

i**Eksamensoppgave i TMA4240 Statistikk****Dato:** 17. desember**Tid:** 15-19**Faglig kontakt:** Jarle Tufto, Jo Eidsvik**Møter i eksamenslokalet:** Nei**Hjelpemiddelkode/Tillatte hjelpemidler:** Hjelpemiddelkode C.

- Tabeller og formler i statistikk (Akademika).
- Én bestemt enkel kalkulator I henhold til NTNUs generelle reglement.
- Matematisk formelsamling, Karl Rottmann.

ANNEN INFORMASJON:

Les oppgavene nøye og gjør dine egne antagelser. Presiser i besvarelsen hvilke forutsetninger du har lagt til grunn i tolkning/avgrensning av oppgaven.

Dersom det er direkte feil eller mangler i oppgavesettet og du ikke kan gjøre dine egne antagelser, vises det til informasjon om klage på formelle feil på NTNU sin nettside om «Begrunnelse og klage».

FAGSPESIFIKK INFORMASJON:

Automatisk rettede oppgaver: Oppgave **1-4** skal kun besvares direkte i Inspira (eventuelle håndtegninger vil ikke bli vurdert). Det er tilstrekkelig å oppgi svar med tre siffrers nøyaktighet.

Papirhåndtegninger: I oppgave **5-7** er det lagt opp til å besvare på ark. Nederst i oppgaven finner du en sjusifret kode. Fyll inn denne koden øverst til venstre på arkene du ønsker å levere. Det anbefales å gjøre dette underveis i eksamen. Dersom du behøver tilgang til kodene etter at eksamenstiden har utløpt, må du klikke «Vis besvarelse».

Du er selv ansvarlig for å fylle inn riktige koder på eventuelle håndtegningsark. Les derfor informasjonen på omslagsarket nøye. Eksamenskontoret kan ikke garantere at feilaktig utfylte ark blir lagt til besvarelsen din.

Merk at det kun er tillatt med blå eller svart kulepenn (ikke rød kulepenn).

Vekting av oppgavene: Maksimalt antall poeng av totalt 100 er angitt for hver oppgave.

Trekk fra/avbrutt eksamen: Dersom du ønsker å levere blankt/avbryte eksamen, gå til "hamburgermenyen" i øvre høyre hjørne og velg «Lever blankt». Dette kan ikke angres selv om prøven fremdeles er åpen.

Tilgang til besvarelse: Etter eksamen finner du besvarelsen din under tidligere prøver i Inspira. Merk at det kan ta én virkedag før eventuelle håndtegninger vil være tilgjengelige under «tidligere prøver».

- 1 La C være hendelsen at en person trukket tilfeldig fra populasjonen er smittet av covid-19 og la T være hendelsen at personen tester positivt på en covid-19 hurtigtest. Anta at $P(C) = 0.03$, at testens sensitivitet $P(T|C) = 0.7$, og at testens spesifisitet $P(T'|C') = 0.99$ der A' er komplementet av hendelsen A .

a) Hva blir $P(T|C')$?

b) Hva blir sannsynligheten for at personen tester positivt?

c) Hva blir $P(C|T)$?

d) Hva blir sannsynligheten for at personen ikke er smittet av covid-19 gitt at hurtigtesten gir positivt resultat?

Maks poeng: 10

- 2 Vi trekker 5 kort fra en vanlig kortstokk bestående av 52 kort.

a) Hvor mange uordnede utvalg kan vi trekke fra kortstokken?

b) På hvor mange måter kan vi få hus (henholdsvis tre og to kort av samme verdi, f.eks. 3 konger og 2 femere)?

c) Hva er sannsynlighet for hus?

Maks poeng: 7.5

- 3 Anta at den stokastisk variabelen X er kontinuerlig fordelt med sannsynlighetstetthet

$$f(x) = \begin{cases} 3x^2 & \text{for } 0 \leq x \leq 1 \\ 0 & \text{ellers.} \end{cases}$$

a) Finn $E(X)$.

b) Finn $E(X^2)$.

c) Finn $\text{Var}(X)$.

d) Finn $P(X \leq 1/2)$.

e) Finn medianen til X .

Maks poeng: 12.5

- 4 Anta at Y er kontinuerlig fordelt med sannsynlighetstetthet

$$f(y) = \frac{1}{\sqrt{50\pi}} e^{-\frac{(y-2)^2}{2 \cdot 5^2}}.$$

a) Hva er standardavviket til Y ?

b) Hva er forventningsverdien til Y ?

c) Finn $P(2 \leq Y \leq 4)$.

d) Finn $P(Y = 3)$.

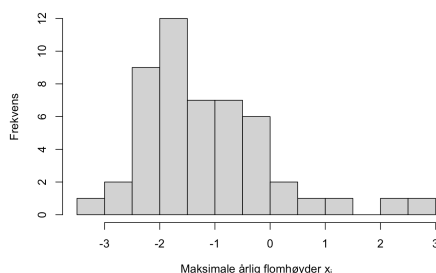
Maks poeng: 10

- 5 Anta at årlig maksimal flomhøyde X (målt i meter) relativt til et fast målepunkt i et vassdrag er fordelt slik at X har kumulativ fordeling $F_X(x) = e^{-e^{-x+\mu}}$ for $-\infty < x < \infty$ der μ er en ukjent parameter.

a) Vi definerer en hunderårsflom som den flomhøyden q som tilfredstiller $P(X \geq q) = 1/100$. Bestem q dersom $\mu = -1.6$.

b) Finn en algoritme for å simulere n realisasjoner av X og spesifiser algoritmen i pseudo- eller python-kode.

Anta at maksimale flomhøyder i ulike år er uavhengige og vi har observert $n = 50$ maksimale flomhøyder $x_1, x_2, \dots, x_n$ som vist i histogrammet under. Det oppgis at $\sum_{i=1}^n x_i = -58.14$ og $\sum_{i=1}^n e^{-x_i} = 253.4$.



c) Vis at X har sannsynlighetstetthet $f_X(x) = e^{-x+\mu-e^{-x+\mu}}$. Utled sannsynlighetsmaksimeringsestimatoren av μ og beregn estimatet av μ for utvalget gitt over.

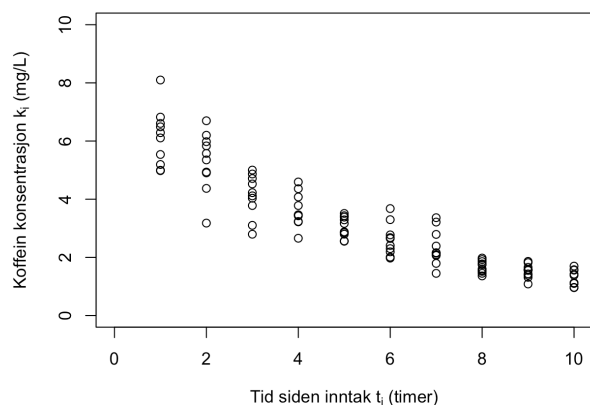
d) Vis at $Y = 2e^{-X+\mu}$ er khi-kvadratfordelt med 2 frihetsgrader. Er $2e^{\mu} \sum_{i=1}^n e^{-x_i}$ en pivotal størrelse? Utled et 95%-konfidensintervall for μ og beregn intervallet for dataene gitt over.

e) Finn et tilsvarende konfidensintervall for hundreårs flomhøyde q .

Skriv svaret ditt på utdelte ark

Maks poeng: 30

- 6 Vi måler koffeinkonsentrasjonen k_i på ulike tidspunkter t_i i et antall forsøkspersoner et visst antall timer etter inntak av en bestemt mengde koffein slik at vi tilsammen har $n = 100$ målinger som vist i følgende figur.



Vi skal anta at koffeinkonsentrasjonen avtar eksponentielt slik at logaritmen til koffeinkonsentrasjonen avtar lineær med tiden t_i . Vi kan da analysere dataene med regresjonsmodellen

$$\ln K_i = \beta_0 + \beta_1 t_i + \epsilon_i$$

der $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1, 2, \dots, n$ og antatt uavhengige. Det oppgis at

$$\sum_{i=1}^n (t_i - \bar{t})^2 = 825, \quad \sum_{i=1}^n (t_i - \bar{t}) \ln k_i = -143.6, \quad \text{og at}$$

$$\sum_{i=1}^n (\ln k_i - \hat{\beta}_0 - \hat{\beta}_1 t_i)^2 = 3.19.$$

- a) Beregn sannsynlighetsmaksimeringsestimatet $\hat{\beta}_1$ av stigningstallet β_1 (bruk uttrykket i formelsamlingen). Vis at

$$\text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (t_i - \bar{t})^2}.$$

Finn et estimat av standardfeilen til $\hat{\beta}_1$.

- b) Finn halveringstiden $t_{1/2}$ til forventet koffeinkonsentrasjonen uttrykt ved parameterene i modellen over og beregn et estimat $\hat{t}_{1/2}$ av halveringstiden. Bruk feilforplantningsformelen til å beregne et tilnærmet estimat av standardfeilen til estimert halveringstid.

Skriv svaret ditt på utdelte ark

Maks poeng: 12

- 7 Basert på historiske data mener Banenor at det langs en jernbanestrekning i gjennomsnitt skjer $\lambda_0 = 9$ viltpåkjørslar per år.

a) La den stokastiske variabel X betegne antall viltpåkjørslar i løpet av $t = 2$ år. Under hvilke forutsetninger vil X være Poissonfordelt? Hva blir forventningsverdien og standardavviket til X ? Finn sannsynligheten for at det skjer minst like mange påkjørslar som forventet i løpet av 2 år.

For å redusere antall påkjørslar bygges det en viltovergang og etter 2 år ønsker Banenor å undersøke om viltovergangen har redusert antall påkjørslar. Det ble registrert tilsammen 12 påkjørslar i de 2 årene etter at viltovergangen ble bygget.

b) Formuler problemet som en hypotesetest. Hva blir null- og alternativ hypotese? Foreslå en testobservator og bestem testens forkastningsregel, enten eksakt eller ved bruk av normaltilnærming, når signifikansnivået $\alpha = 0.05$. Hva blir testens konklusjon? Bestem også testens p -verdi gitt dataene over.

c) Hva blir teststyrken til testen over hvis det i realiteten er slik at $\lambda = 5$ etter at viltovergangen ble etablert. Hvor liten må λ tilnærmet være for at teststyrken skal bli minst 0.95?

Skriv svaret ditt på utdelte ark

Maks poeng: 18