NTNU
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **TMA4240 Statistikk – løsningsskisse**

Academic contact during examination:

Phone:

Examination date: August 2021

Examination time (from–to): 09:00–13:00

Permitted examination support material:

Other information:

Language: English

Number of pages: 13

Number of pages enclosed: 0

Checked by:

Informasjon om trykking av eksamensoppgave
Originalen er:
1-sidig ☐ **2-sidig** ☒
sort/hvit ☒ **farger** ☐
skal ha flervalgskjema ☒

Date Signature

NB: Merk at det ble gitt ulike varianter av oppgavene. I denne løsningsskissen gis det kun detaljert utregning for variant A av hver oppgave. De øvrige oppgavevariantene løses tilsvarende. For disse gis det derfor her kun fasitsvar.

Problem 1

- Utvalgets gjennomsnitt: $\frac{1}{5}(-1+2+3+7+9) = \underline{\underline{4.00}}$
- Utvalgsvarians (empirisk varians):

$$\frac{1}{5-1}((-1-4)^2 + (2-4)^2 + (3-4)^2 + (7-4)^2 + (9-4)^2) = \underline{\underline{16.00}}$$

- Median: $\underline{\underline{3.00}}$

B: 14, 16, 13

C: 7, 37.5, 10

D: 5, 11, 5

Problem 2

- $P(B \cap C) = P(B)P(C|B) = 0.2 \cdot 0.5 = \underline{\underline{0.1}}$
- $P(C|A) = \frac{P(A \cap C)}{P(A)} = \frac{P(A|C)P(C)}{P(A)} \Rightarrow P(C) = \frac{P(A)P(C|A)}{P(A|C)} = \frac{0.08 \cdot 0.25}{0.1} = \underline{\underline{0.20}}$
- $P(A \cap C) = P(A)P(C|A) = 0.08 \cdot 0.25 = 0.02.$

$$P(B|A \cup C) = \frac{P(B \cap (A \cup C))}{P(A \cup C)} = \frac{P(B \cap C)}{P(A) + P(C) - P(A \cap C)} = \frac{0.1}{0.08 + 0.2 - 0.02} = \underline{\underline{0.385}}$$

B: 0.04, 0.50, 0.05

C: 0.09, 0.125, 0.257

D: 0.10, 0.05, 0.690

Problem 3

- $16^4 \cdot 4 = \underline{\underline{262144}}$
- $\binom{6}{3} \cdot 16^3 \cdot 4^3 \cdot 4 = \underline{\underline{20971520}}$

B: 1856465, 974644125

C: 24576, 622080

Problem 4

- $P(X > 1) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) = 0.10 + 0.25 + 0.20 + 0.25 = \underline{\underline{0.800}}$
- $P(X \leq 4 | X \geq 3) = \frac{P(X \leq 4 \cap X \geq 3)}{P(X \geq 3)} = \frac{P(X=3)+P(X=4)}{P(X=3)+P(X=4)+P(X=5)} = \frac{0.25+0.20}{0.25+0.20+0.25} = \underline{\underline{0.643}}$
- $\mathrm{E}[X] = 0 \cdot 0.05 + 1 \cdot 0.15 + 2 \cdot 0.10 + 3 \cdot 0.25 + 4 \cdot 0.20 + 5 \cdot 0.25 = \underline{\underline{3.150}}$

B: 0.75, 0.727, 2.65

C: 0.70, 0.737, 3.15

D: 0.55, 0.833, 2.65

Problem 5

I en eksponensialfordeling har vi at $\mathrm{Var}[X] = \beta^2 = \frac{1}{\lambda^2}$ slik at vi her får $\beta = \frac{1}{3}$, eventuelt $\lambda = 3$. Dessuten vet vi at kumulativ fordelingsfunksjon i en eksponensialfordeling er gitt ved

$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{ellers.} \end{cases}$$

- $\mathrm{E}[X] = \beta = \frac{1}{\lambda} = \underline{\underline{0.333}}$

- $P(X > 1) = 1 - P(X \leq 1) = 1 - (1 - e^{-\lambda \cdot 1}) = e^{-3} = \underline{\underline{0.050}}$
- $P(X \leq 0.5 | X < 1.1) = \frac{P(X \leq 0.5 \cap X < 1.1)}{P(X < 1.1)} = \frac{P(X \leq 0.5)}{P(X \leq 1.1)} = \frac{1 - e^{-3 \cdot 0.5}}{1 - e^{-3 \cdot 1.1}} = \underline{\underline{0.807}}$

B: 0.25, 0.0183, 0.875

C: 0.5, 0.135, 0.711

Problem 6

Vi har oppgitt at

$$\mathrm{E}[X] = \mu, \quad \mathrm{Var}[X] = \sigma^2, \quad \mathrm{E}[Y] = 2\mu \quad \text{og} \quad \mathrm{Var}[Y] = 4\sigma^2.$$

Vi starter med å sjekke hvilke estimatorer som er forventningsrette ved å regne ut forventningsverdien til hver estimator.

$$\begin{aligned}
\mathrm{E}[\widehat{\mu}] &= \mathrm{E}\left[\frac{X+Y}{2}\right] = \frac{1}{2}\mathrm{E}[X+Y] = \frac{1}{2}\left(\mathrm{E}[X] + \mathrm{E}[Y]\right) = \frac{1}{2}(\mu + 2\mu) = \frac{3}{2}\mu \\
\mathrm{E}[\mu^\star] &= \mathrm{E}\left[\frac{2X+Y}{4}\right] = \frac{1}{4}\mathrm{E}[2X+Y] = \frac{1}{4}\left(2\mathrm{E}[X] + \mathrm{Y}\right) = \frac{1}{4}(2\mu + 2\mu) = \mu \\
\mathrm{E}[\widetilde{\mu}] &= \mathrm{E}\left[\frac{3X+Y}{5}\right] = \frac{1}{5}\mathrm{E}[3X+Y] = \frac{1}{5}\left(3\mathrm{E}[X] + \mathrm{E}[Y]\right) = \frac{1}{5}(3\mu + 2\mu) = \mu
\end{aligned}$$

Vi ser at $\mu^\star$ and $\widetilde{\mu}$ er forventningsrette. Finner derfor variansen til disse to for å bestemme hvilken av disse som er mest effisient.

$$\begin{aligned}
\mathrm{Var}[\mu^\star] &= \mathrm{Var}\left[\frac{2X+Y}{4}\right] = \left(\frac{1}{4}\right)^2 \mathrm{Var}[2X+Y] = \frac{1}{16}\left(\mathrm{Var}[2X] + \mathrm{Var}[Y]\right) \\
&= \frac{1}{16}\left(2^2\mathrm{Var}[X] + \mathrm{Var}[Y]\right) = \frac{1}{16}\left(4\sigma^2 + 4\sigma^2\right) = \frac{1}{2}\sigma^2 \\
\mathrm{Var}[\widetilde{\mu}] &= \mathrm{Var}\left[\frac{3X+Y}{5}\right] = \left(\frac{1}{5}\right)^2 \mathrm{Var}[3X+Y] \\
&= \frac{1}{25}\left(\mathrm{Var}[3X] + \mathrm{Var}[Y]\right) = \frac{1}{25}\left(3^2\mathrm{Var}[X] + \mathrm{Var}[Y]\right) = \frac{1}{25}\left(9\sigma^2 + 4\sigma^2\right) = \frac{13}{25}\sigma^2
\end{aligned}$$

Siden $\frac{13}{25} > \frac{1}{2}$ får vi at $\mu^\star$ er mest effisient. Så av de tre estimatorene $\underline{\underline{\text{vil vi foretrekke } \mu^\star}}$.

B: Foretrekker $\widehat{\mu}$.

C: Foretrekker $\widetilde{\mu}$.

Problem 7

Vi finner først marginalfordelingen for X. For $x > 0$ får vi

$$\begin{aligned}
f(x) &= \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\infty}^{\infty} \frac{1}{4}\exp\left\{-\left[x + \frac{1}{2}|y-x|\right]\right\} dy \\
&= \frac{1}{4}e^{-x}\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}|y-x|\right\} dy \\
&= \frac{1}{4}e^{-x}\left[\int_{-\infty}^{x} \exp\left\{-\frac{1}{2}|y-x|\right\} dy + \int_{x}^{\infty} \exp\left\{-\frac{1}{2}|y-x|\right\} dy\right] \\
&= \frac{1}{4}e^{-x}\left[\int_{-\infty}^{x} \exp\left\{\frac{1}{2}(y-x)\right\} dy + \int_{x}^{\infty} \exp\left\{-\frac{1}{2}(y-x)\right\} dy\right] \\
&= \frac{1}{4}e^{-x}\left[\left[2\exp\left\{\frac{1}{2}(y-x)\right\}\right]_{y=-\infty}^{y=x} + \left[-2\exp\left\{-\frac{1}{2}(y-x)\right\}\right]_{y=x}^{y=\infty}\right] \\
&= \frac{1}{4}e^{-x}\left[(2e^0 - 0) + (0 - (-2e^0))\right] \\
&= \frac{1}{4}e^{-x}\cdot 4 = e^{-x}.
\end{aligned}$$

For $x \leq 0$ har vi åpenbart $f(x) = 0$, så marginalfordelingen til x blir

$$\underline{\underline{f(x) = \begin{cases} e^{-x} & \text{for } x > 0, \\ 0 & \text{ellers.} \end{cases}}}$$

Den betingede fordelingen for Y gitt $X = x$ blir da

$$\begin{aligned}
f(y|x) = \frac{f(x,y)}{f(x)} &= \frac{\frac{1}{4}\exp\left\{-\left[x + \frac{1}{2}|y-x|\right]\right\}}{\exp\{-x\}} \\
&= \underline{\underline{\frac{1}{4}\exp\left\{-\frac{1}{2}|y-x|\right\}}}
\end{aligned}$$

for $-\infty < y < \infty$.

B:

$$\begin{aligned}
f(x) &= \begin{cases} \frac{1}{2}e^{-\frac{1}{2}x} & \text{for } x > 0, \\ 0 & \text{ellers.} \end{cases} \\
f(y|x) &= \frac{1}{2}\exp\{-|y-x|\}
\end{aligned}$$

C:

$$f(x) = \begin{cases} 2e^{-2x} & \text{for } x > 0, \\ 0 & \text{ellers.} \end{cases}$$
$$f(y|x) = \exp\{-2|y-x|\}$$

Problem 8

For å finne sannsynlighetsmaksimeringsestimatoren for p starter vi med å etablere rimelighetsfunksjonen,

$$L(p) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \left[(1-p)^{x_i - 1} p\right] = p^n (1-p)^{\sum_{i=1}^{n} x_i - n}.$$

Log-rimelighetsfunksjon blir da

$$\ell(p) = \ln L(p) = n \ln p + \left(\sum_{i=1}^{n} x_i - n\right) \ln(1-p).$$

Deriverer $\ell(p)$ og setter den deriverte lik null for å bestemme for hvilken verdi av p log/rimelighetsfunksjonen har sitt maksimum,

$$\begin{aligned}
\ell'(p) &= \frac{n}{p} + \left(\sum_{i=1}^{n} x_i - n\right) \cdot \frac{1}{1-p} \cdot (-1) = 0 \\
&\Updownarrow \\
n(1-p) &= \left(\sum_{i=1}^{n} x_i - n\right) p \\
&\Updownarrow \\
n - np &= p \sum_{i=1}^{n} x_i - np \\
&\Updownarrow \\
p &= \frac{n}{\sum_{i=1}^{n} x_i}
\end{aligned}$$

Sannsynlighetsmaksimeringsestimatoren for p blir dermed

$$\underline{\underline{\widehat{p} = \frac{n}{\sum_{i=1}^{n} X_i}}},$$

og med de oppgitte observasjonene blir estimatet

$$\frac{6}{1+5+2+10+15+3} = \frac{6}{36} = \underline{\underline{\frac{1}{6}}}.$$

For å finne sannsynlighetsmaksimeringsestimatoren for $\theta = \frac{1}{p}$ kan vi argumentere med at sannsynlighetsmaksimeringsestimatoren er invariant under reparametrisering og dermed direkte få at sannsynlighetsmaksimeringsestimatoren for θ er

$$\underline{\underline{\widehat{\theta}}} = \frac{1}{\widehat{p}} = \frac{1}{\frac{n}{\sum_{i=1}^{n} X_i}} = \underline{\underline{\frac{1}{n}\sum_{i=1}^{n} X_i}}.$$

Dersom man ikke kjenner til at sannsynlighetsmaksimeringsestimatoren er invariant under reparametrisering kan man utlede sannsynlighetsmaksimeringsestimatoren for $\theta = \frac{1}{p} \Leftrightarrow p = \frac{1}{\theta}$ ved å starte med rimelighetsfunksjonen for θ. Denne rimelighetsfunksjonen blir

$$L(\theta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \left[\left(1 - \frac{1}{\theta}\right)^{x_i - 1} \frac{1}{\theta}\right] = \frac{1}{\theta^n}\left(1 - \frac{1}{\theta}\right)^{\sum_{i=1}^{n} x_i - n}.$$

Log-rimelighetsfunksjonen blir da

$$\ell(\theta) = -n \ln \theta + \left(\sum_{i=1}^{n} x_i - n\right) \ln\left(1 - \frac{1}{\theta}\right).$$

Deriverer så denne med hensyn på θ og setter den deriverte lik null for å finne for hvilken verdi av θ log-rimelighetsfunksjonen har sitt maksimum.

$$\begin{aligned}
\ell'(\theta) &= -\frac{n}{\theta} + \left(\sum_{i=1}^{n} x_i - n\right) \cdot \frac{1}{1 - \frac{1}{\theta}} \cdot \left(-\left(-\frac{1}{\theta^2}\right)\right) \\
&= -\frac{n}{\theta} + \left(\sum_{i=1}^{n} x_i - n\right) \cdot \frac{1}{\theta^2 - \theta} = 0 \\
&\Updownarrow \\
n(\theta^2 - \theta) &= \left(\sum_{i=1}^{n} x_i - n\right)\theta
\end{aligned}$$

Siden vi har at $p \in (0,1)$ må vi ha at $\theta = \frac{1}{p} > 1$. Vi er dermed kun interessert i løsninger av ligningen over hvor $\theta > 1$. Løsningen vi er interessert i må dermed ha

$$n(\theta - 1) = \left(\sum_{i=1}^{n} x_i - n\right) \Leftrightarrow \theta = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Sannsynlighetsmaksimeringsestimatoren for θ blir dermed

$$\underline{\underline{\widehat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i.}}$$

For en geometrisk fordeling vet vi at $\mathrm{E}[X_i] = \frac{1}{p} = \theta$. Dermed får vi at

$$\begin{aligned}
\mathrm{E}[\widehat{\theta}] &= \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\mathrm{E}\left[\sum_{i=1}^{n} X_i\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[X_i] \\
&= \frac{1}{n}\sum_{i=1}^{n} \theta \\
&= \frac{1}{n} \cdot n\theta = \theta.
\end{aligned}$$

Dermed ser vi at $\underline{\underline{\widehat{\theta}\text{ er forventningsrett}}}$.

B: $\widehat{p} = \frac{5}{54} = 0.0926$

C: $\widehat{p} = \frac{7}{30} = 0.2333$

Problem 9

For å bestemme for hvilke verdier av X ulikheten er oppfylt løser vi først den tilhørende andregradligningen,

$$\begin{aligned}
&X^2 - \frac{9X}{4} = -\frac{7}{8} \Leftrightarrow X^2 - \frac{9}{4}X + \frac{7}{8} = 0 \\
&X = \frac{-\left(-\frac{9}{4}\right) \pm \sqrt{\left(-\frac{9}{4}\right)^2 - 4 \cdot 1 \cdot \frac{7}{8}}}{2 \cdot 1} = \frac{\frac{9}{4} \pm \sqrt{\frac{81}{16} - \frac{28}{8}}}{2} \\
&= \frac{\frac{9}{4} \pm \sqrt{\frac{81-56}{16}}}{2} = \frac{\frac{9}{4} \pm \sqrt{\frac{25}{16}}}{2} \\
&= \frac{\frac{9}{4} \pm \frac{5}{4}}{2} \Leftrightarrow X = \frac{14}{8} = \frac{7}{4} = 1.75 \quad \vee \quad X = \frac{4}{8} = \frac{1}{2} = 0.5.
\end{aligned}$$

Ved å sette inn ser vi at ulikheten ikke er oppfylt for, for eksempel, $X = 1$,

$$X^2 - \frac{9X}{4} = 1^2 - \frac{9 \cdot 1}{4} = -\frac{5}{4} \not> -\frac{7}{8}.$$

Vi kan dermed konkludere med at ulikheten er oppfylt for $X < \frac{1}{2}$ og for $X > \frac{7}{4}$. Dermed får vi at sannsynligheten det spørres om er gitt ved

$$\begin{aligned}
P\left(X^2 - \frac{9X}{4} > -\frac{7}{8}\right) &= P\left(X < \frac{1}{2}\right) + P\left(X > \frac{7}{4}\right) \\
&= \int_{-\infty}^{\frac{1}{2}} f(x)dx + \int_{\frac{7}{4}}^{\infty} f(x)dx \\
&= \int_{-2}^{\frac{1}{2}} \frac{1}{12}(x+3)dx + \int_{\frac{7}{4}}^{2} \frac{1}{12}(x+3)dx \\
&= \frac{1}{12}\left(\left[\frac{1}{2}x^2 + 3x\right]_{-2}^{\frac{1}{2}} + \left[\frac{1}{2}x^2 + 3x\right]_{\frac{7}{4}}^{2}\right) \\
&= \frac{1}{12}\left(\frac{1}{8} + \frac{3}{2} - 2 + 6 + 2 + 6 - \frac{49}{32} - \frac{21}{4}\right) \\
&= \frac{1}{12} \cdot \frac{4 + 3 \cdot 16 + 12 \cdot 32 - 49 - 21 \cdot 8}{32} \\
&= \underline{\underline{\frac{219}{384} = 0.570.}}
\end{aligned}$$

B: $\frac{271}{320} = 0.847$

C: $\frac{15}{32} = 0.469$

D: $\frac{37}{50} = 0.74$

Problem 10

a)

- Det er rimelig å anta at hver av de 1000 forsøkspersonene får immunitet uavhengig av hverandre og med samme sannsynlighet p. Dermed har vi oppfylt betingelsen for binomisk fordeling, slik at antall forsøkspersoner som oppnår immunitet, X, er binomisk fordelt med $n = 1000$ forsøk og sannsynlighet for suksess lik p.
- Siden vi ønkser å avgjøre om det er grunnlag for åpåstå at den nye vaksinene er bedre enn den gamle velger dette som alternativ hypotese, dvs.

  $$H_0 : p = p_0 \quad \text{mot} \quad H_1 : p > p_0,$$

  der $p_0 = 0.9$.

- Siden $n = 1000$ er stor, og spesielt er $np_0 = 1000 \cdot 0.9 = 900 > 5$ og $n(1 - p_0) = 100 > 5$, vil X være tilnærmet normalfordelt under H_0. Under H_0 har vi dermed tilnærmet at

  $$X \sim N(np_0, np_0(1 - p_0))$$

  og den tilsvarende standardiserte variabelen blir

  $$Z = \frac{X - np_0}{\sqrt{np_0(1 - p_0)}} \sim N(0, 1).$$

  Vi bruker dermed testobservatoren

  $$Z = \frac{X - np_0}{\sqrt{np_0(1 - p_0)}}$$

  og under H_0 er denne (tilnærmet) standard normalfordelt.

- Vi forkaster H_0 dersom $Z \geq k$ der k bestemmes fra kravet

  $$P(\text{Forkast } H_0 | H_0 \text{ er riktig}) = \alpha = 0.05$$
  $$P(Z \geq k | p = 0.9) = 0.05 \Rightarrow k = z_\alpha = z_{0.05} = 1.645.$$

  Vi forkaster dermed H_0 dersom $Z \geq 1.645$. Innsatt $x = 912$ blir observert verdi for testobservatoren

  $$z = \frac{912 - 1000 \cdot 0.9}{\sqrt{1000 \cdot 0.9 \cdot (1 - 0.9)}} = 1.265.$$

  Siden $z = 1.265 \not\geq 1.645$ blir konklusjonen at man ikke skal forkaste H_0. Det er ikke grunnlag for å påstå at den nye vaksinen er bedre enn den gamle.

B: Forkaster H_0 hvis $Z \geq z_{0.01} = 2.326$, observert verdi for testobservator $z = \frac{815 - 1000 \cdot 0.8}{\sqrt{1000 \cdot 0.8 \cdot (1 - 0.80)}} = 1.186$

C: Forkaster H_0 hvis $Z \geq z_{0.05} = 1.645$, observert verdi for testobservator $z = \frac{961 - 1000 \cdot 0.95}{\sqrt{1000 \cdot 0.95 \cdot (1 - 0.095)}} = 1.596$

b)

- Vi forutsetter nå at $p = 0.92$ og ønsker sannsynligheten

  $$P\left(Z \geq 1.645 | p = 0.92\right).$$

Merk at når $p = 0.92$ er H_0 ikke korrekt og dermed er ikke testobservatoren Z standard normalfordelt. Derimot har vi også i denne situasjonen (tilnærmet) at

$$\frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1).$$

Ved å benytte dette får vi at

$$\begin{aligned}
P(Z \geq 1.645 | p = 0.92) &= P\left(\frac{X - np_0}{\sqrt{np_0(1-p_0)}} \geq 1.645 \middle| p = 0.92 \right) \\
&= P\left(X \geq np_0 + 1.645\sqrt{np_0(1-p_0)} \middle| p = 0.92 \right) \\
&= P\left(\frac{X - np}{\sqrt{np(1-p)}} \geq \frac{np_0 + 1.645\sqrt{np_0(1-p_0)} - np}{\sqrt{np(1-p)}} \middle| p = 0.92 \right) \\
&= 1 - P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{np_0 + 1.645\sqrt{np_0(1-p_0)} - np}{\sqrt{np(1-p)}} \middle| p = 0.92 \right) \\
&= 1 - \Phi\left(\frac{np_0 + 1.645\sqrt{np_0(1-p_0)} - np}{\sqrt{np(1-p)}} \right) \\
&= 1 - \Phi\left(\frac{1000 \cdot 0.9 + 1.645\sqrt{1000 \cdot 0.9 \cdot 0.1} - 1000 \cdot 0.92}{\sqrt{1000 \cdot 0.92 \cdot 0.08}} \right) \\
&= 1 - \Phi(-0.51) = 1 - 0.3050 = \underline{\underline{0.695}}.
\end{aligned}$$

- Ønsker nå å bestemme n slik at

  $$P(Z \geq 1.645 | p = 0.92) \geq 0.8.$$

  Benytter at vi fra utregningen over vet at

  $$P(Z \geq 1.645 | p = 0.92) = 1 - \Phi\left(\frac{np_0 + 1.645\sqrt{np_0(1-p_0)} - np}{\sqrt{np(1-p)}} \right),$$

slik at vi trenger å bestemme n slik at

$$1-\Phi\left(\frac{np_0+1.645\sqrt{np_0(1-p_0)}-np}{\sqrt{np(1-p)}}\right)\geq 0.8$$
$$\Phi\left(\frac{np_0+1.645\sqrt{np_0(1-p_0)}-np}{\sqrt{np(1-p)}}\right)\leq 0.2$$
$$\frac{np_0+1.645\sqrt{np_0(1-p_0)}-np}{\sqrt{np(1-p)}}\leq -z_{0.2}=-0.842$$
$$np_0+1.645\sqrt{np_0(1-p_0)}-np\leq -0.842\sqrt{np(1-p)}$$
$$n(p_0-p)+\sqrt{n}\left(1.645\sqrt{p_0(1-p_0)}+0.842\sqrt{p(1-p)}\right)\leq 0$$

Setter vi inn $p_0=0.9$ og $p=0.92$ får vi da ulikheten

$$-0.02n+0.7219\sqrt{n}\leq 0$$
$$\sqrt{n}\left(-0.02\sqrt{n}+0.7219\right)\leq 0$$
$$-0.02\sqrt{n}+0.7219\leq 0$$
$$\sqrt{n}\geq\frac{-0.7219}{-0.02}$$
$$n\geq\left(\frac{0.7219}{0.02}\right)^2=1302.85.$$

Siden antall personer man skal teste nødvendigvis må være et heltall betyr dette at man må teste minst 1303 forsøkspersoner for at teststyrken skal bli minst 0.8.

B: $P(Z\geq 2.326|p=0.83)=0.5199$, $n\geq 1727$

C: $P(Z\geq 1.645|p=0.96)=0.6915$, $n\geq 2741$

Problem 11

Her får vi

$$\bar{X}=\frac{1}{2}(X_1+X_2)\sim N\left(\mu,\frac{\sigma^2}{2}\right)$$

og vi har selvfølgelig

$$X_3\sim N(\mu,\sigma^2).$$

Dermed får vi

$$\bar{X} - X_3 \sim N\left(0, \frac{3}{2}\sigma^2\right).$$

Ved å standardisere denne får vi at

$$\frac{\bar{X} - X_3}{\sqrt{\frac{3}{2}\sigma^2}} \sim N(0,1).$$

Siden σ^2 er ukjent må vi etablere en estimator for denne,

$$\begin{aligned} S^2 &= \frac{1}{2-1}\sum_{i=1}^{2}(X_i - \bar{X})^2 = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 \\ &= \left(X_1 - \frac{1}{2}(X_1 + X_2)\right)^2 + \left(X_2 - \frac{1}{2}(X_1 + X_2)\right)^2 \\ &= \left(\frac{1}{2}(X_1 - X_2)\right)^2 + \left(\frac{1}{2}(X_2 - X_1)\right)^2 \\ &= 2 \cdot \frac{1}{2^2}(X_1 - X_2)^2 = \frac{1}{2}(X_1 - X_2)^2. \end{aligned}$$

Dermed får vi at

$$\begin{aligned} T &= \frac{\bar{X} - X_3}{\sqrt{\frac{3}{2}S^2}} = \frac{\bar{X} - X_3}{\sqrt{\frac{3}{2} \cdot \frac{1}{2}(X_1 - X_2)^2}} \\ &= \frac{\bar{X} - X_3}{\sqrt{\frac{3}{4}(X_1 - X_2)^2}} = \frac{\bar{X} - X_3}{\frac{\sqrt{3}}{2}|X_1 - X_2|} \sim t_{2-1} \end{aligned}$$

som i sin tur gir at

$$P\left(-t_{1,\frac{\alpha}{2}} \leq T \leq t_{1,\frac{\alpha}{2}}\right) = 1 - \alpha.$$

Setter vi inn uttrykket vi har for T og at $1 - \alpha = \frac{1}{3}$ slik at $\alpha = \frac{2}{3}$ og $\frac{\alpha}{2} = \frac{1}{3}$ gir dette

$$P\left(-t_{1,\frac{1}{3}} \leq \frac{\bar{X} - X_3}{\frac{\sqrt{3}}{2}|X_1 - X_2|} \leq t_{1,\frac{1}{3}}\right) = 1 - \frac{1}{3} \tag{1}$$

Setter vi inn den oppgitte kvantilen $t_{1\frac{1}{3}} = \frac{1}{\sqrt{3}}$ gir den venstre ulikheten i dette uttrykket at

$$\begin{aligned} &-\frac{1}{\sqrt{3}} \leq \frac{\bar{X} - X_3}{\frac{\sqrt{3}}{2}|X_1 - X_2|} \\ &-\frac{1}{2}|X_1 - X_2| \leq \bar{X} - X_3 \\ &-\bar{X} - \frac{1}{2}|X_1 - X_2| \leq -X_3 \\ &X_3 \leq \bar{X} + \frac{1}{2}|X_1 - X_2| = \frac{1}{2}(X_1 + X_2) + \frac{1}{2}|X_1 - X_2|. \end{aligned}$$

Ved å betrakte hver av de to tilfelle $X_1 \leq X_2$ og $X_1 > X_2$ hver for seg kan man verifisere at $(X_1 + X_2) + |X_1 - X_2| = 2\max\{X_1, X_2\}$. Ved å benytte dette får vi da at den venstre ulikheten blir

$$X_3 \leq \max\{X_1, X_2\}.$$

For den høyre ulikheten får vi tilsvarende

$$\begin{aligned}
&\frac{\bar{X} - X_3}{\frac{\sqrt{3}}{2}|X_1 - X_2|} \leq \frac{1}{\sqrt{3}} \\
&\bar{X} - X_3 \leq \frac{1}{2}|X_1 - X_2| \\
&-X_3 \leq -\bar{X} + \frac{1}{2}|X_1 - X_2| \\
&X_3 \geq \bar{X} - \frac{1}{2}|X_1 - X_2| = \frac{1}{2}(X_1 + X_2) - \frac{1}{2}|X_1 - X_2|.
\end{aligned}$$

Ved å betrakte hver av de to tilfellene $X_1 \leq X_2$ og $X_1 > X_2$ hver for seg kan man her virifisere at $(X_1 + X_2) - |X_1 - X_2| = \min\{X_1, X_2\}$, og ved å benytte dette får vi at den høyre ulikheten blir

$$X_3 \geq \min\{X_1, X_2\}.$$

Dermed har vi at utrykket i (1) kan skrives som

$$P\left(\min\{X_1, X_2\} \leq X_3 \leq \max\{X_1, X_2\}\right) = 1 - \frac{1}{3},$$

slik at $\left(1 - \frac{1}{3}\right)$-prediksjonsintervall for X_3 blir

$$\underline{\underline{[\min\{X_1, X_2\}, \max\{X_1, X_2\}]}}.$$

For observasjonene $X_1 = 2$ og $x_2 = 5$ blir intervallet $\underline{\underline{[2{,}5]}}$.

Siden variablene X_1, X_2, X_3 kommer fra samme kontinuerlige fordeling vil det være sannsynlighet $\frac{1}{3}$ for at X_3 er den minste av disse variablene, det vil være sannsynlighet $\frac{1}{3}$ for at X_3 vil være den største av variablene, og det vil være sannsynlighet $\frac{1}{3}$ for at X_3 vil være den midterste verdien. Prediksjonsintervallet som ble utledet over er en måte å utrykke den siste av disse sannsynlighetene.

B: $[\min\{X_1, X_2\}, \max\{X_1, X_2\}] = [12, 15]$

C: $[\min\{X_1, X_2\}, \max\{X_1, X_2\}] = [1, 6]$