Institutt for matematiske fag

Eksamensoppgave i TMA4245 Statistikk

Eksamensdato: 15.05.2020

Eksamenstid (fra-til): 09:00–13:00

Hjelpemiddelkode/Tillatte hjelpemidler: A / Alle hjelpemidler tillatt

Faglig kontakt under eksamen: Geir-Arne Fuglstad og Thea Bjørnland
Tlf.: 45 27 08 06 og 41 12 38 49

Teknisk hjelp under eksamen: NTNU Orakel
Tlf: 73 59 16 00

ANNEN INFORMASJON:

Gjør dine egne antagelser og presiser i besvarelsen hvilke forutsetninger du har lagt til grunn i tolkning/avgrensing av oppgaven. Faglig kontaktperson skal kun kontaktes dersom det er direkte feil eller mangler i oppgavesettet.

Lagring: Besvarelsen din i Inspera Assessment lagres automatisk. Jobber du i andre programmer – husk å lagre underveis.

Juks/plagiat: Eksamen skal være et individuelt, selvstendig arbeid. Det er tillatt å bruke hjelpemidler. Alle besvarelser blir kontrollert for plagiat. *Du kan lese mer om juks og plagiering på eksamen her*.

Varslinger: Hvis det oppstår behov for å gi beskjeder til kandidatene underveis i eksamen (f.eks. ved feil i oppgavesettet), vil dette bli gjort via varslinger i Inspera. Et varsel vil dukke opp som en dialogboks på skjermen i Inspera. Du kan finne igjen varselet ved å klikke på bjella øverst i høyre hjørne på skjermen. Det vil i tillegg bli sendt SMS til alle kandidater for å sikre at ingen går glipp av viktig informasjon. Ha mobiltelefonen din tilgjengelig.

Vekting av oppgavene: Det er 20 deloppgaver som teller likt ved sensur.

Fritekstoppgaver: Følg instruksene og beskriv med egne ord de stegene du har tatt for å komme frem til svaret.

OM LEVERING:

Besvarelsen din leveres automatisk når eksamenstida er ute og prøven stenger, forutsatt at minst én oppgave er besvart. Dette skjer selv om du ikke har klikket «Lever og gå tilbake til Dashboard» på siste side i oppgavesettet. Du kan gjenåpne og redigere besvarelsen din så lenge prøven er åpen. Dersom ingen oppgaver er besvart ved prøveslutt, blir ikke besvarelsen din levert.

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

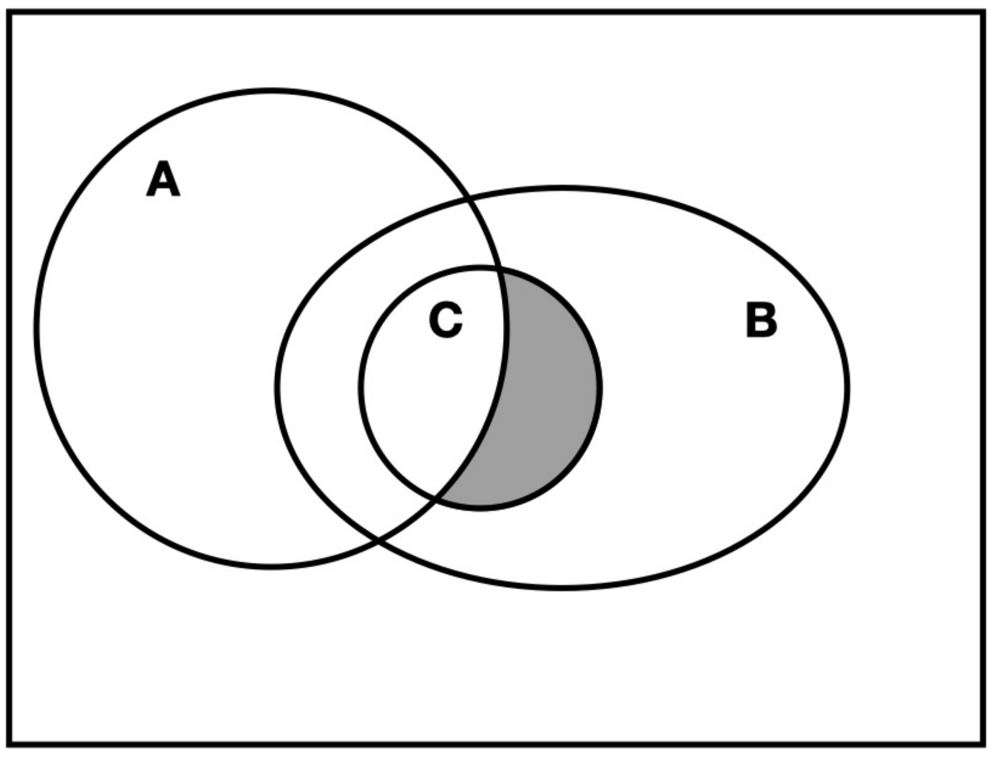
Tilgang til besvarelse: Du finner besvarelsen din i Arkiv etter at sluttida for eksamen er passert.

1

Oppgave 1

Hendelser og sannsynlighet

(a)



Hvilken av disse hendelsene svarer til det skraverte området i venndiagrammet?
Velg ett alternativ

- $C \cap A'$
- $B \cap C$
- $(B \cup C) \cap A$
- $B \cap A'$

Maks poeng: 5

(b) Det er 100 personer tilstede på en forelesing. Av de 100 personene er 55 kvinner, 60 er økonomistudenter og 20 er kvinnelige økonomistudenter. Anta at vi velger en person tilfeldig. La hendelsen A være at personen er en kvinne, og la hendelsen B være at personen er en økonomistudent. Hva er P(B|A)?
Velg ett alternativ

- 0.600
- 0.200
- 0.333
- 0.364

Maks poeng: 5

(c) En urne inneholder 20 kuler. Av de 20 kulene er 14 blå og 6 røde. Vi skal trekke tre kuler fra urnen *uten* tilbakelegging og telle hvor mange kuler vi har trukket av hver farge. Hva er sannsynligheten for at vi trekker to blå kuler og én rød kule?

Velg ett alternativ

- 0.479
- 0.147
- 0.160
- 0.441

Maks poeng: 5

(d) En urne inneholder 20 kuler. Av de 20 kulene er 14 blå og 6 røde. Vi skal trekke tre kuler fra urnen *med* tilbakelegging og telle hvor mange kuler vi har trukket av hver farge. Hva er sannsynligheten for at vi trekker to blå kuler og én rød kule?

Velg ett alternativ

- 0.160
- 0.441
- 0.479
- 0.147

Maks poeng: 5

2

Oppgave 2

Vi har to stokastiske variabler, X_1 og X_2, der

- X_1 har forventningsverdi $\mu_1 = 1$ og varians $\sigma_1^2 = 2$.
- X_2 har forventningsverdi $\mu_2 = 1$ og varians $\sigma_2^2 = 3$.

I tillegg er kovariansen mellom X_1 og X_2 gitt ved $\text{Cov}[X_1, X_2] = 1$.
La $Y = \frac{1}{3}(2X_1 - 5X_2)$.

(a) Regn ut forventningsverdien til Y.

Du skal bare gi tallsvaret. E[Y] = ☐.

Maks poeng: 5

(b) Regn ut variansen til Y.

Du skal bare gi tallsvaret. Var[Y] = ☐.

Maks poeng: 5

3

Oppgave 3

En frisørsalong tar imot "drop-in" kunder i åpningstiden 09:00–17:00.

Eieren av salongen ønsker å modellere ankomst av kunder, og antar at kunder ankommer frisørsalongen ifølge en poissonprosess med rate λ.

Basert på erfaring vet eieren at det i gjennomsnitt kommer omtrent $\lambda = 2$ kunder per time.

(a) Er det grunn til å forvente færre kunder mellom klokka 09:00 og 10:00, enn mellom klokka 12:00 og 13:00? Begrunn svaret ditt ved å ta utgangspunkt i egenskapene til prosessen.

Skriv ditt svar her...

Maks poeng: 5

(b) En mandag morgen klokka 10:00 har det ennå ikke kommet noen kunder. Hva er sannsynligheten for at det kommer minst 4 kunder før klokken 12:00?

Velg ett alternativ

- 0.371
- 0.715
- 0.849
- 0.567

Maks poeng: 5

4

Oppgave 4

La X være en kontinuerlig stokastisk variabel med kumulativ fordelingsfunksjon

$$F(x) = \begin{cases} 1 - \left(\frac{3}{x}\right)^2, & \text{for } x \geq 3, \\ 0, & \text{ellers} \end{cases}$$

(a) Hva er $\mathrm{P}(3 < X \leq 5)$?

Velg ett alternativ

- 0.640
- 0.203
- 0.360
- 0.400

Maks poeng: 5

(b) Hva er $\mathrm{P}(X > 6 | X > 4)$?

Velg ett alternativ

- 0.338
- 0.250
- 0.444
- 0.313

Maks poeng: 5

5

Oppgave 5

Norge har et omfattende barnevaksinasjonsprogram som dekker en stor andel av befolkningen. En av vaksinene som inngår i vaksinasjonsprogrammet er MMR-vaksinen. Denne vaksinen beskytter blant annet mot meslinger og vi tror derfor at en stor andel av Norges befolkning er immune mot meslinger.

En undersøkelse fra 2016 viste at 91% av alle 16-åringer i Norge hadde tatt MMR-vaksinen. Det er viktig at en høy vaksinasjonsgrad opprettholdes slik at vi ikke risikerer et epidemisk utbrudd med ukontrollert smittespredning. I 2020 vil forskere derfor gjøre en ny undersøkelse for å bestemme om andelen vaksinerte 16-åringer har blitt lavere enn 0.91. De ønsker å utføre en hypotesetest med signifikansnivå $\alpha = 0.05$.

Forskerne gjør et tilfeldig utvalg av 100 16-åringer, og for hver 16-åring registrerer de om han/hun er vaksinert eller ikke vaksinert. La $X_1, X_2, \ldots, X_{100}$ angi vaksinasjonsstatus for de 100 tilfeldig valgte 16-åringene slik at

$$X_i = \begin{cases} 1, & \text{vaksinert} \\ 0, & \text{ikke vaksinert} \end{cases}, \qquad i = 1, 2, \ldots, 100.$$

(a) Hvilken fordeling kan vi anta for hver X_i ($i = 1, 2, \ldots, 100$), og hva beskriver parameteren i fordelingen?

Skriv ditt svar her...

Maks poeng: 5

(b) Vi ønsker å undersøke om andelen vaksinerte 16-åringer har blitt lavere enn 0.91. Hvilken nullhypotese, H_0, og alternativ hypotese, H_1, er best egnet for dette?

Skriv ditt svar her...

Maks poeng: 5

(c) Vi tar utgangspunkt i observatoren $\hat{P} = \frac{1}{100}\sum_{i=1}^{100} X_i$. Dersom nullhypotesen er sann, hva er den tilnærmede fordelingen til denne observatoren? Hva er tallverdien(e) til parameteren(e) i denne fordelingen når nullhypotesen er sann?

Skriv ditt svar her...

Maks poeng: 5

(d) Forskerne observerer at 89 av 100 16-åringer er vaksinerte. Hva er konklusjonen av hypotesetesten? Oppgi hvilken testobservator og forkastningsregel du har brukt.

Skriv ditt svar her...

Maks poeng: 5

6

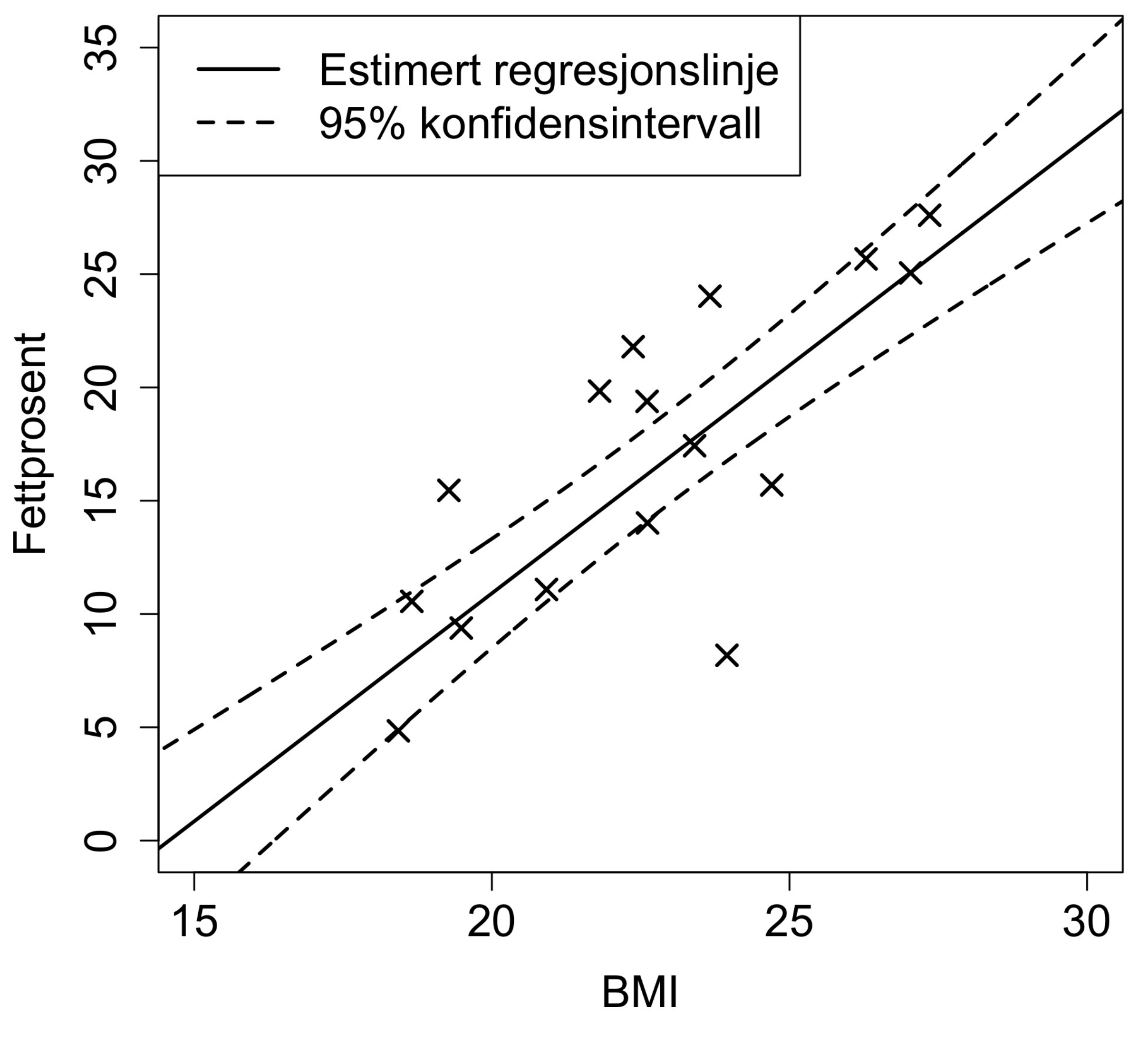
Oppgave 6

For n tilfeldig valgte personer skal vi modellere sammenhengen mellom fettprosent og BMI (body mass index) ved

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

der Y_i og x_i er henholdsvis fettprosent og BMI for person i, β_0 og β_1 er regresjonskoeffisienter, og residualene $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ er identiske og uavhengige normalfordelte med forventningsverdi 0 og varians σ^2.

Vi gjør målinger for $n = 20$ personer og får verdiene som er vist i figuren under. Kryssene viser målingene, den heltrukne linjen viser den estimerte regresjonslinjen, og de stiplede linjene viser nedre og øvre grenser for punktvise 95% konfidensintervaller for regresjonslinjen.



(a) Diskuter om antagelsene i den lineære regresjonsmodellen virker å være tilfredstilte for målingene vist i figuren.

Skriv ditt svar her...

Maks poeng: 5

(b) Minste-kvadraters-metode gir estimatene $b_0 = -29.3$ og $b_1 = 2.02$ for henholdsvis β_0 og β_1.

Hva blir predikert fettprosent for en person med BMI = 20, og hva blir predikert fettprosent for en person med BMI = 30?

Forklar hvorfor vi er mer usikre på prediksjonen av fettprosent for personer med BMI = 30, enn for personer med BMI = 20.

Skriv ditt svar her...

Maks poeng: 5

7

Oppgave 7

La $X_1, X_2, \ldots, X_{10}$ være et tilfeldig utvalg fra en normal-populasjon (Populasjon 1) med forventningsverdi μ_1 og kjent varians $\sigma^2 = 4$, og la $Y_1, Y_2, \ldots, Y_{25}$ være et tilfeldig utvalg fra en normal-populasjon (Populasjon 2) med forventningsverdi μ_2 og kjent varians $\sigma^2 = 4$. De to utvalgene er uavhengige.

For å estimere differansen $d = \mu_1 - \mu_2$ ser vi på følgende estimatorer:

1. $\hat{d} = \frac{1}{10}\sum_{i=1}^{10} X_i - \frac{1}{25}\sum_{i=1}^{25} Y_i$
2. $d^* = \frac{1}{10}\sum_{i=1}^{10}(X_i - Y_i)$
3. $\tilde{d} = \frac{1}{10}\sum_{i=1}^{10}(X_i - Y_i) - \frac{1}{15}\sum_{i=11}^{25} Y_i$

(a) Hvilke(n) av de tre estimatorene er forventningsrett(e) for d?

Velg ett alternativ

- Bare d^*
- Alle
- Bare $\hat{d}$ og d^*
- Bare $\hat{d}$ og $\tilde{d}$

Maks poeng: 5

(b) Hvilken av estimatorene foretrekker vi? Begrunn svaret ditt.

Skriv ditt svar her...

Maks poeng: 5

(c) Vi har gjort ti observasjoner $x_1, x_2, \ldots, x_{10}$ fra Populasjon 1 og finner $\sum_{i=1}^{10} x_i = 25.92$. Videre har vi gjort 25 observasjoner $y_1, y_2, \ldots, y_{25}$ fra Populasjon 2 og finner $\sum_{i=1}^{25} y_i = 16.92$.

Ta utgangspunkt i estimatoren $\hat{d}$ og regn ut et 95% konfidensintervall for $d = \mu_1 - \mu_2$. Oppgi tallsvar, og forklar de ulike stegene du brukte for å regne ut konfidensintervallet.

Skriv ditt svar her...

Maks poeng: 5

(d) For å finne et mer presist estimat av $d = \mu_1 - \mu_2$ får vi muligheten til å gjøre ytterlige 15 observasjoner. Diskuter hvordan vi bør fordele disse 15 nye observasjonene mellom Populasjon 1 og Populasjon 2 for å få et smalest mulig konfidensintervall.

Skriv ditt svar her...

Maks poeng: 5