Løsningsforslag Eksamen i Statistikk Aug 1999

Oppgave 1)

a)

Vekten på en moden kål, $X \sim n(x; 2, 0.5)$

$P(X < 1.5) = P(\frac{X-2}{0.5} < \frac{1.5-2}{0.5}) = \Phi(-1) = 1 - 0.84 = \underline{\underline{0.16}}$

$$\begin{aligned} P(2 < X < 2.5) &= P(X < 2.5) - P(X < 2) \\ &= P(\frac{X-2.5}{0.5} < \frac{2-2.5}{0.5}) - P(\frac{X-2}{0.5} < \frac{2-2}{0.5}) \\ &= \Phi(1) - \Phi(0) \\ &= 0.84 - 0.5 = \underline{\underline{0.34}} \end{aligned}$$

Vi lar Y være differansen mellom to kålhoder X_1 og X_2, $Y = X_1 - X_2$. Da er $Y \sim n(y; 0, \sqrt{2 \cdot 0.5^2})$. Ved symmetri er $P(Y < -1) = P(Y > 1) = 1 - P(Y < 1)$.

$$\begin{aligned} P(Y > 1) + P(Y < -1) &= 2(1 - P(Y < 1)) \\ &= 2(1 - \Phi(\frac{1}{0.5\sqrt{2}})) \\ &= 2(1 - \Phi(1.41)) = 2(1 - 0.92) = \underline{\underline{0.16}} \end{aligned}$$

b)

$P(2 < X < 2.5 | X > 1.5) = \frac{P(2<X<2.5)}{P(X>1.5)} = \frac{0.34}{1-0.16} = \underline{\underline{0.40}}$

Lar så Z være vekten av et klasse 1 kålhode og la $F_X(x)$ være den kumulative fordelingsfunksjonen til X. Vi setter dette inn for $P(X \leq x | X > 1.5)$:

$P(X \leq x | X > 1.5) = \frac{P(1.5<X<x)}{P(X>1.5)} = \frac{P(X<x)-P(X<1.5)}{0.84} = \frac{F_X(x)-0.16}{0.84}$ for $x > 1.5$.

Dvs

$$F_Z(z) = P(Z \leq z) = \begin{cases} \frac{F_X(z)-0.16}{0.84} & \text{for } z > 1.5 \\ 0 & \text{ellers} \end{cases}$$

c)

Rimelig estimator for μ: $\hat{\mu}_Y = \underline{\underline{\bar{Y}}}$

Estimat: $\hat{\mu}_Y = \bar{y} = \underline{\underline{2.33}}$.

Vi tar videre utgangspunkt i at $U = \frac{\sqrt{n}(\bar{Y}-\mu_Y)}{\sigma_Y}$ er normalfordelt med forventning 0 og varians 1.

Konfidensintervallet finnes da ved:

$$\begin{aligned} P(-1.645 < \frac{\sqrt{n}(\bar{Y} - \mu_Y)}{\sigma_Y} < 1.645) &= 0.90 \\ P(\bar{Y} - 1.645\frac{\sigma_Y}{\sqrt{n}} < \mu_Y < \bar{Y} + 1.645\frac{\sigma_Y}{\sqrt{n}}) &= 0.90 \end{aligned}$$

Dvs et 90%-konfidensintervall er gitt ved:

$$\underline{\underline{[\bar{Y} - 1.645\frac{\sigma_Y}{\sqrt{n}}, \bar{Y} + 1.645\frac{\sigma_Y}{\sqrt{n}}]}}$$

Innsatt tall får vi intervallet $\underline{\underline{[2.07, 2.59]}}$

Lengden på konfidensintervallet er $L = 2 \cdot 1.645\frac{\sigma_Y}{\sqrt{n}} = \underline{\underline{0.52}}$

Finner tilstrekkelig antall planter, n, for at lengden av konfidensintervallet, L, skal bli mindre enn 0.2 kg:

$$\begin{aligned} 2 \cdot 1.645 \cdot \frac{0.5}{\sqrt{n}} &< 0.2 \\ \frac{1.645}{0.2} &< \sqrt{n} \\ 67.65 &< n \end{aligned}$$

Hvis $\underline{\underline{n = 68}}$ blir lengden mindre enn den gitte grensa.

d)

For å estimere variansen bruker vi $\underline{\underline{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$

Innsatt tall får vi estimatet $\underline{\underline{s^2 = 0.585^2}}$.

Siden variansen er ukjent tar vi utgangspunkt i T-observatoren. Vi har fra pensum at $T = \sqrt{n}\frac{\bar{X}-2}{S}$ er Student T-fordelt med $n-1$ frihetsgrader under H_0 at $\mu_Y = 2$. Vi forkaster H_0 dersom $T > t_{\alpha,n-1}$ der

$$P(\text{forkaste } H_0 | H_0) = P(T > t_{\alpha,n-1}) = \alpha$$

Med $\alpha = 0.1$ og $n = 10$ får vi $t_{\alpha,n-1} = t_{0.10,9} = 1.38$

Observert: $\sqrt{10}\frac{2.33-2}{0.585} = 1.78 > 1.38$, dvs vi $\underline{\underline{\text{forkaster } H_0}}$. Dataene gir grunnlag for å påstå at den nye kålen er tyngre.

Oppgave 2

Opplysningene i oppgaven tilsier at: $P(J) = 0.8$, $P(\bar{J}) = 1 - 0.8 = 0.2$, $P(R|J) = 0.02$, $P(R|\bar{J}) = 0.05$.

Andel innringere som mener JA er ved Bayes formel:

$$P(J|R) = \frac{P(R|J)P(J)}{P(R|J)P(J) + P(R|\bar{J})P(\bar{J})} = \underline{\underline{0.62}}$$

Sammenliknet med den store gruppen som mener JA er det få som ringer inn, dermed blir resultatet for JA dårligere.

Oppgave 3

a)

La N være antall alarmer per døgn. N er da Poisson-fordelt med parameter λt, der $\lambda = 1.5$ og $t = 1$. Dvs N er Poisson-fordelt med parameter 1.5.

$$P(N \geq 2) = 1 - P(N < 2) = 1 - P(N \leq 1) = 1 - 0.56 = \underline{\underline{0.44}}$$

Dette finnes fra tabell, eller ved å bruke Poisson-punktsannsynligheten direkte.

$$P(N \geq 2 | N \geq 1) = \frac{P(N \geq 2 \cap N \geq 1)}{P(N \geq 1)} = \frac{P(N \geq 2)}{P(N \geq 1)} = \frac{1 - P(N \leq 1)}{1 - P(N \leq 0)} = \frac{1 - 0.56}{1 - 0.22} = \underline{\underline{0.57}}$$

La M være antall alarmer i løpet av 2 timer $= \frac{1}{12}$ døgn. M er da Poisson-fordelt med parameter λt, der $\lambda = 1.5$ og $t = \frac{1}{12}$. Dvs M er Poisson-fordelt med parameter 0.125.

$$P(M \geq 1) = 1 - P(M = 0) = 1 - e^{-0.125} = \underline{\underline{0.12}}$$

b)

Vi vet den momentgenererende funksjonen (MGF) til T_i fra tabell: $M_{T_i}(t) = E(e^{tT_i}) = \frac{\lambda}{\lambda - t}$. Vi bruker dette og uavhengighet av T_i-ene for å finne MGF til Z:

$$M_Z(t) = \mathrm{E}(e^{tZ}) = \mathrm{E}(e^{t2\lambda \sum_{i=1}^{n} T_i}) \overset{uavh.}{=} \prod_{i=1}^{n} \mathrm{E}(e^{2\lambda t T_i}) = M_T(2\lambda t)^n = (\frac{\lambda}{\lambda - 2\lambda t})^n = \underline{\underline{(\frac{1}{1-2t})^n}}$$

Fra tabellen har vi at MGF for en χ^2_ν-fordelt variabel er $(\frac{1}{1-2t})^{\nu/2}$. Dvs vi ser at MGF for Z er lik MGF for χ^2_{2n}-fordelingen, dvs Z er χ^2_{2n}-fordelt som skulle vises.

c) Finne konfidensintervall for λ. Vi bruker resultatet fra b) om at $Z \sim \chi^2_{2n}$.

$$\begin{aligned} P(\chi^2_{1-\frac{\alpha}{2},2n} < Z < \chi^2_{\frac{\alpha}{2},2n}) &= 1-\alpha \\ P(\chi^2_{1-\frac{\alpha}{2},2n} < 2\lambda \sum_{i=1}^{n} T_i < \chi^2_{\frac{\alpha}{2},2n}) &= 1-\alpha \\ P(\frac{\chi^2_{1-\frac{\alpha}{2},2n}}{2\sum_{i=1}^{n} T_i} < \lambda < \frac{\chi^2_{\frac{\alpha}{2},2n}}{2\sum_{i=1}^{n} T_i}) &= 1-\alpha \end{aligned}$$

Dvs et $(1-\alpha)100\%$ konfidensintervall for λ blir:

$$\underline{\underline{[\frac{\chi^2_{1-\frac{\alpha}{2},2n}}{2\sum_{i=1}^{n} T_i}, \frac{\chi^2_{\frac{\alpha}{2},2n}}{2\sum_{i=1}^{n} T_i}]}}$$

Ved å sette inn $\chi^2_{1-\frac{\alpha}{2},2n} = \chi^2_{0.95,20} = 10.85$, $\chi^2_{\frac{\alpha}{2},2n} = \chi^2_{0.05,20} = 31.41$ og $\sum_{i=1}^{10} t_i = 7.5$ får vi 90% konfidensintervallet: $(\frac{10.85}{2\cdot 7.5}, \frac{31.41}{2\cdot 7.5}) = \underline{\underline{(0.72, 2.09)}}$.

d) Det følger fra sentralgrenseteoremet at en sum av uavhengige identisk fordelte variable $\sum_{i=1}^{n} T_i$ vil være tilnærmet normalfordelt.

$\mathrm{E}(\sum_{i=1}^{n} T_i) = \sum_{i=1}^{n} \mathrm{E}(T_i) = \frac{n}{\lambda}$ og $\mathrm{Var}(\sum_{i=1}^{n} T_i) \overset{uavh}{=} \sum_{i=1}^{n} \mathrm{Var}(T_i) = \sum_{i=1}^{n} \frac{1}{\lambda^2} = \frac{n}{\lambda^2}$, dvs

$$\frac{\sum_{i=1}^{n} T_i - \frac{n}{\lambda}}{\sqrt{\frac{n}{\lambda^2}}} \approx n(z;0,1)$$

Vi finner et tilnærmet $(1-\alpha)100\%$ konfidensintervall ved:

$$\begin{aligned} P(-z_{\frac{\alpha}{2}} < \frac{\sum_{i=1}^{n} T_i - \frac{n}{\lambda}}{\sqrt{\frac{n}{\lambda^2}}} < z_{\frac{\alpha}{2}}) = 1-\alpha \\ P(n - z_{\frac{\alpha}{2}}\sqrt{n} < \lambda \sum_{i=1}^{n} T_i < n + z_{\frac{\alpha}{2}}\sqrt{n}) &= 1-\alpha \\ P(\frac{n - z_{\frac{\alpha}{2}}\sqrt{n}}{\sum_{i=1}^{n} T_i} < \lambda < \frac{n + z_{\frac{\alpha}{2}}\sqrt{n}}{\sum_{i=1}^{n} T_i}) &= 1-\alpha \end{aligned}$$

Dvs et tilnærmet $(1-\alpha)100\%$ konfidensintervall for λ blir:

$$\underline{\underline{[\frac{n - z_{\frac{\alpha}{2}}\sqrt{n}}{\sum_{i=1}^{n} T_i}, \frac{n + z_{\frac{\alpha}{2}}\sqrt{n}}{\sum_{i=1}^{n} T_i}]}}$$

Innsatte data får vi tilnærmet 90% konfidensintervall: $(\frac{10-1.645\sqrt{10}}{7.5}, \frac{10+1.645\sqrt{10}}{7.5}) = \underline{\underline{(0.64, 2.03)}}$

Vi ser at overenstemmelsen med det eksakte intervallet er god. Intervallet avviker bare litt fra det eksakte intervallet over.

e)

Finner likelihoodfunksjonen.

$$\begin{aligned} L(\lambda; t_1, ..t_n) &= \prod_{i=1}^{n}(\lambda e^{-\lambda t_i}) = \lambda^n e^{-\lambda \sum_{i=1}^{n} t_i} \\ l(\lambda) &= \ln L(\lambda; t_1, ..t_n) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} t_i \\ \frac{dl(\lambda)}{d\lambda} &= \frac{n}{\lambda} - \sum_{i=1}^{n} t_i = 0 \qquad \Rightarrow \lambda = \frac{n}{\sum_{i=1}^{n} t_i} \end{aligned}$$

Dvs SME er gitt som: $\hat{\lambda} = \underline{\underline{\frac{n}{\sum_{i=1}^{n} T_i}}}$

Estimat: $\hat{\lambda} = \frac{10}{7.5} = \underline{\underline{1.33}}$.

For å finne forventningen til $\hat{\lambda}$, trenger vi $\mathrm{E}(\frac{1}{Z})$:

$$\begin{aligned} \mathrm{E}(\frac{1}{Z}) &= \int_0^\infty \frac{1}{z} \frac{z^{n-1}}{2^n \Gamma(n)} e^{-\frac{z}{2}} dz \\ &= \frac{1}{2(n-1)} \int_0^\infty \frac{z^{n-1-1}}{2^{n-1}\Gamma(n-1)} e^{-\frac{z}{2}} dz \\ &= \frac{1}{2(n-1)} \cdot 1 \end{aligned}$$

Det siste integralet er integralet over $\chi^2_{2(n-1)}$ fordelingen som er 1.

$$\mathrm{E}(\hat{\lambda}) = n\,\mathrm{E}(\frac{1}{\sum_{i=1}^{n} T_i}) = n\,\mathrm{E}(\frac{2\lambda}{2\lambda \sum_{i=1}^{n} T_i}) = 2n\lambda\,\mathrm{E}(\frac{1}{Z}) = 2n\lambda \cdot (\frac{1}{2(n-1)}) = \lambda \frac{n}{n-1}$$

Dvs $\hat{\lambda}$ er ikke forventningsrett.
Vi innfører $\lambda^* = \frac{n-1}{n}\hat{\lambda}$.

$$\mathrm{E}(\lambda^*) = \frac{n-1}{n}\,\mathrm{E}(\hat{\lambda}) = \frac{n-1}{n}\frac{n}{n-1}\lambda = \lambda$$

Dvs λ^* er en forventningsrett estimator for λ.