$

(b) Todennäköisyys, että pilaantunut annos on valmistanut tehdas X

Tämä voidaan laskea käyttämällä Bayesin teoreemaa, joka voidaan esittää seuraavasti:

$$P(\text{X}|\text{Pilaantunut}) = \frac{P(\text{Pilaantunut}|\text{X})P(\text{X})}{P(\text{Pilaantunut})}$$

Kun arvot korvataan, saadaan

$$P(\text{X}|\text{Pilaantunut}) = \frac{0.0004 \cdot 0.3}{0.00034} = 0.3529$$

(c) Valinta tehtaasta, jonka kanssa katkaista toimitussopimus

Valinta tehdään sen perusteella, mikä tehdas tuottaa suurimman osuuden pilaantuneista annoksista. Tämä voidaan laskea käyttämällä ehdollisen todennäköisyyden kaavaa jokaiselle tehtaalle:

$$P(\text{Pilaantunut}|\text{X})P(\text{X}) = 0.0004 \cdot 0.3 = 0.00012$$
$$P(\text{Pilaantunut}|\text{Y})P(\text{Y}) = 0.0003 \cdot 0.4 = 0.00012$$
$$P(\text{Pilaantunut}|\text{Z})P(\text{Z}) = 0.0003 \cdot 0.3 = 0.00009$$

Näistä suurin on tehtaiden X ja Y tuottama osuus. Koska kysymys ei määrittele, miten valita tehdas, jos useampi tehdas tuottaa saman osuuden pilaantuneista annoksista, voidaan valita jompikumpi näistä tehtaista.

Problem 4

(a) 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).
(3p)
(b) Terveysvalmisteen vaikuttavan aineen pitoisuus $X \sim N\left(\mu, \sigma^2\right)$. Vaikuttavan aineen pitoisuus valmisteissa on keskimäärin $\mu = 100$mg. Tuottaja pystyy tuotanto-prosessin viilauksella säätämään hajontaa σ. Minkä suuruiseksi σ on asetettava, jotta vähintään 90% valmisteista sisältää vaikuttavaa ainetta vähintään 80mg, mutta korkeintaan 120 mg? Käytä normalisointia!
(3p)

Problem 4 solution

(a) Todennäköisyyksien määrittäminen

Koska Z on standardinormaalijakaumaa noudattava satunnaismuuttuja, tiedämme, että sen jakauma on symmetrinen nollan ympärillä. Tämä tarkoittaa, että $P(Z < 1) = P(Z > -1)$ ja $P(Z < -1) = P(Z > 1)$.

Lisäksi tiedämme, että $P(-1 \leq Z \leq 1) = 0.6827$. Koska jakauma on symmetrinen ja tämä todennäköisyys kattaa suurimman osan jakaumasta, voimme päätellä, että $P(Z < 1) = P(Z > -1) = 0.5 \times 0.6827 = 0.34135$ ja $P(Z < -1) = P(Z > 1) = 0.5 - 0.34135 = 0.15865$.

Lopuksi, $P(|Z| > 1)$ on todennäköisyys, että Z on suurempi kuin 1 tai pienempi kuin -1. Tämä voidaan laskea seuraavasti:

$$P(|Z| > 1) = P(Z > 1) + P(Z < -1) = 2 \times P(Z > 1) = 2 \times 0.15865 = 0.3173$$

(b) Hajonnan määrittäminen

Koska haluamme, että vähintään 90 prosenttia valmisteista sisältää vaikuttavaa ainetta vähintään 80 mg, mutta korkeintaan 120 mg, meidän on löydettävä hajonta σ, joka saa tämän tapahtumaan.

Koska tiedämme, että X noudattaa normaalijakaumaa, voimme normalisoida tämän ongelman käyttämällä standardinormaalijakaumaa. Tämä tarkoittaa, että meidän on löydettävä σ, joka saa seuraavat ehdot täyttymään:

$$P(80 \leq X \leq 120) = 0.9$$

Kun normalisoidaan, saadaan

$$P\left(\frac{80-\mu}{\sigma} \leq Z \leq \frac{120-\mu}{\sigma}\right) = 0.9$$

Koska tiedämme, että $\mu = 100$, voimme laskea nämä arvot:

$$P\left(\frac{-20}{\sigma} \leq Z \leq \frac{20}{\sigma}\right) = 0.9$$

Koska tiedämme, että $P(-1 \leq Z \leq 1) = 0.6827$ ja $P(-1.645 \leq Z \leq 1.645) = 0.9$, voimme asettaa nämä arvot yhtä suuriksi ja ratkaista σ:

$$\frac{20}{\sigma} = 1.645$$

Ratkaisemalla tämän saadaan $\sigma = \frac{20}{1.645} \approx 12.16$. Joten hajonnan on oltava noin 12.16 mg, jotta vähintään 90

Problem 5

(a) Kurssilla käsitellyistä testeistä neljässä (testi odotusarvolle, testi suhteelliselle osuudelle, odotusarvojen parivertailu, sekä suhteellisten osuuksien parivertailu) p-arvo ja/tai kriittiset rajat laskettiin normaalijakaumasta. Perustele lyhyesti omin sanoin, miksi normaalijakauman käyttö kyseisissä testeissä on perusteltua.

(b) Yrityksen johtaja väittää, että vain yhdessä prosentissa tuotteista esiintyy vikoja. Perusteluiksi poimittiin satunnaisesti 50 tuotetta suhteellisen osuuden testiä varten. Nollahypoteesina oli, että viallisten osuus on 1%, ja testissä käytettiin kaksisuuntaista vaihtoehtoista hypoteesia. Testisuureen arvoksi saatiin 3.03 , josta edelleen normaalijakaumasta p-arvoksi noin 0.2%. Yritysjohtaja hehkuttaa olleensa oikeassa. Esitä ja perustele huolellisesti analyysisi testausasetelmasta ja sen tuloksista. Ovatko tulokset mielestäsi luotettavia? Löydätkö mahdollisia virhelähteitä?

Problem 5 solution

(a) Normaalijakauman käyttö testeissä

Normaalijakauman käyttö on perusteltua näissä testeissä, koska se on yleisimmin käytetty jakauma tilastollisessa analyysissä. Tämä johtuu siitä, että monien satunnaismuuttujien, kuten odotusarvon ja suhteellisen osuuden, uskotaan noudattavan normaalijakaumaa suurissa otoksissa keskeisen raja-arvolauseen mukaan. Lisäksi normaalijakauma on symmetrinen, mikä tekee siitä sopivan parivertailuihin, joissa tarkastellaan kahden ryhmän välistä eroa.

(b) Analyysi testausasetelmasta ja sen tuloksista

Testin tulokset viittaavat siihen, että yritysjohtajan väite ei pidä paikkaansa, koska p-arvo on pienempi kuin yleisesti käytetty merkitsevyystaso 0.05. Tämä tarkoittaa, että nollahypoteesi, jonka mukaan viallisten osuus on 1 %, voidaan hylätä.

Kuitenkin, on tärkeää huomata, että testin luotettavuus riippuu otoskoon suuruudesta ja otannan satunnaisuudesta. Tässä tapauksessa otoskoko on melko pieni (50 tuotetta), mikä saattaa vaikuttaa testin tarkkuuteen ja luotettavuuteen. Lisäksi, jos otanta ei ole satunnaista, se saattaa vääristää tuloksia.

Lisäksi, vaikka p-arvo on pienempi kuin 0.05, se ei ole paljon pienempi. Tämä tarkoittaa, että on olemassa noin 0.2 % todennäköisyys, että saadut tulokset johtuvat sattumasta, jos nollahypoteesi on totta. Tämä ei ole merkittävästi pienempi kuin 0.05, joten tuloksia tulisi tulkita varoen.

Lopuksi, on tärkeää huomata, että vaikka testi osoittaa, että yritysjohtajan väite ei pidä paikkaansa, se ei kerro, onko viallisten osuus suurempi vai pienempi kuin 1%. Tämä olisi selvitettävä lisätutkimuksilla.