Løsningsforslag Eksamen i Statistikk Mai 2000

Oppgave 1

a)

$$\begin{aligned}
P(X < 6.74) &= P(\frac{X - 6.8}{0.06} < \frac{6.74 - 6.8}{0.06}) \\
&= \Phi(-1) = 1 - \Phi(1) \\
&= 1 - 0.841 = 0.159
\end{aligned}$$

$$\begin{aligned}
P(6.74 < X < 6.86) &= P(X < 6.86) - P(X < 6.74) \\
&= P(\frac{X - 6.8}{0.06} < \frac{6.86 - 6.8}{0.06}) - 0.159 \\
&= \Phi(1) - 0.159 = 0.841 - 0.159 = 0.682
\end{aligned}$$

$$\begin{aligned}
P(|X - \mu|) > 0.06) &= P(X - \mu < -0.06) + P(X - \mu > 0.06) \\
&= P(\frac{X - \mu}{0.06} < -1) + P(\frac{X - \mu}{0.06} > 1) \\
&= \Phi(-1) + 1 - \Phi(1) = 2(1 - \Phi(1)) = 0.318
\end{aligned}$$

Eventuelt

$$\begin{aligned}
P(|X - \mu|) > 0.06) &= 1 - P(6.74 < X < 6.86) \\
&= 1 - 0.682 = 0.318
\end{aligned}$$

b) $Y \sim N(\mu, \frac{\sigma^2}{5})$

$$\begin{aligned}
P(|Y - \mu|) > 0.06) &= 2P(Y - \mu > 0.06) \\
&= 2(1 - P(\frac{Y - \mu}{\frac{0.06}{\sqrt{5}}} \leq 1)) \\
&= 0.026
\end{aligned}$$

$Y = \frac{1}{5}\sum_{i=1}^{5} X_i$ er lineærkombinasjon av uavhengige normalfordelte variable. Dermed er Y normalfordelt med $E(Y) = \mu$ og $Var(Y) = \frac{\sigma^2}{5}$

$\Longrightarrow \frac{Y-\mu}{\frac{\sigma}{\sqrt{5}}} \sim N(0,1)$
$\Longrightarrow P(-Z_{0.025} < \frac{y-\mu}{\frac{\sigma}{\sqrt{5}}} < Z_{0.025}) = 0.95$

$P(Y - Z_{0.025}.\frac{\sigma}{\sqrt{5}} < \mu < Y + Z_{0.025}.\frac{\sigma}{\sqrt{5}}) = 0.95$
D.v.s 95% konf. int. blir:

$$[Y - Z_{0.025}.\frac{\sigma}{\sqrt{5}}, Y + Z_{0.025}\frac{\sigma}{\sqrt{5}}]$$

Innsatt tall:

$y = \bar{x} = 6.76, \sigma = 0.06, z_{0.025} = 1.96$

$[6.76 - (1.96)\frac{0.06}{\sqrt{5}}, 6.76 + (1.96)\frac{0.06}{\sqrt{5}}] = [6.707, 6.813]$

Oppgave 2

$P(E_2|E_1) \neq P(E_2) \Longrightarrow$ ikke uavhengige

$$\begin{aligned} P(E_1 \cap E_2) &= P(E_2|E_1).P(E_1) \\ &= (0.1)(0.01) = 0.001 > 0 \end{aligned}$$

dvs $E_1 \cap E_2 \neq \emptyset \Longrightarrow$ ikke disjunkte

$$\begin{aligned} P(E_1 \cup E_2) &= P(E_1) + P(E_2) - P(E_1 \cap E_2) \\ &= 2(0.01) - (0.001) = 0.019 \end{aligned}$$

La $F =$ minst to av tre pumper er i feiltilstand ser ved bruk av addisjonssetningen at:

$$\begin{aligned} P(F) &= P(E_1 \cap E_2) + P(E_1 \cap T) + P(E_1 \cap T) - 2.P(E_1 \cap E_2 \cap T) \\ &= P(E_1 \cap E_2) + P(E_1).P(T) + P(E_1).P(T) - 2.P(E_1 \cap E_2).P(T) \\ &= 0.001 + 0.01(0.04) + 0.01(0.04) - 2(0.001)(0.04) = 0.00172 \end{aligned}$$

Oppgave 3

a) $Y \sim Poisson(5t)$

$t = 1 : P(Y > 12) = 1 - P(Y \leq 12) = 1 - 0.99798 = 0.002$
$t = 0.5 : P(Y > 6) = 1 - P(Y \leq 6) = 1 - 0.99581 = 0.014$

P(minst en av 10 prøver > 12) $= 1 - P$(ingen av 10 prøver > 12) $= 1 - P$(alle prøver ≤ 12) $= 1 - P(\text{prøve} \leq 12)^{10} = 1 - (0.99798)^{10} = 0.020$

b) $Y_i \sim Poisson(\mu t_i)$

$$\Longrightarrow L(\mu) = \prod_{i=1}^{n} f(y_i; t_i, \mu) = \prod_{i=1}^{n} \frac{(\mu t_i)^{y_i}}{(y_i)!} e^{-\mu t_i}$$

$$ln(L(\mu)) = \sum_{i=1}^{n} y_i ln(\mu t_i) - ln(\prod_{i=1}^{n} y_i!) - \mu \sum_{i=1}^{n} t_i$$

$$\frac{\partial ln L(\mu)}{\partial \mu} = \sum_{i=1}^{n} \frac{y_i}{\mu} - \sum_{i=1}^{n} t_i$$

$$\Longrightarrow \mu = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} t_i}$$

D.v.s SME blir : $\hat{\mu} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} t_i}$

$E(\hat{\mu}) = \frac{1}{\sum_{i=1}^{n} t_i} \sum_{i=1}^{n} E(Y_i) = \frac{1}{\sum_{i=1}^{n} t_i} \sum_{i=1}^{n} \mu t_i = \mu$
$Var(\hat{\mu}) = (\frac{1}{\sum_{i=1}^{n} t_i})^2 Var(\sum_{i=1}^{n} Y_i) \overset{\text{uavh.}}{=} (\frac{1}{\sum_{i=1}^{n} t_i})^2 \sum_{i=1}^{n} Var(Y_i)$
$= (\frac{1}{\sum_{i=1}^{n} t_i})^2 . \sum_{i=1}^{n} \mu t_i = \frac{\mu}{\sum_{i=1}^{n} t_i}$

c) $H_o : \mu$ mot $H_1 : \mu > 6$
$\hat{\mu} = \frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} t_i}$
Siden Y_i er Poissonfordelt med parameter μt_i, vil $\sum_{i=1}^{n} Y_i$ selv være Poissonfordelt med parameter $\mu \sum_{i=1}^{n} t_i$. Under H_o er $\mu = 6$, dvs dersom $\sum_{i=1}^{n} t_i > 2.5$ har vi fra oppgitt resultat at $\sum_{i=1}^{n} Y_i$ vil være tilnærmet normalfordelt. Dermed vil ogsa $\hat{\mu}$ være tilnærmet normalfordelt med :

$E(\hat{\mu}) = \frac{1}{\sum_{i=1}^{n} t_i} \sum_{i=1}^{n} E(Y_i) = \frac{1}{\sum_{i=1}^{n} t_i} \sum_{i=1}^{n} \mu t_i = \mu = 6$

$Var(\hat{\mu}) = (\frac{1}{\sum_{i=1}^{n} t_i})^2 Var(\sum_{i=1}^{n} Y_i) \overset{\text{uavh.}}{=} (\frac{1}{\sum_{i=1}^{n} t_i})^2 \sum_{i=1}^{n} Var(Y_i)$

$= (\frac{1}{\sum_{i=1}^n t_i})^2 \cdot \sum_{i=1}^n \mu t_i = \frac{\mu}{\sum_{i=1} {}^n t_i} = \frac{6}{\sum_{i=1}^n t_i}$
$\Longrightarrow Z = \frac{\hat{\mu}-6}{\sqrt{\frac{6}{\sum_{i=1}^n t_i}}}$
er tilnærmet $N(0,1)$ under H_o.
Forkaster H_o dersom $Z \geq z_{0.05}$ der
P(forkaste $H_o|H_o) = P(Z \geq z_{0.05}|\mu = 6) = 0.05$

Observert : $z_{obs} = \frac{65/10-6}{\sqrt{6/10}} = 0.645 > z_{0.05} = 1.645 \Longrightarrow$ Forkaster ikke H_o.
Eventuelt : $p - verdi = P(Z \geq z_{obs}) = P(Z \geq 0.645)$
$= 1 - \phi(0.645) = 1 - 0.741 = 0.259$

$p - verdi > 0.05 \Longrightarrow$ Forkaster ikke H_o.

d)

$X|Y = y \sim bin(y,p)$
- y bakterier oppdages(oversees) uavh. av hverandre
- har at bakterier vil enten oppdages eller ikke
- $p =$ P(bakterier oppdages) er den samme for alle bakterier

$\Longrightarrow X =$ antall av y bakterier som oppdages er $bin(y,p)$

$$\begin{aligned}
f(x,y) &= P(X = x, Y = y) \\
&= P(X = x|Y = y) \cdot P(Y = y) \\
&= \binom{y}{x} p^x (1-p)^{y-x} \frac{(\mu t)^y}{y!} e^{-\mu t}
\end{aligned}$$

$$\begin{aligned}
f(x) &= \sum_{i=1}^{\infty} f(x,y) \\
&= \sum_{i=1}^{\infty} \frac{y!}{x!(y-x)!} p^x (1-p)^{y-x} \frac{(\mu t)^{y-x} (\mu t)^x}{y!} e^{-\mu t} \\
&= \sum_{i=1}^{\infty} \frac{((1-p)\mu t)^{y-x}}{(y-x)!} \cdot \frac{(p\mu t)^x}{x!} \cdot e^{-\mu t} \\
&= e^{(1-p)\mu t} \cdot e^{-\mu t} \cdot \frac{(p\mu t)^x}{x!} \\
&= e^{-p\mu t} \cdot \frac{(p\mu t)^x}{x!}
\end{aligned}$$

Dvs $X \sim Poisson(\mu t p)$

Rimelig estimator :

$$\hat{\mu} = \frac{X_1 + \cdots + X_n}{p(t_1 + \cdots + t_n)}$$

Oppgave 4

a)

$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{-52.57}{60} = -0.876$
$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{x} = \frac{360.37}{9} + 0.876 \cdot \frac{369}{9} = 75.96$

$\hat{\sigma^2} = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2$

b)

$Var(\hat{\beta}) = Var(\frac{\sum_{i=1}^n (x_i - \bar{x})Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}) = \frac{1}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} \cdot Var(\sum_{i=1}^n (x_i - \bar{x})Y_i) = \frac{1}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} \cdot \sum_{i=1}^n (x_i - \bar{x}) Var(Y_i) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$

$H_0 : \beta = 0 \qquad H_1 : \beta \neq 0$

Under H_0 har vi følgende observator og fordeling: $T = \frac{\hat{\beta}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim t_{n-2}$

Vi forkaster H_0 dersom $T > t_{n-2,\alpha/2}$ eller om $T < -t_{n-2,\alpha/2}$. Med $\alpha = 0.01$, $n = 9$ har vi at $t_{7,0.005} = 3.5$.

$T = \frac{-0.876}{\frac{1.568}{\sqrt{60}}} = -4.33 < -3.50$
Vi forkaster H_0 på nivå $\alpha = 0.01$.

Tolkningen blir at alder har betydning for løypetiden.

c)

Predikert tid er $\hat{Y}_0$.
$\hat{Y}_0 = \hat{\alpha} - \hat{\beta} x_0 = 75.96 - 0.876 \cdot 46 = 35.66$

$\hat{Y}_0$ er et estimat på sann verdi Y_0. Siden Y_i ene er uavhengige og $\hat{Y}_0$ er basert på andre Y_i er enn Y_0, så er $\hat{Y}_0$ og Y_0 uavhengige.
$E(\hat{Y}_0 - Y_0) = E(\hat{\alpha}) + E(\hat{\beta}) x_0 - \alpha - \beta x_0 = 0$
$Var(\hat{Y}_0 - Y_0) = Var(\hat{Y}_0) + Var(Y_0) = \sigma^2 (\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}) + \sigma^2 = \sigma^2 \cdot v$

Vi har at observatoren $S = \frac{\hat{Y}_0 - Y_0}{\hat{\sigma} \cdot \sqrt{v}} \sim t_{n-2}$

$P(-t_{n-2,\alpha/2} < S < t_{n-2,\alpha/2}) = 1 - \alpha$
$P(\hat{Y}_0 - t_{n-2,\alpha/2}\hat{\sigma}\sqrt{v} < Y_0 < \hat{Y}_0 + t_{n-2,\alpha/2}\hat{\sigma}\sqrt{v}) = 1 - \alpha$

Innsatt verdier $n = 9$, $\alpha = 0.05$ og $t_{7,0.0025} = 2.36$, er et 0.95 prediksjonsintervall for Y_0 gitt ved:
$\{35.66 - 2.36 \cdot 1.568 \cdot \sqrt{1 + \frac{1}{9} + \frac{(46-41)^2}{60}}, 35.66 + 2.36 \cdot 1.568 \cdot \sqrt{1 + \frac{1}{9} + \frac{(46-41)^2}{60}}\} = \{31.09, 40.23\}$

Ekstrapolasjon så langt frem i tid bør ikke gjøres. Det er ikke sikkert at modellen holder utover de x verdiene hvor vi har data. Løperen vil neppe fortsette å forbedre seg i all fremtid.