# MAT122 Oblig

Thobias Høivik

## Oppgave 1

*Løsning for a)*

La $M = \{\text{studerer matematikk}\}$ og $I = \{\text{studerer informatikk}\}$.

$$\begin{aligned} |M| &= 120 \\ |I| &= 60 \\ |M \cap I| &= 30 \end{aligned}$$

Da blir mengden av de som studerer matematikk, men ikke informatikk $M \setminus I = M \setminus (M \cap I)$ og kardinaliteten av denne mengden blir 90.

Dermed er sjansen for at en student studerer matematikk, men ikke informatikk

$$P(\text{matematikk, men ikke informatikk}) = \frac{90}{200} = 45\%$$

For å finne sjansen for at en studerer informatikk, men ikke matematikk bruker vi samme fremgansmåte:

$$\begin{aligned} I \setminus M &= I \setminus (M \cap I) \\ P(\text{informatikk, men ikke matematikk}) &= \frac{30}{200} = 15\% \end{aligned}$$

*Løsning for b)*

$$P(\text{matematikk eller informatikk}) = \frac{150}{200} = 75\%$$

Med andre ord:

$$\frac{|M \cup I| \setminus |M \cap I|}{200}$$

for å ungå dobbeltelling.

*Løsning for c)*

$$P(\text{verken mat eller inf}) = P(1 - \text{mat eller inf}) = 40\%$$

## Oppgave 2

*Løsning for a)*

Ved bruk av Bayes' teorem får vi at

$$P(\text{Korona} \mid \text{Positiv test}) = \frac{P(\text{Positiv Test} \mid \text{Korona})P(\text{Korona})}{P(\text{Positiv test})}$$

*Løsning for b)*

$$\begin{aligned}
P(\text{Positiv test}) &= P(\text{Positiv test} \mid \text{Korona})P(\text{Korona}) + P(\text{Positiv test} \mid \overline{\text{Korona}})P(\overline{\text{Korona}}) \\
P(\text{Positiv test}) &= 0.911 \cdot 0.01 + 0.004 \cdot (0.99) = 0.01307
\end{aligned}$$

*Løsning for c)*

$P(\text{Korona}) = 0.1$ og $P(\text{Positiv} \mid \overline{\text{Korona}}) = 0.004$.

$$\begin{aligned}
P(\text{Positiv}) &= 0.911 \cdot 0.1 + 0.004 \cdot (1 - 0.1) \\
&= 0.0911 + 0.0036 \\
&= 0.0947
\end{aligned}$$

*Løsning for d)*

1. **1.** For $P(\text{Korona}) = 0.01$:

$$P(\text{Korona} \mid \text{Positiv}) = \frac{0.911 \cdot 0.01}{0.01307} \approx 69.7\%$$

2. **2.** For $P(\text{Korona}) = 0.1$:

$$P(\text{Korona} \mid \text{Positiv}) = \frac{0.911 \cdot 0.1}{0.0947} \approx 96.2\%$$

*Løsning for e)*

Selv om testen har høg sensitivitet på 91.1% og lav falsk-positiv rate på 0.4% så vil en lav prevalens på 1% føre til en lavere positivt prediktiv verdi på 69.7% kontra den 96.2% vi får når en høg andel er smittet (10%).

## Oppgave 3

*Løsning for a)*

Vi snakker om antall forekomster av en hendelse i et tidsinterval. Vi blir gitt at hendelsene er tilfeldig fordelt i dette intervallet og uavhengige. Vi kan også gjør en simplifiserende antagelse at ingen hendelser skjer på akkurat samme tid. Dermed får vi at

$$X \sim \text{poisson}(\lambda t)$$

Siden antall forekomster er tilfeldig fordelt iløpet av et 8 timer tidsinterval kan vi fint anta at vi da forventer $32/8 = 4$. Så det er naturlig å snakke om $t$ timer, også siden de følgende oppgavene snakker om individuelle timer.

*Løsning for b)*

Hva er sjansen for 0 forespørseler på en time?

For $t = 8$ har forventer vi 32. Dermed har vi at på en time:

$$E(X) = \lambda t = 4$$

så $\lambda = 4$.

$$P(X = 0) = \frac{(\lambda t)^x}{x!} e^{-\lambda t} = e^{-4} = 0.01831563888 \approx 1.8\%$$

*Løsning for c)*

Det minste antallet forespørseler vi kan ha er 0. En negativ verdi gir ikke mening.

Dermed får vi at

$$P(X \geq 1) = 1 - P(X = 0)$$

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - 0.01831563888 \\ &= 0.98168436112 \\ &\approx 98\% \end{aligned}$$

*Løsning for d)*

Vi ønsker å finne $P(X \geq 32)$ for $t = 8$. Det blir $1 - P(X \leq 31)$. $P(X \leq 31)$ er nøyaktig

$$\sum_{x=0}^{31} P(X = x)$$

som er plagsomt å finne for hand. Vi kan merke at poissonfordelingen vår konvergerer her mot en normalfordeling (som er symmetrisk) og siden 32 er middelverdien vil $P(X \leq 31)$ være ca. halvparten av arealet under normalfordelingen (50%) , men en gøyere måte å finne den er å ta integralet

$$\int_0^{31} \frac{(\lambda t)^x}{x!} e^{-\lambda t} dx$$

Dette integralet blir nødvendighvis mindre enn eller lik den faktiske sannsynligheten siden

$$\int_0^N f(x)dx = \sum_{n=1}^{N} \int_{n-1}^{n} f(x)dx \leq \sum_{n=0}^{N} f(n)$$

for en $f : [0, \infty) \to [0, \infty)$.

De betyr at 1 minus dette integralet vil være høyere enn eller lik $1 - P(X \leq 31)$.

Dermed kan vi gi en lavere-bund for $P(X \leq 31) = I$ hvor

$$I = \int_0^{31} \frac{32^x}{\Gamma(x+1)} e^{-32} dx$$

hvor $\Gamma(x+1) = \int_0^\infty e^{-z} z^{x-1} dz = x!$, når $x \in \mathbb{N}$.

Vi får at

$$I \approx 0.441226\cdots \leq P(X \leq 31)$$

Dermed får vi at

$$P(X \geq 32) \leq 1 - 0.441226\cdots = 0.558774$$

## Oppgave 4

Lavere standardavvik betyr høyere konsentrasjon ved middelverdien. Dermed er den blå kurven $\sigma = 0.5$ og den oransje er $\sigma = 0.9$.

## Oppgave 5

$X \sim \mathcal{N}(3510, 385)$.

*Løsning for a)*

$$P(X \geq 4500) = P(X > 4500) = 1 - G\left(\frac{4500 - 3510}{385}\right) = 0.0051$$

*Løsning for b)*

$$P(2500 < X < 4500) = G\left(\frac{4500 - 3510}{385}\right) - G\left(\frac{2500 - 3510}{385}\right) = 0.9906$$

## Oppgave 6

*Løsning for a)*

$$\begin{cases} H_0 : \mu \geq= 400 \\ H_1 : \mu < 400 \end{cases}$$

Siden vi ikke er misnøgde med en batterilengde på mer en det som er reklamert.

Vi har et estimat for $X$ og et estimat for standardavviket $S$. Vi bruker en $T$-test med 19-frihetsgrader.

Forkaster $H_0$ viss $T < -t_\alpha$.

*Løsning for b)*

$\alpha = 0.05$

$$|T| = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{380 - 400}{10 - \sqrt{20}} \approx -3.618 < -2.433 = -t_{0.025}$$

*Løsning for c)*

$$\begin{aligned} [a, b] &= \overline{X} \pm t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} \\ &= 380 \pm 2.433 \cdot \frac{10}{\sqrt{20}} \\ &= 380 \pm 5.436 \\ &= [374.564, 385.436] \end{aligned}$$

Eit 90% konfidensinterval vil være smalere (eller like stort) siden et mindre "vindu" for korrekte verdier gir en lavere sjanse for at vi treffer.

## Oppgave 7

*Løsning for a)*

$$\begin{cases} H_0 : \mu = 26 \\ H_1 : \mu \neq 26 \end{cases}$$

Vi kjenner $\sigma$ og kjenner ikkje $\mu$ så vi gjøre en $Z$-test.

$$Z = \frac{\overline{X} - 26}{\sigma / \sqrt{n}}$$

Forkast $H_0$ viss $|Z| > z_{\alpha/2}$.

*Løsning for b)* $\alpha = 0.05 \Rightarrow \alpha/2 = 0.025$.

$$Z = \frac{28 - 26}{4/\sqrt{10}} = 1.58 < 1.96 = z_{\alpha/2}$$

*Løsning for c)*

$$\begin{aligned} [a, b] &= 28 \pm 1.96 \frac{4}{\sqrt{10}} \\ &= 28 \pm 2.48 \\ &= [25.52, 30.48] \end{aligned}$$

Igjen, 90% gir et smalere interval.

*Løsning for d)*

9 frihetsgrader.

$$\begin{aligned} 28 \pm t_{\alpha/2} \frac{4}{\sqrt{10}} &= 28 \pm 2.262 \frac{4}{\sqrt{10}} \\ &= 28 \pm 2.86 \\ &= [25.14, 30, 86] \end{aligned}$$

Intervallet blir større fordi det er mer usikkerhet ($\sigma$ ukjent).

## Oppgave 8

*Løsning for a)*

$X \sim \text{bin}(n, p)$ hvor $n$ er antall forsøk (1200) og $p$ er sannsynet for positivt utfall (1/6).

$$E(X) = np = 1200/6 = 200$$

$$Var(X) = np(1-p) = 200 \cdot \frac{5}{6} = 166.\overline{6}$$

*Løsning for b)*

$$\begin{cases} H_0 : \mu = \frac{1}{6} \\ H_1 : \mu \neq \frac{1}{6} \end{cases}$$

$\alpha = 0.05 \Rightarrow \alpha/2 = 0.025$.

Vi får $\widehat{p} = 172/1200 = 0.14\overline{3}$.

$$Z = \frac{0.14\overline{3} - \frac{1}{6}}{\sqrt{\frac{\frac{1}{6}(\frac{5}{6})}{n}}} \approx -2.17$$

$$Z_{0.025} = 1.96 < |Z|$$

Dermed forkaster vi $H_0$ og sier med 95% sikkerhet at terningen ikke er rettferdig.

## Oppgave 9

*Løsning for a)*

Siden vi snakker om antall hendelser innen et tidsrom og det er rimelig å anta at om en gitt pasient blir henvist eller ikke har ingen påvirkning på om en annen pasient blir henvist samt at å anta at ingen to pasienter blir henvist på akkurat samme tid, kan vi anta at antall henvisninger er poissonfordelt.

$$p = \frac{24300}{650000} = 0.0373846$$

For et fastlegekontor med $n = 6200$ pasienter har vi

$$\lambda = np = 231.7846$$

$$\begin{aligned} E(X) = Var(X) = \sigma^2 = \lambda = 231.7846 \\ \sigma = 15.22 \end{aligned}$$

*Løsning for b)*

$$\begin{cases} H_0 : p = 0.0373846 \\ H_1 : p \neq 0.0373846 \end{cases}$$

Viss $H_0$ ikke blir forkastet forventer vi rundt $\lambda$ henvisninger.

$X = 276$.

For stor $\lambda$ går det mot normalfordelt.

$$Z = \frac{276 - 231.78}{15.22} \approx N(0,1)$$

*Løsning for c)* $\alpha = 0.05 \Rightarrow \alpha/2 = 0.025$.

$$|Z| = |Z| \approx 2.9 > 1.96$$

Forkast $H_0$.

## Oppgave 10

$$\hat{y} = 19.83 + 0.85x$$

180 tonn fordelt tilfeldig utover 20 mål gir i gjennomsnitt 9 tonn per mål.

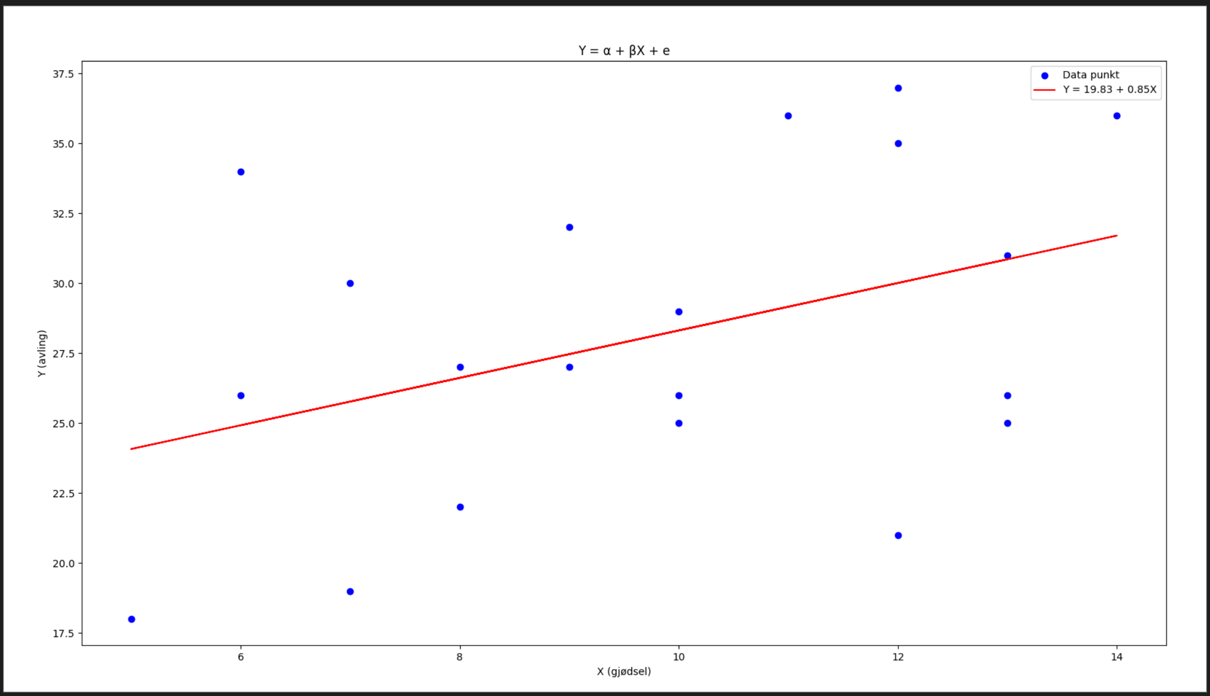
$$\hat{y}(9) = 27.48 \text{ tonn per mål}$$

Til sammen over 20 mål har vi

$$20 \cdot 27.48 = 549.6 \text{ tonn avlinger}$$

Listing 1: Kode

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression


X = np.array([8, 7, 13, 9, 6, 5, 11, 9, 14, 13, 6, 10, 12, 13, 12, 10,
    7, 8, 12, 10]).reshape(-1, 1)
Y = np.array([27, 30, 26, 32, 26, 18, 36, 27, 36, 31, 34, 29, 21, 25,
    35, 26, 19, 22, 37, 25])


model = LinearRegression()
model.fit(X, Y)

alpha = model.intercept_
beta = model.coef_[0]
print(f"a:{alpha:.2f}")
print(f"b:{beta:.2f}")


Y_pred = model.predict(X)

plt.scatter(X,Y,color='blue',label='Data-punkt')

plt.plot(X,Y_pred,color='red',label=f'Y={alpha:.2f}+{beta:.2f}X')

plt.xlabel('X(gjodsel)')
plt.ylabel('Y(avling)')
plt.title('Y=a+bX+e')
plt.legend()
plt.show()
```


Y = α + βX + e
Data punkt
Y = 19.83 + 0.85X
Y (avling)
X (gjødsel)
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
6
8
10
12
14