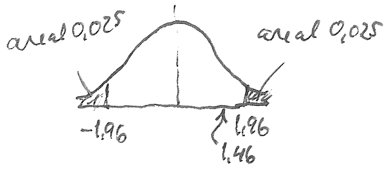
MNFSIB1, 2002H. Løsningsskisse

1(a) Punktestimat: $p' = \frac{188}{1000} = 0{,}188$. $1-\alpha$-konf.intervall har grenser $p' \pm z(\alpha/2)\sqrt{\frac{p'(1-p')}{n}}$.

90%-konf.int. har grenser $0{,}188 \pm 1{,}65\sqrt{\frac{0{,}188\cdot 0{,}812}{1000}} = 0{,}188 \pm 0{,}020$, dvs. fra 0,168 til 0,208.

(b) La p_1 = andel jenter og p_2 = andel gutter. $H_0: p_1 = p_2$, $H_a: p_1 \neq p_2$.

$p_1' = \frac{103}{500} = 0{,}206$, $p_2' = \frac{85}{500} = 0{,}170$. $p_p' = \frac{103+85}{1000} = 0{,}188$.



$z(\alpha/2) = z(0{,}025) = 1{,}96$. Forkastn.område: $< -1{,}96$ el. $> 1{,}96$.

$$z^* = \frac{p_1' - p_2'}{\sqrt{p_p'(1-p_p')\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0{,}206 - 0{,}170}{\sqrt{0{,}188\cdot 0{,}812\cdot\left(\frac{1}{500}+\frac{1}{500}\right)}} = 1{,}46,$$

ikke i forkastn.omr. Forkaster ikke H_0 på 0,05-signifikansnivå. Har ikke grunnlag for å si at andel jenter som nasker er forskj. fra andel gutter som nasker.

[Alt.: $P = P(z < -1{,}46) + P(z > 1{,}46) = 2\cdot P(z > 1{,}46) = 2\cdot(0{,}5 - 0{,}4279) = 0{,}144 > 0{,}05 = \alpha$.]

(c) La x = antall av de fire som nasker. x er binomisk fordelt med $n=4$, $p=0{,}20$.

$P(x \geq 2) = 1 - P(x<2) = 1 - P(x \leq 1) = 1 - P(x=0) - P(x=1) = 1 - \binom{4}{0}\cdot 0{,}2^0\cdot 0{,}8^4 - \binom{4}{1}\cdot 0{,}2^1 \cdot 0{,}8^3$

$= 1 - 0{,}4096 - 0{,}4096 = 0{,}1808$.

[Alt: $P(x\geq 2) = P(x=2) + P(x=3) + P(x=4) = \binom{4}{2}\cdot 0{,}2^2\cdot 0{,}8^2 + \binom{4}{3}\cdot 0{,}2^3\cdot 0{,}8 + \binom{4}{4}\cdot 0{,}2^4 \cdot 0{,}8^0$

$= 0{,}1536 + 0{,}0256 + 0{,}0016 = 0{,}1808$.]

(d) $P(x \geq 2 \mid x \geq 1) = \frac{P(x\geq 2 \text{ og } x \geq 1)}{P(x\geq 1)} = \frac{P(x \geq 2)}{P(x\geq 1)} = \frac{0{,}1808}{1-P(x=0)} = \frac{0{,}1808}{1-0{,}8^4} = 0{,}306$.

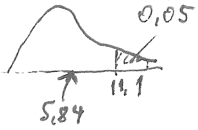
Hvis én bestemt har nasket, er hendelsen at minst to har nasket det samme som hendelsen at minst én av de tre andre har nasket. Antallet y av de tre andre som har nasket er binomisk fordelt med parametre 3 og 0,2. $P(y \geq 1) = 1 - P(y=0) = 1 - 0{,}8^3 = 0{,}488$.

(e) Antallet x som har nasket er binomisk fordelt med $n=40$ og $p=0{,}20$. For å begrense regnearbeidet, antar vi at $P(x \geq 15) \approx P(y \geq 14{,}5)$, der y er normalfordelt med μ og σ lik hhv. forvent.verdi og standardavvik for x, dvs. $\mu = np = 40\cdot 0{,}20 = 8$, $\sigma = \sqrt{np(1-p)} = \sqrt{40\cdot 0{,}20 \cdot 0{,}80} = \sqrt{6.4}$ $= 2{,}530$. $P(x \geq 15) \approx P(y \geq 14{,}5) = P\left(\frac{y-8}{\sqrt{6.4}} \geq \frac{14{,}5-8}{\sqrt{6.4}}\right) = P(z \geq 2{,}57)$ $= 0{,}5 - 0{,}4949 = 0{,}0051$ (der z er standard normalfordelt). (Forutsetningene $np = 40\cdot 0.2 = 8 \geq 5$ og $n(1-p) = 40\cdot 0{,}8 = 32 \geq 5$ er oppfylt.)

2(a) $P(12 < x < 14) = P\left(\frac{12-14}{4} < \frac{x-14}{4} < \frac{14-14}{4}\right) = P(-0{,}5 < z < 0) = P(0 < z < 0{,}5) = 0{,}1915$,

$P(x > 18) = P\left(\frac{x-14}{4} > \frac{18-14}{4}\right) = P(z > 1) = 0{,}5 - 0{,}3413 = 0{,}1587$ (tabell 3), der x = ant. timer sett på fjernsyn, z standard normalfordelt.

(b) H_0: multinomiske sanns. som i tabellen, H_a: de er forskjellige.



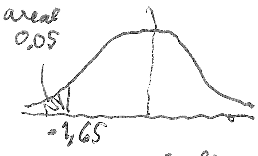
$$\chi^{2*} = \sum\frac{(O-E)^2}{E} = \frac{(12-16)^2}{16} + \frac{(16-15)^2}{15} + \frac{(25-19)^2}{19} + \frac{(22-19)^2}{19} + \frac{(9-15)^2}{15} + \frac{(16-16)^2}{16}$$

$= 1{,}000 + 0{,}067 + 1{,}895 + 0{,}474 + 2{,}400 + 0{,}000 = 5{,}84$. $df = 5$.

Kritisk verdi 11,1 (tabell 6). χ^{2*} ikke i forkastn.området. Forkaster ikke H_0. Har på 0,05-sig.nivå ikke grunnlag for å si at andelene er gale.

(c) La μ = gj.sn. tid en med tre års ekstra utd. ser på fjernsyn. $H_0: \mu = 14$, $H_a: \mu < 14$.

$\bar{x} = \Sigma x/n = 1345/105 = 12{,}81$. $s = \sqrt{\frac{\Sigma x^2 - (\Sigma x)^2/n}{n-1}} = \sqrt{\frac{18462 - 1345^2/105}{104}} = \sqrt{11{,}858} = 3{,}44$.



$t^* = \frac{\bar{x} - 14}{s/\sqrt{n}} = \frac{12{,}81 - 14}{3{,}44/\sqrt{105}} = -3{,}54$, $df = 104$. Forkast H_0 hvis t^* er liten.

Krit. verdi $-1{,}65$ (tabell 6). Forkaster H_0. På 0,05-signivå har vi grunnlag for å si at de med tre års ekstra utdannelse ser mindre enn 14 timer på fjernsyn.

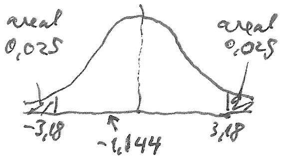
[Alt.: $P = P(t < -3{,}5) = 0{,}001$, tabell 7] Merk: σ ukjent i pop. av de med tre års ekstra utd.

3(a) $\bar{x} = \Sigma x/n_x = 44{,}7/4 = \underline{11{,}175}$, $s_x = \sqrt{\frac{\Sigma x^2 - (\Sigma x)^2/n}{n-1}} = \sqrt{\frac{502{,}61 - 44{,}7^2/4}{3}} = \sqrt{1{,}029} = \underline{1{,}014}$.

$\bar{y} = 47{,}7/4 = \underline{11{,}925}$, $s_y = \sqrt{\frac{570{,}89 - 47{,}7^2/4}{3}} = \sqrt{0{,}689} = \underline{0{,}830}$.

(b) La μ_x, μ_y være de forventede tidene. $H_0: \mu_x = \mu_y$, $H_a: \mu_x \neq \mu_y$ (dvs. hhv. $\mu_x - \mu_y = 0$, $\mu_x - \mu_y \neq 0$).

$t^* = \frac{\bar{x} - \bar{y}}{\sqrt{s_x^2/n_x + s_y^2/n_y}} = \frac{11{,}175 - 11{,}925}{\sqrt{1{,}029/4 + 0{,}689/4}} = -1{,}144$, $df = 3$ iflg. læreboka



Forkaster H_0 hvis t^* er liten el. stor. Kritiske verdier: $\pm 3{,}18$ (tabell 6).
Forkaster ikke H_0. Kan på 0,05-signif.nivå ikke påstå at metodene fører til forskjellig tid brukt på lekser. [$P \approx 2 \cdot P(t > 1{,}1) = 2 \cdot 0{,}176 = 0{,}35$.]

(c) $H_0: \mu_x = \mu_y = \mu_z$, H_a: minst én er forskjellig fra de andre. Variansanalyse.

Total sum og kvadratsum: $\Sigma t = \Sigma x + \Sigma y + \Sigma z = 44{,}7 + 47{,}7 + 51{,}2 = 143{,}6$, $\Sigma t^2 = \Sigma x^2 + \Sigma y^2 + \Sigma z^2 = 502{,}61 + 570{,}89 + 656{,}04 = 1729{,}54$.

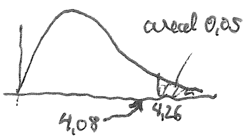
Kolonnesummer: $C_1 = \Sigma x = 44{,}7$, $C_2 = \Sigma y = 47{,}7$, $C_3 = \Sigma z = 51{,}2$.

$SS(\text{total}) = \Sigma t^2 - (\Sigma t)^2/n = 1729{,}54 - 143{,}6^2/12 = 11{,}127$.

$SS(\text{faktor}) = C_1^2/k_1 + C_2^2/k_2 + C_3^2/k_3 - (\Sigma t)^2/n = 44{,}7^2/4 + 47{,}7^2/4 + 51{,}2^2/4 - 143{,}6^2/12 = 5{,}292$

$SS(\text{feil}) = SS(\text{total}) - SS(\text{faktor}) = 11{,}127 - 5{,}292 = 5{,}835$. $df(\text{faktor}) = 2$, $df(\text{feil}) = 9$.

$F^* = \frac{SS(\text{faktor})/df(\text{faktor})}{SS(\text{feil})/df(\text{feil})} = \frac{5{,}292/2}{5{,}835/9} = 4{,}08$. Forkaster hvis F^* er stor.



Kritisk verdi: $F(2, 9, 0{,}05) = 4{,}26$. Forkaster ikke H_0.
Vi kan på 0,05-signivå ikke påstå at metodene fører til forskjellig tid til lekselesing.

(d) $SS(x) = \Sigma x^2 - (\Sigma x)^2/n = 56 - 24^2/12 = 8$, $SS(xy) = \Sigma xy - (\Sigma x)(\Sigma y)/n = 293{,}7 - 24 \cdot 143{,}6/12 = 6{,}5$.

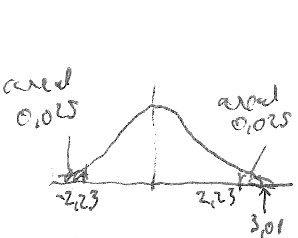
$b_1 = SS(xy)/SS(x) = 6{,}5/8 = 0{,}8125$, $b_0 = \frac{\Sigma y - b_1 \Sigma x}{n} = \frac{143{,}6 - 0{,}8125 \cdot 24}{12} = 10{,}3417$.

Estimert regresjonslinje: $\underline{\hat{y} = 10{,}34 + 0{,}81x}$

$s_e^2 = \frac{\Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy}{n-2} = \frac{1729{,}54 - 10{,}3417 \cdot 143{,}6 - 0{,}8125 \cdot 293{,}7}{10} = 0{,}5845$,

$s_{b_1}^2 = s_e^2/SS(x) = 0{,}5845/8 = 0{,}0731$, $s_{b_1} = \sqrt{s_{b_1}^2} = \sqrt{0{,}0731} = 0{,}2703$.

Regresjonsmodellen: $\mu_y = \beta_0 + \beta_1 x$. $H_0: \beta_1 = 0$, $H_a: \beta_1 \neq 0$.



$t^* = \frac{b_1 - 0}{s_{b_1}} = \frac{0{,}8125}{0{,}2703} = 3{,}01$, $df = n - 2 = 10$.

Forkaster H_0 når t^* er liten eller stor. Kritiske verdier: $\pm 2{,}23$ (tabell 6). Forkaster H_0. Vi har på 0,05-signifikansnivå grunn til å si at mengde veiledning har innvirkning på tida brukt på lekser.
[$P \approx 2P(t > 3{,}0) = 2 \cdot 0{,}007 = 0{,}014$, tabell 7.]