Institutt for matematiske fag, NTNU

Eksamensoppgave i **TMA4240 Statistikk**

Faglig kontakt under eksamen: Sara Martino og Håkon Tjelmeland
Tlf: 44903330, 48221896

Eksamensdato: 30. november 2019
Eksamenstid (fra-til): 09:00 - 13:00
Hjelpemiddelkode/Tillatte hjelpemidler: Hjelpemiddelkode C.

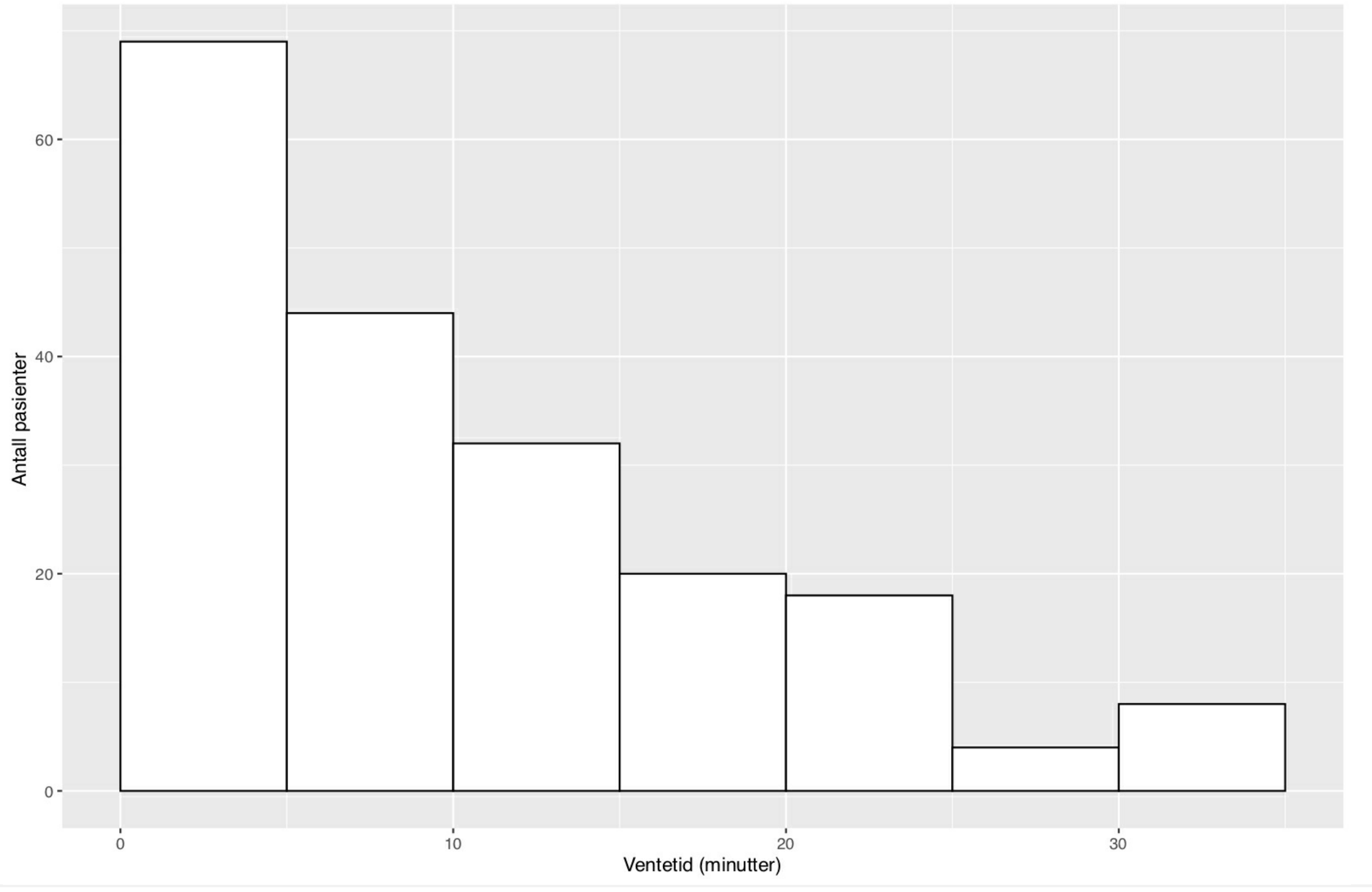
- *Tabeller og formler i statistikk* (Fagbokforlaget)
- Et gult ark (A5 med stempel) med egne håndskrevne formler og notater
- Bestemt, enkel kalkulator

Annen informasjon:

- For oppgavene som ikke er flervalgsoppgaver, må alle svar begrunnes og besvarelsen må inneholde mellomregning slik at det er helt klart hvordan man har tenkt.
- Vekt ved sensur for hver deloppgave er angitt i oppgavesettet.

Målform/språk: Bokmål

1(a)



Flervalgsoppgaver

Innledning: Histogrammet over viser frekvensfordelingen for ventetider for 200 pasienter på et tannlegekontor.

Oppgave: Hva er (omtrentlig) medianen av disse ventetidene?

Velg ett alternativ

- 2.50
- 7.25
- 12.25
- 15.00
- 17.50

Vekt ved sensur: 4 %.

Maks poeng: 4

1(b) **Innledning**: Du skal utføre en ensidig hypotesetest hvor nullhypotesen er at forventningsverdien er lik 532 og alternativ hypotese er at forventningsverdien er mindre enn 532. Anta at gjennomsnittet av dine observasjoner er 529 og at p-verdien er 0.01.

Oppgave: Hvilke av følgende utsagn er da korrekte?

Velg ett eller flere alternativer

- Sannsynligheten for at forventningsverdien er mindre enn 529 er 0.01.
- Hvis forventningsverdien er 532 er sannsynligheten for å observere et gjennomsnitt som er mindre enn eller lik 529 lik 0.01.
- Sannsynligheten for at forventningsverdien er mindre enn eller lik 532 er 0.01.
- Hvis signifikansnivået er 0.05, vil man ikke forkaste H_0.
- Ingen av utsagnene over er korrekte.

Vekt ved sensur: 4 %.

Maks poeng: 4

1(c) **Innledning**: En produsent av ballonger hevder at andelen, p, av ballongene som sprekker når de blåses opp til å ha en diameter på $30\ cm$, er mindre enn 0.05. En del kunder har klaget på at andelen av ballongene som sprekker er høyere enn han hevder.

Oppgave: Dersom disse kundene ønsker å gjennomføre en undersøkelse for å teste produsentens påstand, hvilke av følgende hypoteser vil være riktige?

Velg ett eller flere alternativer

- $H_0: p \neq 0.05$ mot $H_1: p = 0.05$.
- $H_0: p = 0.05$ mot $H_1: p \neq 0.05$.
- $H_0: p = 0.05$ mot $H_1: p < 0.05$.
- $H_0: p = 0.05$ mot $H_1: p > 0.05$.
- $H_0: p < 0.05$ mot $H_1: p = 0.05$.

Vekt ved sensur: 4 %.

Maks poeng: 4

1(d) **Innledning**: Anta at man i en undersøkelse fant et 95 %-konfidensintervall for andelen av nordmenn som trener regelmessig til å gå fra 0.29 til 0.37.

Oppgave: Hvilke av følgende utsagn er da FEIL?

Velg ett eller flere alternativer

- Det er rimelig å si at mer enn 25 % av den norske befolkningen trener regelmessig.
- Det er rimelig å si at mer enn 40 % av den norske befolkningen trener regelmessig.
- En hypotese om at mer enn 33 % av den norske befolkningen trener regelmessig kan ikke forkastes.
- Det er rimelig å si at mindre enn 40 % av den norske befolkningen trener regelmessig.

Vekt ved sensur: 4 %.

Maks poeng: 4

1(e) **Innledning**: La X og Y være to uavhengige stokastiske variabler som begge er geometrisk fordelt med parameter p, dvs

$$X \sim g(x;p) = p(1-p)^{x-1}; x = 1,2,\dots \text{ og } Y \sim g(y;p) = p(1-p)^{y-1}; y = 1,2,\dots$$

La $Z = \min(X,Y)$. Du kan i denne oppgaven få bruk for formelen for en geometrisk rekke,

$$\sum_{k=0}^{n} a^k = \frac{1-a^{n+1}}{1-a}$$

Oppgave: Hva er punktsannsynligheten $f_Z(z)$ for Z?

Velg ett alternativ

- $f_Z(z) = (1-p)^{2(z-1)}p(2-p); z = 1,2,\dots$
- $f_Z(z) = 1-(1-p)^{2z}; z = 1,2,\dots$
- $f_Z(z) = (1-(1-p)^z)^2; z = 1,2,\dots$
- $f_Z(z) = p^2(1-p)^{2(z-1)}; z = 1,2,\dots$
- $f_Z(z) = -2(1-p)^{2z}\ln(1-p); z = 1,2,\dots$

Vekt ved sensur: 4 %.

Maks poeng: 4

2 Normalfordeling

Innledning: La X og Y være to uavhengige og normalfordelte stokastiske variabler. Anta at X har forventningsverdi lik 0 og standardavvik lik 2, mens Y har forventningsverdi lik 1 og standardavvik lik 1.

Oppgave:

- Skisser sannsynlighetstetthetene til X og Y i et felles plott.
- Finn følgende sannsynligheter,

$$P(X \leq 1), \quad P(Y \geq -1) \quad \text{og} \quad P(X - Y \leq 0)$$

Skriv ditt svar her...

Format

Words: 0

Vekt ved sensur: 10 %.

Maks poeng: 10

3(a) Sigdcelleanemi

Innledning: Sigdcelleanemi er en alvorlig sykdom som forårsaker anemi, dvs lav blodprosent. Sigdcelleanemi er arvelig, og et barn får sykdommen dersom det arver et bestemt recessivt gen (a) fra både mor og far. Barnet får ikke sykdommen dersom det arver det dominante genet (A) fra minst en av sine foreldre.

En person har altså enten genotype AA, Aa eller aa. Personer med genotype aa har sigdcelleanemi, mens personer med AA eller Aa har ikke sykdommen. Et barn arver et gen fra hver av sine foreldre. Hvis en forelder har genotype Aa, vil et barn arve enten a eller A fra denne forelderen med sannsynlighet 0.5 for hver. Et barn arver gen fra mor og far uavhengig av hverandre.

Man sier at personer som har genotype Aa er bærere av sykdommen, de har ikke selv sigdcelleanemi men kan få barn som har sykdommen.

Betrakt nå et par hvor hverken mannen eller kvinnen har sigdcelleanemi, men vi vet ikke om de er bærere av sykdommen. Anta at mannen og kvinnen hver har en sannsynlighet på $8\ \%$ for å være bærere av sykdommen (dette er andelen av den afroamerikanske befolkningen som er bærere) og at mannen og kvinnen er bærere eller ikke uavhengig av hverandre.

Vi skal i denne oppgaven regne på sannsynligheten for at dette paret får barn som har sigdcelleanemi eller som er bærere av sykdommen. Definer hendelsene

- M: Mannen er bærer av sigdcelleanemi,
- K: Kvinnen er bærer av sigdcelleanemi,
- D: Det første barnet paret får, har sigdcelleanemi, og
- B: Det første barnet paret får, er bærer av sigdcelleanemi.

Oppgave:

- Tegn opp hendelsene M, K, D og B i et venndiagram.
- Regn ut sannsynligheten for at det første barnet paret får vil ha sigdcelleanemi.
- Regn ut sannsynligheten for at det første barnet paret får vil være bærer av sigdcelleanemi.

Skriv ditt svar her...

Format B I U x₂ x² Ω Σ

Words: 0

Vekt ved sensur: 10 %.

Maks poeng: 10

3(b) **Innledning**: Vi skal så anta at paret allerede har fått et barn som ikke har sigdcelleanemi, og at de nå planlegger å få et barn til.

Oppgave:

- Hva er da sannsynligheten for at det andre barnet vil ha sigdcelleanemi?
- Hva er da sannsynligheten for at det andre barnet vil være bærer av sigdcelleanemi?

Skriv ditt svar her...

Format

Words: 0

Vekt ved sensur: 10 %.

Maks poeng: 10

4(a) Binomisk fordelte variabler

Innledning: Anta at vi har to uavhengige stokastiske variabler X og Y, der $X \sim b(x; n, p)$ og $Y \sim b(y; n, 2p)$. X er altså antall suksesser i n uavhengige forsøk der hvert forsøk har sannsynlighet p for å gi suksess, mens Y er antall suksesser i n andre uavhengige forsøk der hvert forsøk gir suksess med sannsynlighet $2p$.

Oppgave:

- Når $n = 12$ og $p = 0.2$, finn følgende sannsynligheter,

$$P(X \leq 3),\ \ P(Y \geq 4 | X \leq 3) \ \text{ og } \ P(X + Y \leq 1).$$

Skriv ditt svar her...

Format B I U x₂ x² Ω Σ

Words: 0

Vekt ved sensur: 10 %.

Maks poeng: 10

4(b) **Innledning**: I resten av oppgaven skal vi anta at verdien av parameteren $p \in [0, 0.5]$ er ukjent og skal estimeres basert på X og Y. Følgende tre estimatorer er foreslått,

$$\hat{p} = \frac{X+Y}{2n}, \;\; \tilde{p} = \frac{X+Y}{3n} \quad \text{og} \quad p^{\star} = \frac{X}{2n} + \frac{Y}{4n} \tag{1}$$

Oppgave:

- Hvilken av de tre estimatorene vil du foretrekke? Begrunn svaret.

Skriv ditt svar her...

Format B I U x_2 x^2 I_x Ω Σ

Words: 0

Vekt ved sensur: 10 %.

Maks poeng: 10

4(c) **Innledning**: Anta at vi også ønsker å estimere p ved å benytte sannsynlighetsmaksimeringsprinsippet, fremdeles basert på X og Y.

Oppgave:

- Finn et uttrykk for log-rimelighetsfunksjonen for p, $l(p)$.
- Bestem hva estimatet for p blir ved å benytte sannsynlighetsmaksimeringsprinsippet når $n = 25$, og man observerer $x = 3$ og $y = 8$.

Skriv ditt svar her...

Format

Words: 0

Vekt ved sensur: 10 %.

Maks poeng: 10

4(d) **Innledning**: Anta så at vi ønsker å benytte X og Y til å teste

$$H_0 : p = 0.2 \quad \text{mot} \quad H_1 : p > 0.2.$$

Vi skal i resten av oppgaven anta at n er stor nok til at man med god approksimasjon kan anta at X og Y er normalfordelte.

Oppgave:

- Ta utgangspunkt i den av de tre estimatorene i (1) du fant var best, velg en testobservator og bestem et (tilnærmet) forkastningskriterium når signifikansnivået velges som $\alpha = 0.05$. Hvis du ikke konkluderte med hvilken av de tre estimatorene i (1) som var best, kan du selv velge hvilken av estimatorene du vil ta utgangspunkt i.

Skriv ditt svar her...

Format

Words: 0

Vekt ved sensur: 10 %.

Maks poeng: 10

4(e) **Oppgave**:

- Hvis man bruker signifikansnivå $\alpha = 0.05$ i hypotesetesten over, hvor stor må n velges for at man skal ha minst sannsynlighet 0.9 for å oppdage at H_0 er feil når $p = 0.25$?

Skriv ditt svar her...

Format

Words: 0

Vekt ved sensur: 10 %.

Maks poeng: 10