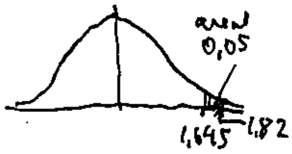
ST0201 2009V. Løsningsskisse

1a. $\hat p = 24/50 = 0{,}48$, $n\hat p(1-\hat p) = 50\cdot 0{,}48\cdot 0{,}52 = 12{,}48 \ge 5$. Bruker normaltilnærmelse.

Grenser for konf.-int.: $\hat p \pm z_{\alpha/2}\sqrt{\frac{\hat p(1-\hat p)}{n}}$. $\alpha = 0{,}05$, $z_{\alpha/2} = 1{,}960$ gir grenser $0{,}48 \pm 1{,}960\sqrt{\frac{0{,}48\cdot 0{,}52}{50}} = 0{,}48 \pm 0{,}14$. 95%-konf.int.: $[0{,}34,\ 0{,}62]$

b. Fra side 233 i læreboka: $n \ge \left(\frac{z_{\alpha/2}}{L}\right)^2 = \left(\frac{1{,}96}{0{,}1}\right)^2 = 384{,}16$. Minst 385 forsøk.

c. $H_0: p \le 0{,}357$, $H_1: p > 0{,}357$. Testobservator $Z = \frac{\hat p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$, der $p_0 = 0{,}357$. Forkaster H_0 når Z er stor. Her: $z = \frac{0{,}48 - 0{,}357}{\sqrt{\frac{0{,}357\cdot 0{,}643}{50}}} = 1{,}82$, som ligger i forkastn.området. Forkaster H_0 – på 0,05-nivå er det grunn til å si at den nye prosedyren har høyere suksessrate.



[P-verdi: $P(Z \ge 1{,}82) = 0{,}0344$.]

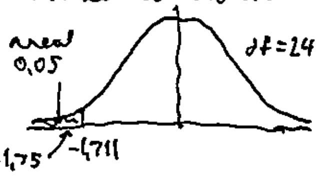
d. Hvis suksessraten er 0,5, så er $\frac{\hat p - 0{,}5}{\sqrt{\frac{0{,}5\cdot 0{,}5}{50}}}$ tilnærmet $N(0,1)$. Vi forkaster Ho hvis $\frac{\hat p - 0{,}357}{\sqrt{\frac{0{,}357\cdot 0{,}643}{50}}} > 1{,}645$, dvs. $\hat p > 0{,}4685$. Sanns. for dette er $P(\hat p > 0{,}4685)$

$$= P\left(\frac{\hat p - 0{,}5}{\sqrt{\frac{0{,}5\cdot 0{,}5}{50}}} > \frac{0{,}4685 - 0{,}5}{\sqrt{\frac{0{,}5\cdot 0{,}5}{50}}}\right) \approx P(Z > -0{,}45) = 1 - 0{,}3264 = 0{,}6736.$$

2a. Grenser for konf.int.: $\bar X \pm t_{\alpha/2}\frac{S}{\sqrt n}$. $\alpha = 0{,}05$, $t_{\alpha/2} = 2{,}201$ gir grenser $46{,}0 \pm 2{,}201\sqrt{\frac{116{,}0}{12}} = 46{,}0 \pm 6{,}84$. 95%-konf.int.: $[39{,}2,\ 52{,}8]$ $(df = 11)$

b. La forv.verdiene være hhv. μ_1 og μ_2. $H_0: \mu_1 \ge \mu_2$, $H_1: \mu_1 < \mu_2$. Testobs.: $T = \frac{\bar X - \bar Y}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$, der $\bar X$ og $\bar Y$ er gjennomsnittene av utvalgene med hhv. $n_1 = 12$ og $n_2 = 14$ observasjoner, og $S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$, der S_1^2 og S_2^2 er utvalgsvariansene.

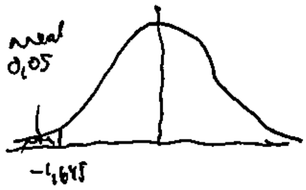
Forkaster Ho når T er liten. Kritisk verdi $-1{,}711$ $(df = 12+14-2 = 24)$.



Her: $s_p^2 = \frac{11\cdot 116{,}0 + 13\cdot 82{,}6}{24} = 97{,}91$. $t = \frac{46{,}0 - 52{,}8}{\sqrt{97{,}91}\sqrt{\frac{1}{12}+\frac{1}{14}}} = -1{,}75$.

Forkaster Ho. På 0,05-nivå er det grunnlag for å si at den andre populasjonen har en større forv. tilbakelagt distanse.

c. Uparet Mann-Whitney-Wilcoxon-test: Rangsum for den første populasjonen er $w = 131{,}5$ (observasjonen 48,3 er på delt 12-13-plass). Testobservator $Z = \frac{W - EW}{\sqrt{\mathrm{Var}\,W}}$, der $EW = \frac{n_1(n+1)}{2} = \frac{12\cdot 27}{2} = 162$ og $\mathrm{Var}\,W = \frac{n_1 n_2(n+1)}{12} = \frac{12\cdot 14\cdot 27}{12} = 378$. Z er tilnærmet $N(0,1)$ hvis Ho er riktig $(n_1 > 10,\ n_2 > 10)$, forkaster Ho når Z er liten. Her:



$z = \frac{131{,}5 - 162}{\sqrt{378}} = -1{,}57$. Forkaster ikke Ho på 0,05-nivå, dvs. motsatt konklusjon enn vi fikk i (b), noe som samsvarer med at ikkeparametriske tester generelt har lavere teststyrke.

3a. Kumulativ fordeling: $F(x) = \int_4^x \frac{1}{\theta} e^{-(t-4)/\theta}\,dt = \left[-e^{-(t-4)/\theta}\right]_4^x = 1 - e^{-(x-4)/\theta}$, $x>4$.

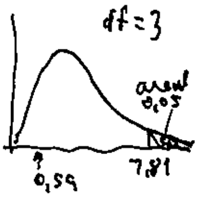
$P(4 < X < 4{,}2) = F(4{,}2) - F(4) = 1 - e^{-0{,}2/0{,}21} = 0{,}614,$

$P(4{,}2 \le X < 4{,}4) = F(4{,}4) - F(4{,}2) = e^{-0{,}2/0{,}21} - e^{-0{,}4/0{,}21} = 0{,}237,$

$P(4{,}4 \le X < 4{,}6) = F(4{,}6) - F(4{,}4) = e^{-0{,}4/0{,}21} - e^{-0{,}6/0{,}21} = 0{,}091,$

$P(X \ge 4{,}6) = 1 - F(4{,}6) = e^{-0{,}6/0{,}21} = 0{,}057$

hvis $\theta = 0{,}21$. Vi gjør en modelltest (χ^2-test). Forventede antall i de fire klassene er hhv. $100 \cdot 0{,}614 = 61{,}4$, $23{,}7$, $9{,}1$ og $5{,}7$. Vi forkaster H_0 hvis χ^2-observatoren er stor. Her:



$$\chi^2 = \frac{(64-61{,}4)^2}{61{,}4} + \frac{(21-23{,}7)^2}{23{,}7} + \frac{(10-9{,}1)^2}{9{,}1} + \frac{(5-5{,}7)^2}{5{,}7} = 0{,}59.$$

Kritisk verdi er 7,81 – forkaster ikke H_0.

b. $L = \prod \frac{1}{\theta} e^{-(x_i-4)/\theta} = \frac{1}{\theta^n} e^{-\frac{1}{\theta}\sum(x_i-4)} = \frac{1}{\theta^n} e^{-\frac{1}{\theta}(\sum x_i - 4n)}$,

$\ln L = -n \ln\theta - \frac{1}{\theta}(\sum x_i - 4n)$, $\frac{d \ln L}{d\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}(\sum x_i - 4n)$.

$\frac{d\ln L}{d\theta} = 0$: $n\theta = \sum x_i - 4n$, $\theta = \frac{\sum x_i}{n} - 4$. Dvs. sannsynlighetsmaks.estimatoren er $\hat\theta = \bar X - 4$.

Med tall fra oppgaven: $\hat\theta = 4{,}21 - 4 = 0{,}21$.

Legg merke til at $Y = X-4$ er eksp. fordelt med forventningsverdi θ:

$F_Y(y) = P(Y \le y) = P(X \le y+4) = F(y+4)$, dvs. $f_Y(y) = \frac{d}{dy}F(y+4) = f(y+4) = \frac{1}{\theta}e^{-y/\theta}$, dvs. eksp. ford. med forv. θ. Så $E\hat\theta = E(\bar X - 4) = E\bar X - 4 = E(X-4) = EY = \theta$, og

$\mathrm{Var}\,\hat\theta = \mathrm{Var}(\bar X - 4) = \mathrm{Var}\,\bar X = \frac{1}{n}\mathrm{Var}\,X = \frac{1}{n}\mathrm{Var}(X-4) = \frac{1}{n}\mathrm{Var}\,Y = \theta^2/n.$

c. $2 \cdot \frac{1}{\theta}\sum(X_i - 4)$ er χ^2-fordelt med $2n$ frihetsgrader (notat om gammafordeling).

Dvs. $P\left(\chi_{0{,}975} \le \frac{2}{\theta}(\sum X_i - 4n) \le \chi_{0{,}025}\right) = 0{,}95$. $\chi_{0{,}975} \le \frac{2}{\theta}(\sum X_i - 4n) \le \chi_{0{,}025}$ er ekvivalent med $\theta \le \frac{2(\sum X_i - 4n)}{\chi_{0{,}975}}$ og $\frac{2(\sum X_i - 4n)}{\chi_{0{,}025}} \le \theta$, dvs. $\frac{2n(\bar X - 4)}{\chi_{0{,}025}} \le \theta \le \frac{2n(\bar X - 4)}{\chi_{0{,}975}}$.

Med tallene fra oppgaven får vi da dette 95%-konfidensintervallet:

$$\left[\frac{200 \cdot 0{,}21}{241{,}1}, \frac{200 \cdot 0{,}21}{162{,}7}\right] = [0{,}174,\ 0{,}258]$$