Eksamensoppgåve i TMA4240 Statistikk
Dato: 17. desember
Tid: 15-19
Fagleg kontakt: Jarle Tufto, Jo Eidsvik
Møter i eksamenslokalet: Nei

Hjelpemiddelkode/Tillatne hjelpemiddel: Hjelpemiddelkode C.

- Tabeller og formler i statistikk (Akademika).
- Ein bestemd enkel kalkulator i henhald til NTNU sitt generelle reglement.
- Matematisk formelsamling, Karl Rottmann.

ANNAN INFORMASJON:

Les oppgåvene nøye og gjer di eigen formeining. Presiser kva du har lagd til grunn i tolking av oppgåva. Om det er direkte feil eller manglar i oppgåvesettet og du ikkje kan gjere din eigen formeining, viser ein til informasjon om klage på formelle feil på NTNU si nettside «Begrunnelse og klage».

FAGSPESIFIKK INFORMASJON:

Automatisk retta oppgåver: Oppgåve **1-4** skal kun svarast direkte i Inspera (eventuelle handteikninger vert ikkje vurdert). Det er tilstrekkeleg å oppgje svar med tre siffers nøyaktigheit.

Papirhandteikninger: I oppgåve **5-7** er det lagd opp til å svara på ark. Nedst i oppgåva finn du ein sjusifra kode. Fyll inn denne koden øverst til venstre på arkene du ønskjer å levera. Det er anbefala å gjere dette undervegs i eksamen. Om du behøver tilgang til kodane etter at eksamenstida har løpt ut, må du klikke «Vis besvarelse».

Du er sjølv ansvarleg for å fylle inn riktige kodar på eventuelle handteikningsark. Les derfor informasjonen på omslagsarka nøye. Eksamenskontoret kan ikkje garantere at ark som er feilaktig fyld ut vert lagd til besvarelsen din.

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

Vekting av oppgåvene: Maksimalt antall poeng av totalt 100 er angjeve for kvar oppgåve.

Trekk frå/avbrutt eksamen: Dersom du ønskjer å levera blankt/avbryte eksamen, gå til "hamburgermenyen" i øvre høgre hjørne og velg «Lever blankt». Dette kan ikkje angrast sjølv om prøven framleis er open.

Tilgang til svar: Etter eksamen finn du svara dine under tidlegare prøver i Inspera. Merk at det kan ta ein virkedag før eventuelle handteikninger vil vere tilgjengeleg under «tidligere prøver».

1 La C vere hendinga at ein person trekt tilfeldig frå populasjonen er smitta av covid-19 og la T vere hendinga at personen testar positivt på ein covid-19 hurtigtest. Anta at $P(C) = 0.03$, at testen sin sensitivitet $P(T|C) = 0.7$, og at testen sin spesifisitet $P(T'|C') = 0.99$ der A' er komplementet av hendinga A.

a) Kva vert $P(T|C')$?

b) Kva vert sannsynet for at personen testar positivt?

c) Kva vert $P(C|T)$?

d) Kva vert sannsynet for at personen ikkje er smitta av covid-19 gjeve at hurtigtesten gjer positivt resultat?

Maks poeng: 10

2 Vi trekkjer 5 kort frå ein vanleg kortstokk bestående av 52 kort.

a) Kor mange uordna utval kan vi trekkje frå kortstokken?

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

c) Kva er sannsynet for hus?

Maks poeng: 7.5

3 Anta at den stokastisk variabelen X er kontinuerleg fordelt med sannsynlighetstettleik

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

a) Rekn ut $E(X)$.

b) Rekn ut $E(X^2)$.

c) Rekn ut $\mathbf{Var}(X)$.

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

e) Rekn ut medianen til X.

Maks poeng: 12.5

4 Anta at Y er kontinuerleg fordelt med sannsynlighetstettleik

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

a) Kva er standardavviket til Y?

b) Kva er forvetningsverdien til Y?

c) Rekn ut $P(2 \leq Y \leq 4)$.

d) Rekn ut $P(Y = 3)$.

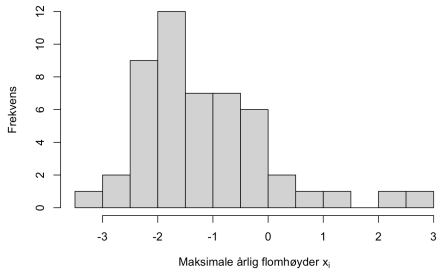
Maks poeng: 10

5 Anta at årleg maksimal flomhøgde X (i meter) relativt til eit fast målepunkt i eit vassdrag er fordelt slik at X har kumulativ fordeling $F_X(x) = e^{-e^{-x+\mu}}$ for $-\infty < x < \infty$ der μ er ein ukjend parameter.

a) Vi definerer ein hunderårsflom som den flomhøgda q som tilfredsstiller $P(X \geq q) = 1/100$. Bestem q om $\mu = -1.6$.

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

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



c) Vis at X har sannsynlighetstettleik $f_X(x) = e^{-x+\mu-e^{-x+\mu}}$. Utled sannsynlighetsmaksimeringsestimatoren av μ og berekn estimatet av μ for utvalet gjeve over.

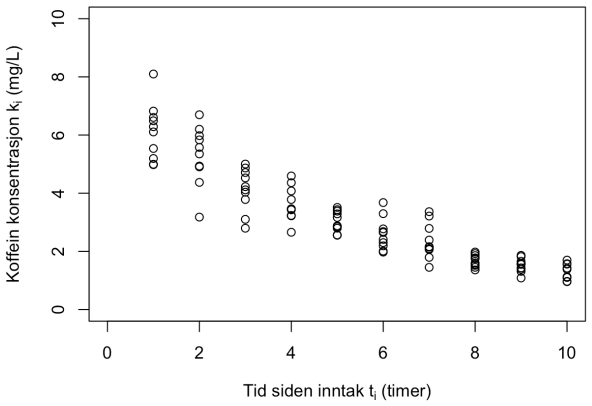
d) Vis at $Y = 2e^{-X+\mu}$ er khi-kvadratfordelt med 2 frihetsgrader. Er $2e^{\mu}\sum_{i=1}^{n} e^{-X_i}$ ein pivotal storleik? Utled eit 95%-konfidensintervall for μ og berekn intervallet for data gjeve over.

e) Finn eit tilsvarande konfidensintervall for hundreårs flomhøgde q.

Skriv svaret ditt på utdelte ark

Maks poeng: 30

6 Vi måler koffeinkonsentrasjonen k_i på ulike tidspunkt t_i i eit antall forsøkspersonar eit visst antall timar etter inntak av ei bestemd mengde koffein slik at vi tilsaman har $n = 100$ målinger som vist i følgjande figur.



Vi skal anta at koffeinkonsentrasjonen avtek eksponentielt slik at logaritmen til koffeinkonsentrasjonen avtek lineært med tida t_i. Vi kan da analysere data med regresjonsmodellen

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

der $\epsilon_i \sim N(0, \sigma^2)$ for $i = 1, 2, \ldots, n$ og anteke uavhengige. Det vert oppgjeve at $\sum_{i=1}^{n}(t_i - \bar{t})^2 = 825$, $\sum_{i=1}^{n}(t_i - \bar{t})\ln k_i = -143.6$, og at $\sum_{i=1}^{n}(\ln k_i - \hat{\beta}_0 - \hat{\beta}_1 t_i)^2 = 3.19$.

a) Berekn sannsynlighetsmaksimeringsestimatet $\hat{\beta}_1$ av stigningstallet β_1 (bruk uttrykk i formelsamlinga). Vis at

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

Rekn ut eit estimat av standardfeilen til $\hat{\beta}_1$.

b) Finn halveringstida $t_{1/2}$ til forventa koffeinkonsentrasjonen uttrykt ved parameterane i modellen over og berekn eit estimat $\hat{t}_{1/2}$ av halveringstida. Bruk feilforplantningsformelen til å berekne eit tilnærma estimat av standardfeilen til estimert halveringstid.

Skriv svaret ditt på utdelte ark

Maks poeng: 12

7 Frå historiske data meiner Banenor at det langs ein jernbanestrekning i gjennomsnitt skjer $\lambda_0 = 9$ viltpåkjørsler kvart år.

a) La stokastisk variabel X vere talet på viltpåkjørsel i løpet av $t = 2$ år. Under kva forutsetninger vil X være Poissonfordelt? Kva vert forventningsverdien og standardavviket til X? Rekn ut sannsynet for at det skjer minst like mange påkjørser som forventa i løpet av 2 år.

For å redusere antal påkjørsel byggjer ein viltovergang og etter 2 år ønskjer Banenor å undersøkje om viltovergangen har redusert antal påkjørslar. Det vert registrert tilsaman 12 påkjørsel i dei 2 åra etter at viltovergangen vert bygd.

b) Formuler problemet som ein hypotesetest. Kva vert null- og alternativ hypotese? Foreslå ein testobservator og bestem testen sin forkastingsregel, enten eksakt eller ved bruk av normaltilnærming, når signifikansnivået $\alpha = 0.05$. Kva blir testen sin konklusjon? Bestem også testen sin p-verdi gjeve data over.

c) Kva blir teststyrken til testen over om det i realiteten er slik at $\lambda = 5$ etter at viltovergangen vert etablert. Kor liten λ må ein tilnærma ha for at teststyrken skal vera minst 0.95?

Skriv svaret ditt på utdelte ark

Maks poeng: 18