TMA4240 Statistikk
Høst 2020

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Skriftlig innlevering 5

Oppgave 1

En bestemt målemetode for bestemmelse av pH-verdien i en løsning gir måleresultater som antas å være uavhengige og normalfordelte, med forventningsverdi μ lik virkelig pH og varians $\sigma^2 = 0.060^2$. La $X_1, \ldots, X_n$ være uavhengige målinger av pH i en bestemt løsning.

a) Anta (bare i dette punktet) at den virkelige pH-verdien i en løsning er 6.8.

Hva er sannsynligheten for at en bestemt måling gir et resultat som er under 6.74?

Hva er sannsynligheten for at en bestemt måling gir et resultat som er mellom 6.74 og 6.86?

Hva er sannsynligheten for at en bestemt måling, X, gir et resultat som avviker mer enn 0.06 fra μ, dvs bestem $P(|X - \mu| > 0.06)$?

Du skal estimere pH i en løsning, og bruker gjennomsnittet av 5 uavhengige målinger som estimat. La Y være gjennomsnittet av 5 uavhengige målinger.

b) Hva er sannsynligheten for at Y avviker mer enn 0.06 fra μ?

Utled et 95% konfidensintervall for μ. Hva blir konfidensintervallet når gjennomsnittet av fem uavhengige målinger ble 6.76?

Oppgave 2

Det vurderes å utbedre en lite trafikkert, men farlig veistrekning i Trondheimsområdet. I den sammenheng blir du bedt om å analysere hvor mye trafikk det er på veien. La X være antall biler som passerer et bestemt punkt på veistrekningen fra kl 16:00 til kl 18:00 på en tilfeldig valgt hverdag. Vi antar at X er poissonfordelt med parameter λ, dvs

$$f(x) = \frac{\lambda^x}{x!} \exp\{-\lambda\}, \quad x = 0, 1, 2, \ldots$$

a) Anta bare i dette punktet at $\lambda = 15$.

Regn ut $\mathrm{P}(X > 20)$ og $\mathrm{P}(10 \leq X < 20)$.

Finn E(X).

Anta at verdien til λ er ukjent i resten av oppgaven. Vi ønsker nå å finne realistiske verdier for λ. Vi observerer derfor antall passerende biler fra kl 16:00 til kl 18:00 på n tilfeldig valgte hverdager, $X_1, X_2, \ldots, X_n$. Vi antar at observasjonene er uavhengige og identisk fordelt med samme poissonfordeling som tidligere beskrevet. Resultatet av målingene for $n = 30$ tilfeldige dager, $x_1, x_2, \ldots, x_{30}$, gir $\sum_{i=1}^{30} x_i = 359$.

b) Definér estimatoren

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Bruk sentralgrenseteoremet til å utlede et tilnærmet $(1-\alpha)\cdot 100\%$ konfidensintervall for λ basert på estimatoren $\hat{\lambda}$.

Regn ut konfidensintervallet for λ med de gitte dataene og $\alpha = 0.01$.

Kommunen ønsker å samle inn mer data om hvor trafikkert veien er. Derfor registreres det også hvor mange biler som passerer på veien fra kl 18:00 til kl 20:00 på m tilfeldig valgte hverdager, $Y_1, Y_2, \ldots, Y_m$. Vi antar at observasjonene er uavhengige og identisk poissonfordelte med parameter $\lambda/2$, altså halv intensitet i forhold til fra kl 16:00 til kl 18:00. Anta også at $Y_1, Y_2, \ldots, Y_m$ er uavhengige av $X_1, X_2, \ldots, X_n$.

c) Utled sannsynlighetsmaksimeringsestimatoren (maximum likelihood estimatoren) for λ basert på observasjonene $X_1, X_2, \ldots, X_n$ og $Y_1, Y_2, \ldots, Y_m$.

Oppgave 3

I situasjoner der det er uklart hvem som er den biologiske faren til et barn kan farskapet avklares ved å sammenligne DNA-prøver fra barnet med mulige fedre. For en mulig far gjøres dette ved å sammenligne n ulike deler av DNA-strukturen til mannen med de samme n deler av DNA-strukturen hos barnet. De n undersøkte delene av DNA-strukturen antas uavhengige.

Hos et barn og en tilfeldig valgt mann (som ikke er biologisk far) er det for hver enkel del av DNA-strukturen som undersøkes en sannsynlighet $p = 0.15$ for at delen er sammenfallende hos barnet og mannen. Anta videre at en biologisk far alltid har alle de undersøkte delene av DNA-strukturen sammenfallende med barnets (dvs. vi ser bort fra mutasjoner o.l.), slik at hver undersøkte del av DNA-strukturen hos biologisk far og barn er sammenfallende med sannsynlighet $p = 1$.

La X være antall sammenfallende deler i DNA-strukturen hos et barn og en tilfeldig valgt mann (som ikke er biologisk far).

a) Begrunn at X er binomisk fordelt med parametre n og $p = 0.15$.

Dersom $n = 5$, beregn sannsynlighetene $P(X = 2)$, $P(X \geq 2)$ og $P(X = 2 | X \geq 2)$.

I en farsskapssak blir en mann erklært å være biologisk far dersom alle undersøkte deler av DNA-strukturen er sammenfallende hos mannen og barnet. Dette kan vi se på som en hypotesetest der vi tester

$$H_0: \quad p = 0.15 \text{ (ikke far)} \qquad \text{mot} \qquad H_1: \quad p = 1.0 \text{ (far)}$$

der H_0 forkastes (dvs. mannen erklæres som far til barnet) dersom $X = n$.

b) For $n = 5$, finn sannsynligheten for å begå type 1 feil i testen over.

For $n = 5$, finn sannsynligheten for å begå type 2 feil i testen over.

Hvor mange ulike deler, n, av DNA-strukturen må man minst sammenligne dersom man ønsker at sannsynligheten for feilaktig å erklære en mann som far skal være mindre enn 0.000001?

Oppgave 4

Ved innsjekking på rutefly veies normalt ikke passasjerer med håndbagasje. Vekten av de enkelte passasjerer med håndbagasje, X_1, $X_2, ...$, antas uavhengige og normalfordelte (Gaussisk fordelte), hvor forventningsverdi μ og varians σ^2 normalt er kjent fra internasjonale statistikker. En bestemt flyrute betjenes av en flytype med 16 passasjerplasser.

a) La $X_1, ..., X_{16}$ være vekten av 16 passasjerer med håndbagasje, og la $Y = X_1 + ... + X_{16}$. Anta at $\mu = 80$ (kg) og $\sigma^2 = 18^2$ (kg^2). Beregn følgende:

i) $P(X_1 > 90)$ ii) $E(Y)$ iii) $\mathrm{Var}(Y)$ iv) $P(Y > 16 \cdot 90)$

Ville noen av de fire beregningene du har gjort vært korrekte uten å anta uavhengighet men med antagelse om normalfordeling? Angi i så fall hvilke.

Ville noen av de fire beregningene du har gjort vært korrekte uten å anta normalfordeling men med antagelse om uavhengighet? Angi i så fall hvilke.

I resten av oppgaven antas μ og σ^2 å være ukjente. For å skaffe informasjon om vekten av passasjerer med håndbagasje på denne flyruten, veies 20 tilfeldig trukne passasjerer med håndbagasje, med følgende resultat: $x_i, \ i = 1, 2, ..., 20$:

76, 97, 82, 73, 65, 74, 104, 55, 97, 69, 78, 86, 73, 90, 69, 86, 76, 66, 107, 68.

Disse gir: $\sum_{i=1}^{20} x_i = 1591, \ \sum_{i=1}^{20} x_i^2 = 130221.$

b) Beregn estimat for μ og σ^2 basert på dette datasettet. Finn et 90% konfidensintervall for μ.

c) Flyselskapet vil undersøke om $\sigma^2 < 15^2$ på denne flyruten. I så fall kan det eventuelt tillates økt vekt av frakt. Selskapet setter opp hypotesen

$$H_0: \quad \sigma^2 \geq 15^2 \qquad\qquad H_1: \quad \sigma^2 < 15^2.$$

Hva blir konklusjonen av en test med 1% signifikansnivå, ut fra datasettet ovenfor?

Fasit

1. **a**) $0.159, 0.682, 0.318$ **b**) $0.026, [6.707, 6.813]$

2. **a**) 0.0830, 0.8053 **b**) [10.34, 13.59]

3. **a**) $0.138, 0.165, 0.836$ **b**) $0.000076, 0, n = 8$

4. **a**) i) 0.289, ii) 1280, iii) 72^2, iv) 0.013 **b**) 79.55, 13.87^2, (74.2,84.9) **c**) Ikke forkast H_0