Løsningsforslag øving 5

Oppgaver fra boka:

8.26 / 8.5:10

La X være tid brukt av en tilfeldig kunde. Merk at vi ikke har oppgitt hvilken fordeling X har! Vi har imidlertid oppgitt at X har forventning $\mu = 3.2$ og standardavvik $\sigma = 1.6$. La videre $\bar{X} = \frac{1}{64}\sum_{i=1}^{64} X_i$ være gjennomsnittstiden for 64 tilfeldige kunder. Sentralgrenseteoremet (SGT) gir nå at:

$$Z = \frac{\bar{X} - \mathrm{E}(\bar{X})}{\sqrt{\mathrm{Var}(\bar{X})}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - 3.2}{1.6/8} = \frac{\bar{X} - 3.2}{0.2} \approx N(0,1)$$

dvs Z er tilnærmet standard normalfordelt.

a)

$$\begin{aligned} P(\bar{X} \le 2.7) &= P(\frac{\bar{X} - 3.2}{0.2} \le \frac{2.7 - 3.2}{0.2}) \\ &= P(Z \le -2.5) \\ &\overset{SGT}{\approx} \underline{\underline{0.0062}} \end{aligned}$$

b)

$$\begin{aligned} P(\bar{X} > 3.5) &= 1 - P(Z \le \frac{3.5 - 3.2}{0.2}) \\ &\overset{SGT}{\approx} 1 - 0.9332 \\ &= \underline{\underline{0.0668}} \end{aligned}$$

c)

$$\begin{aligned} P(3.2 \le \bar{X} \le 3.4) &= P(\bar{X} < 3.4) - P(\bar{X} < 3.2) \\ &= P(Z < \frac{3.4 - 3.2}{0.2}) - P(Z < \frac{3.2 - 3.2}{0.2}) \\ &\overset{SGT}{\approx} 0.8413 - 0.5 \\ &= \underline{\underline{0.3413}} \end{aligned}$$

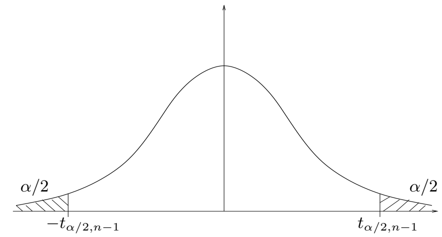
9.15 / 9.7: 15

Målinger av hardhet: $X_1, \ldots, X_n$ u.i.f. $\mathrm{N}(\mu, \sigma^2)$. Både μ og σ ukjente.

Konfidensintervall for μ:

$$\hat{\mu} = \bar{X}$$

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$



$$\begin{aligned} P(-t_{\alpha/2,n-1} \le T \le t_{\alpha/2,n-1}) &= 1 - \alpha \\ P(-t_{\alpha/2,n-1} \le \frac{\bar{X} - \mu}{S/\sqrt{n}} \le t_{\alpha/2,n-1}) &= 1 - \alpha \\ P(-t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \le \bar{X} - \mu \le t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}) &= 1 - \alpha \\ P(\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}) &= 1 - \alpha \end{aligned}$$

Dvs et $(1 - \alpha)100\%$ konfidensintervall for μ er gitt ved:

$$[\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}]$$

Observert:

$n = 12$, $\bar{x} = 48.50$, $s = 1.5$. $\alpha = 0.1 \Rightarrow t_{\alpha/2,n-1} = t_{0.05,11} = 1.796$. Innsatt gir dette 90% konfidensintervall for μ:

$$[48.50 - 1.796\frac{1.5}{\sqrt{12}}, 48.50 + 1.796\frac{1.5}{\sqrt{12}}] = \underline{\underline{[47.72, 49.28]}}$$

9.76 / 9.13: 6

Målinger av hardhet: $X_1, \ldots, X_n$ u.i.f. $\mathrm{N}(\mu, \sigma^2)$.

Konfidensintervall for σ^2, variansen i hardheten (som det spørres om i 9.13:6):

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\text{Teorem 8.4/tabell s. 27 gir at: } (n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\begin{aligned}
\Rightarrow \qquad P(\chi^2_{1-\alpha/2,n-1} \leq (n-1)\frac{S^2}{\sigma^2} \leq \chi^2_{\alpha/2,n-1}) &= 1-\alpha \\
P\left(\frac{\chi^2_{1-\alpha/2,n-1}}{(n-1)S^2} \leq \frac{1}{\sigma^2} \leq \frac{\chi^2_{\alpha/2,n-1}}{(n-1)S^2}\right) &= 1-\alpha \\
P\left(\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}} \geq \sigma^2 \geq \frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}\right) &= 1-\alpha \\
P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right) &= 1-\alpha
\end{aligned}$$

Konfidensintervall for σ (som det spørres om i 9.76):

$$P\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}}} \leq \sigma \leq \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}}\right) = 1-\alpha$$

Observert: $n = 12$ og $s = 1.5$

$\alpha = 0.10 \quad \Rightarrow \quad \chi^2_{0.05,11} = 19.675$ og $\chi^2_{0.95,11} = 4.575$.

90% konfidensintervall for σ^2: $\left[\frac{11\cdot 1.5^2}{19.675}, \frac{11\cdot 1.5^2}{4.575}\right] = \underline{\underline{[1.26, 5.41]}}$.

90% konfidensintervall for σ: $\left[\sqrt{\frac{11\cdot 1.5^2}{19.675}}, \sqrt{\frac{11\cdot 1.5^2}{4.575}}\right] = \underline{\underline{[1.12, 2.33]}}$.

10.25 / 10.10: 7

X="oljevolum i beholder"$\sim \mathrm{N}(\mu, \sigma^2)$ der både μ og σ er ukjente. Vi skal teste

$$H_0: \ \mu = 10 \qquad \text{mot} \qquad H_1: \ \mu \neq 10$$

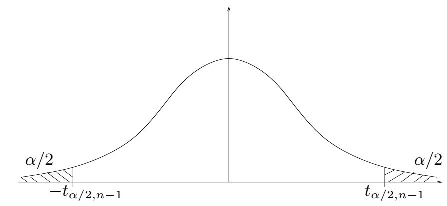
Når σ er ukjent har vi fra pensum at:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1} \text{ under } H_0$$

Med $\alpha = 0.01$ og $n = 10$ forkaster vi nullhypotesen dersom $T \geq t_{\alpha/2,n-1} = t_{0.005,9} = 3.25$ eller $T \leq -t_{\alpha/2,n-1} = -t_{0.005,9} = -3.25$

Observert: $\bar{x} = 10.06$, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2) = 0.0604$, dvs $s = 0.246$ og

$$t_{obs} = \frac{10.06 - 10}{0.246/\sqrt{10}} = 0.77$$



Dvs, $\underline{\underline{\text{vi forkaster ikke } H_0}}$ på 1% nivå. Dataene gir ingen grunn for å påstå at forventa volum er ulik 10.

10.68 / 10.13: 2

$$H_0: \ \sigma = 6 \qquad \text{mot} \qquad H_1: \ \sigma < 6$$

Dette er det samme som å teste:

$$H_0: \ \sigma^2 = 36 \qquad \text{mot} \qquad H_1: \ \sigma^2 < 36$$

$$\text{Estimator:} \qquad \hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Teorem 8.4/tabell s. 27 gir at:

$$(n-1)\frac{S^2}{\sigma_0^2} = \sum_{i=1}^{n}\frac{(X_i - \bar{X})^2}{\sigma_0^2} \sim \chi^2_{n-1} \quad \text{under } H_0$$

Vi forkaster H_0 dersom $(n-1)\frac{S^2}{\sigma_0^2} \leq \chi^2_{1-\alpha,n-1}$.

Med $\alpha = 0.05$ og $n = 20$ blir $\chi^2_{1-\alpha,n-1} = \chi^2_{0.95,19} = 10.117$

Observert:

$$(n-1)\frac{s^2}{\sigma_0^2} = (20-1)\frac{4.51^2}{36} = 10.74 > \chi^2_{0.95,19} = 10.117$$

dvs vi $\underline{\underline{\text{forkaster ikke } H_0}}$ på 5% nivå. Dataene gir ikke grunnlag for å hevde at $\sigma < 6$.

Dvs $\hat{\sigma} = s = 4.51$ er ikke så mye mindre enn 6 at det gir grunnlag for å hevde at $\sigma < 6$.

Oppgave 1:

a) Den eksakte fordelingen til X vil være en hypergeometrisk fordeling med parametre N, k og n der N er antall som stemmer, k er antallet som stemmer på Ap og $n = 1000$ er antallet som blir spurt i meningsmålingen. P.g.a. at $n << N$ vil X være tilnærmet binomisk fordelt med parametre $n = 1000$ og $p = k/N$.

b) Siden $np > 5$ og $n(1-p) > 5$ kan vi bruke tilnærming til normalfordeling:

$$\begin{aligned} P(X > 300) &= 1 - P(X \le 300) = 1 - P(Z \le \frac{300 + 0.5 - 1000 \cdot 0.275}{\sqrt{1000 \cdot 0.275 \cdot (1 - 0.275)}}) \\ &= 1 - P(Z \le 1.81) = 1 - 0.9649 = \underline{\underline{0.0351}} \end{aligned}$$

Dvs ca 3.5% sannsynlighet for at meningsmålingen indikerer en oppslutning på over 30% dersom virkeligheten er 27.5% oppslutning.

c) I binomisk fordeling har vi $f(x; p) = \binom{n}{x} p^x (1-p)^{n-x}$, og siden vi har gjort kun ett binomisk forsøk (kun en X) blir $L(p; x) = f(x; p)$, dvs

$$\begin{aligned} L(p;x) &= \binom{n}{x} p^x (1-p)^{n-x} \\ l(p;x) &= \ln L(p;x) = \ln\binom{n}{x} + x \ln p + (n-x)\ln(1-p) \\ \frac{\partial l(p;x)}{\partial p} &= \frac{x}{p} - \frac{n-x}{1-p} = 0 \\ &\Rightarrow \quad (1-p)x - p(n-x) = 0 \\ & \quad x - px - pn + px = 0 \\ &\Rightarrow \quad \underline{\underline{\hat{p} = \frac{X}{n}}} \end{aligned}$$

Sjekker at vi har funnet maksimum:

$$\frac{\partial^2 l(p; x_1, \ldots, x_n)}{\partial p^2} = -\frac{1}{p^2}x - \frac{1}{(1-p)^2}(n-x) < 0 \qquad \text{dvs maksimum!}$$

$$\begin{aligned} \mathrm{E}(\hat{p}) &= \mathrm{E}(\frac{X}{n}) = \frac{1}{n}\mathrm{E}(X) = \frac{1}{n}np = \underline{\underline{p}} \\ \mathrm{Var}(\hat{p}) &= \frac{1}{n^2}\mathrm{Var}(X) = \frac{1}{n^2}np(1-p) = \underline{\underline{\frac{p(1-p)}{n}}} \end{aligned}$$

d) $X \sim \mathrm{bin}(n, p)$. Konfidensintervall for p:

$$\hat{p} = \frac{X}{n}$$

$$\frac{\hat{p} - \mathrm{E}(\hat{p})}{\sqrt{\mathrm{Var}(\hat{p})}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{SGT}{\approx} \mathrm{N}(0,1)$$

For å forenkle utregningen av konfidensintervallet bruker vi videre tilnærmingen at for stor n er $p(1-p) \approx \hat{p}(1-\hat{p})$, dvs vi tar utgangspunkt i:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx \mathrm{N}(0,1)$$

$$\begin{aligned} P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) &= 1 - \alpha \\ P(-z_{\alpha/2} \le \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \le z_{\alpha/2}) &= 1 - \alpha \\ &\vdots \\ P(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}) &= 1 - \alpha \end{aligned}$$

Observert: $n = 1000$ og $x = 297$ gir $\hat{p} = \frac{297}{1000} = 0.297$.

$\alpha = 0.05 \quad \Rightarrow \quad z_{\alpha/2} = z_{0.025} = 1.96$.

Innsatt gir dette tilnærmet 95% konf. int. for p:

$$[0.297 - 1.96\sqrt{\frac{0.297 \cdot 0.703}{1000}}, 0.297 + 1.96\sqrt{\frac{0.297 \cdot 0.703}{1000}}] = \underline{\underline{[0.269, 0.325]}}$$

e)

$$H_0: \; p = 0.327 \qquad \text{mot} \qquad H_1: \; p < 0.327$$

Tar utgangspunkt i

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{\hat{p} - 0.327}{\sqrt{0.327(1-0.327)/n}} \approx N(0,1) \quad \text{under } H_0$$

Vi forkaster H_0 dersom $Z \le -z_\alpha = -z_{0.05} = -1.645$.

Observerte data gir:

$$z = \frac{0.297 - 0.327}{\sqrt{0.327(1-0.327)/1000}} = -2.02$$

dvs vi $\underline{\underline{\text{forkaster } H_0}}$ på 5% nivå. Dataene gir grunnlag for å hevde at $p < 0.327$, dvs at oppslutningen om Ap har avtatt.

Oppgave 2:

c) Siden $\hat{\mu} = \frac{6}{7}\bar{X} + \frac{1}{7}\bar{Y}$ er en lineærkombinasjon av uavhengige normalfordelte variable vil $\hat{\mu}$ være normalfordelt. I 1 a) (øving 4) regnet vi ut at $\mathrm{E}(\hat{\mu}) = \mu$ og $\mathrm{Var}(\hat{\mu}) = 0.000857$. Vi får dermed:

$$Z = \frac{\hat{\mu} - \mathrm{E}(\hat{\mu})}{\sqrt{\mathrm{Var}(\hat{\mu})}} = \frac{\hat{\mu} - \mu}{\sqrt{0.000857}} \sim N(0,1)$$

$$\begin{aligned} P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) &= 1 - \alpha \\ P(-z_{\alpha/2} \le \frac{\hat{\mu} - \mu}{\sqrt{0.000857}} \le z_{\alpha/2}) &= 1 - \alpha \\ P(-z_{\alpha/2}\sqrt{0.000857} \le \hat{\mu} - \mu \le z_{\alpha/2}\sqrt{0.000857}) &= 1 - \alpha \\ P(\hat{\mu} - z_{\alpha/2}\sqrt{0.000857} \le \mu \le \hat{\mu} + z_{\alpha/2}\sqrt{0.000857}) &= 1 - \alpha \end{aligned}$$

Innsatt $\alpha = 0.05 \quad \Rightarrow \quad z_{\alpha/2} = z_{0.025} = 1.96$ og $\hat{\mu} = \frac{6}{7} \cdot 6.12 + \frac{1}{7} \cdot 6.05 = 6.11$ gir dette at et 95% konfidensintervall for μ er gitt ved:

$$[6.11 - 1.96\sqrt{0.000857},\ 6.11 + 1.96\sqrt{0.000857}] = \underline{\underline{[6.05,\ 6.17]}}$$

d)

$$H_0: \quad \mu = 6.0 \qquad \text{mot} \qquad H_1: \quad \mu > 6.0$$

$$Z = \frac{\hat{\mu} - \mathrm{E}(\hat{\mu})}{\sqrt{\mathrm{Var}(\hat{\mu})}} = \frac{\hat{\mu} - 6.0}{\sqrt{0.000857}} \sim N(0,1) \ \text{ under } H_0$$

Vi forkaster H_0 dersom $Z \geq z_\alpha = z_{0.05} = 1.645$.

Observert:

$$z = \frac{\hat{\mu} - 6.0}{\sqrt{0.000857}} = \frac{6.11 - 6.0}{\sqrt{0.000857}} = 3.76 > 1.645$$

dvs vi $\underline{\underline{\text{forkaster } H_0}}$. Dataene gir grunnlag for å hevde at man har lyktes med å få pH-verdien over 6.0.

Oppgave 3:

a)

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

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

Estimat: $\hat{\lambda} = \frac{12}{108.5} = \underline{\underline{0.11}}$.

b) Vi begynner med å se på fordelingen til $Y = 2\lambda T_i$. Med $Y = 2\lambda T_i$ får vi $T_i = Y/(2\lambda) = w(Y)$ og dermed $w'(Y) = 1/(2\lambda)$ slik at transformasjonsformelen gir:

$$g(y) = f(w(y))|w\prime(y)| = \lambda e^{-\lambda(y/(2\lambda))} \frac{1}{2\lambda} = \underline{\underline{\frac{1}{2} e^{-y/2}}}$$

Dersom vi setter $\nu = 2$ i sannsynlighetstettheten til χ^2-fordelingen får vi:

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} e^{-x/2} \overset{\nu=2}{=} \frac{1}{2^{2/2}\Gamma(2/2)} x^{2/2-1} e^{-x/2} = \frac{1}{2} e^{-x/2}$$

dvs vi ser at fordelingen til Y er en χ^2_2-fordeling. Videre har vi at siden en sum av uavhengige χ^2-fordelte variable er χ^2-fordelt med parameter ("frihetsgrader") lik summen av parametrene i fordelingene til enkeltvariablene, så vil vi ha at $Z = 2\lambda \sum_{i=1}^n T_i = \sum_{i=1}^n 2\lambda T_i$ er $\underline{\underline{\chi^2_{2n}\text{-fordelt}}}$ (sum av n uavh. χ^2_2-fordelte variable).

c)

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

Dvs $\hat{\lambda}$ er $\underline{\underline{\text{ikke forventningsrett}}}$.

Vi prøver estimatoren $\lambda^* = \frac{n-1}{n}\hat{\lambda} = \frac{n-1}{\sum_{i=1}^n T_i}$.

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

Dvs $\lambda^* = \frac{n-1}{\sum_{i=1}^n T_i}$ er en forventningsrett estimator for λ.

Denne gir estimatet: $\lambda^* = \frac{11}{108.5} = 0.10$.

d)

Ut fra at

$$Z = 2\lambda \sum_{i=1}^{n} T_i = 2\lambda(n-1)\frac{\sum_{i=1}^n T_i}{n-1} = 2(n-1)\frac{\lambda}{\lambda^*} \sim \chi^2_{2n}$$

får vi

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

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

$$\underline{\underline{\left[\frac{\lambda^*}{2(n-1)}\chi^2_{1-\frac{\alpha}{2},2n},\ \frac{\lambda^*}{2(n-1)}\chi^2_{\frac{\alpha}{2},2n}\right]}}$$

Ved å sette inn $\alpha = 0.05$ som gir $\chi^2_{1-\frac{\alpha}{2},2n} = \chi^2_{0.975,24} = 12.401$, $\chi^2_{\frac{\alpha}{2},2n} = \chi^2_{0.025,24} = 39.364$ og $\lambda^* = \frac{n-1}{\sum_{i=1}^n T_i} = \frac{12-1}{108.5} = 0.101$ får vi 95% konfidensintervallet:

$$[\frac{0.101}{2 \cdot 11} \cdot 12.401,\ \frac{0.101}{2 \cdot 11} \cdot 39.364] = \underline{\underline{[0.057,\ 0.181]}}$$