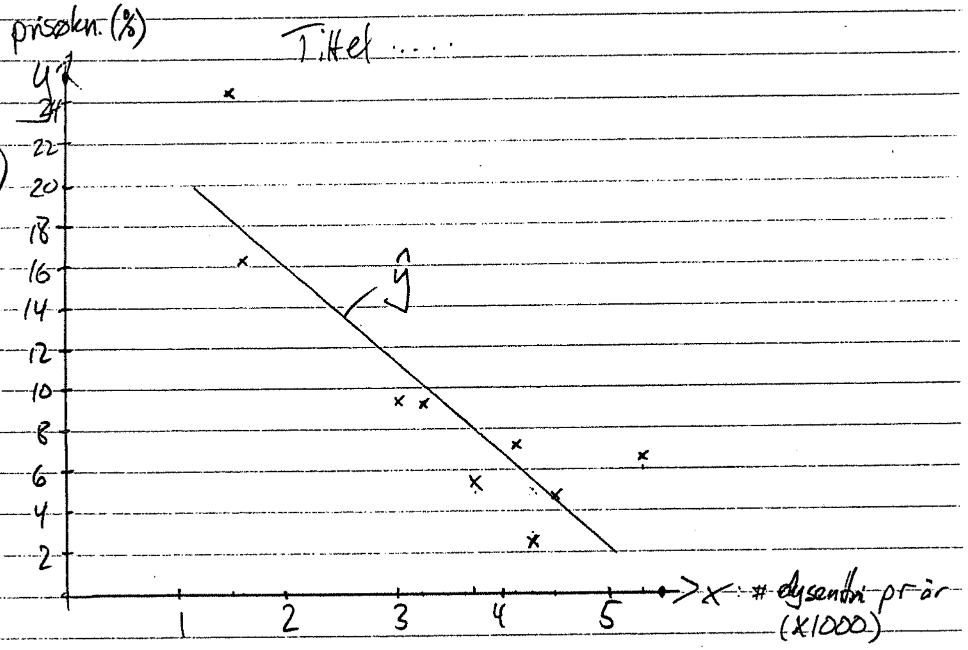
(S1)

Løsningsforslag
SIB1 - H97

1) A)



Beregn lineær korr. koeff:

$$r = \frac{SS(xy)}{\sqrt{SS(x)\,SS(y)}} \quad , \quad \begin{aligned} SS(xy) &= \Sigma xy - \frac{\Sigma x \Sigma y}{n} = -60.06 \\ SS(x) &= \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 13.22 \\ SS(y) &= 363.94 \end{aligned}$$

$$\Rightarrow r = -0.8658 \approx \underline{\underline{-0.87}}$$

B)

$$\left.\begin{aligned} b_1 &= \frac{SS(xy)}{SS(x)} = \underline{-4.5429} \\ b_0 &= \frac{1}{n}(\Sigma y - b_1 \Sigma x) = \underline{25.19} \end{aligned}\right\} \quad \hat{y} = \underline{\underline{25.2 - 4.54x}}$$

(S2)

SIB H-97

1) B) forts: Test om regresjonslinja bør benyttes dvs tester om $\beta_1 = ?$

$$H_0: \beta_1 = 0 \quad \text{mot} \quad H_a: \beta_1 \neq 0$$

Velger $\alpha = 0.05$

Testobservator: $t = \frac{b_1 - \beta_1}{s_{b_1}}$, $df = n - 2 = 7$

$$s_{b_1}^2 = \frac{s_e^2}{SS(x)} \quad , \quad s_e^2 = \frac{\Sigma y^2 - b_0 \Sigma y - b_1 \Sigma xy}{n-2} = \underline{13.05}$$

$$\Rightarrow s_{b_1}^2 = \underline{0.987}$$

$$\Rightarrow t^* = \frac{-4.5429}{\sqrt{0.987}} \approx \underline{4.57}$$

P-verdi; $p = 2 \cdot P(t > 4.57 \mid df = 7) < 2 \cdot 0.003 = 0.006$ (Ta...

$\Rightarrow p < \alpha = 0.05 \Rightarrow$ Forkaster H_0

Konklusjon: På et 0.05 signifikansnivå kan vi påstå at stigningstallet i regresjonslinja er ulik null

C) Vi finner her at vi kan forsvare å benytte regresjonslinja selv om de to variablene tilsynelatende ikke har noe med hverandre å gjøre.
Dette kan forklares med at regresjon ikke impliserer kausalitet dvs vi behøver nødvendigvis ikke ha en årsak-effekts sammenheng. En mulig forklaring her kan være at vi har en tredje ukjent variabel som påvirker begge.

(s.3) SIBI -H-97

(2)

A) Kommentarer:

- Gjennomføring;
  - Svært lite utvalg; 225 når vet-ikke-gruppen er trukket ifra
  - Ujevn avstand i tid, men det blir det nødvendigvis i en slik meningsmåling

:

- Presentasjon;
  Stort sett ok.

B) Standardfeil;

$$Ap: S_p = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.438(1-0.438)}{225}} = \underline{\underline{0.033}}$$

$$RV: S_{p'} = \sqrt{\frac{p'(1-p')}{n}} = \sqrt{\frac{0.019(1-0.019)}{225}} = \underline{\underline{0.0091}}$$

Sammenlign med feilmargin: 6% = 0.06

Norfaktas feilmargin er sannsynligvis målefeilen i et konfidensintervall.

Muligens et 95% KI for største andel i utvalget dvs $p'_{AP} = 0.438 \Rightarrow E = Z(\alpha/2) \cdot S_{p'} = 1.96 \cdot 0.033 = \underline{0.064}$

Eventuelt 95% KI for $p = \frac{1}{2}$ som gir maksimal feil;
$$E = Z(\alpha/2) \cdot S_{p=\frac{1}{2}} = 1.96 \sqrt{\frac{\frac{1}{2}\cdot\frac{1}{2}}{225}} = \underline{0.0653}$$

Evt. 90% KI for p=½ : E = 1.65 · √... = 0.055

(s.4) SIBI H-97

C) Ap fikk 40.1 % ved valget

$\Rightarrow H_0: p_{AP} = 0.401$ ved sept.måling

mot $H_a: p_{AP} \neq 0.401$ — " —

$\alpha = 0.05$

Antagelse: Tilfeldig utvalg og uavhengige målinger
Antar tilnærmet normalfordeling, siden $n = 225$ og både np og $n(1-p) \geq 5$

Observert: $n = 225$, $p'_{AP} = 0.438$

Testobservator: $$Z^* = \frac{p'_{AP} - p_{AP}}{\sqrt{\frac{p_{AP}(1-p_{AP})}{n}}} = \frac{0.438-0.401}{\sqrt{\frac{0.401(1-0.401)}{225}}} = \underline{1.13}$$

P-verdi; $p = 2 \cdot P(Z > 1.13) = 2 \cdot 0.1292 \approx \underline{0.26}$

$p > \alpha \Rightarrow$ Forkaster ikke H_0

Konklusjon; På et 0.05 signifikansnivå kan vi ikke påstå at det har vært meningsendringer i Nord-Trøndelag fra septembermålingen til valgdagen

(3)

$P(\text{Hvit}) = 0.86$, $P(\text{Svart}) = 0.14$

$P(\text{kvinne}) = P(\text{mann}) = 0.5$

A) $P(\text{Hvit og kvinne}) = \overset{\text{uavh.}}{P(\text{Hvit})P(\text{kvinne})} = 0.86 \cdot 0.5 = \underline{\underline{0.43}}$

$P(\text{Hvit og mann}) = 0.86 \cdot 0.5 = \underline{0.43}$

$P(\text{Svart og kvinne}) = 0.14 \cdot 0.5 = \underline{0.07}$

$P(\text{Svart og mann}) = 0.14 \cdot 0.5 = \underline{0.07}$

Oppgir så:

$P(\text{Mordoffer} \mid H.K.) = 1/495$
$P(M.O. \mid H.M) = 1/179$
$P(M.O. \mid S.K) = 1/132$
$P(M.O. \mid S.M) = 1/30$

$P(\text{Mordoffer}) = 1/163$

B) $P(HK \mid M.O) = \dfrac{P(HK \text{ og } MO)}{P(MO)} = \dfrac{P(HK) \cdot P(MO \mid HK)}{P(MO)}$

$= \dfrac{0.43 \cdot 1/495}{1/163} = 0.1416 \approx \underline{\underline{0.14}}$

$P(HM \mid MO) = \dfrac{P(HM) \cdot P(MO \mid HM)}{P(MO)} = \dfrac{0.43 \cdot 1/179}{1/163} = 0.3916 \approx \underline{\underline{0.39}}$

$P(SK \mid MO) = \dfrac{0.07 \cdot 1/132}{1/163} = 0.08644 \approx \underline{0.09}$

$P(SM \mid MO) = \dfrac{0.07 \cdot 1/30}{1/163} = 0.3803 \approx \underline{\underline{0.38}}$

(3)

c) Observerer 63 m.o. hvorav 5 HK, 17 HM, 9 SK og 32 SM

$\Rightarrow \chi^2$-test ang. andel innen de 4 kategoriene

H_0: Andelene er $p_{HK} = 0.14$, $p_{HM} = 0.39$, $p_{SK} = 0.09$ og $p_{SM} = 0.38$

mot H_a: Andelene ikke slik

Antagelse: Utvalget tilfeldig samlet inn.
Denne antagelsen brytes, men vi blir spurt om å teste om utvalget stemmer med hva vi forventer for et tilf. utv.

	O	E	O-E	$(O-E)^2/E$
HK	5	$p_{HK} \cdot 63 = 8.82$	-3.82	1.65
HM	17	24.57	-7.57	2.33
SK	9	5.67	3.33	1.96
SM	32	23.94	8.06	2.71
	63	63		8.65

Testobs; $\chi^2 = \sum \frac{(O-E)^2}{E} = \underline{8.65}$, $df = 4-1 = \underline{3}$

Velger $\alpha = 0.05$

P-verdi; $p = P(\chi^2 > 8.65 \mid df = 3)$

Tab 8: $\left.\begin{array}{l} P(\chi^2 > 7.82 \mid df=3) = 0.05 \\ P(\chi^2 > 9.35 \mid df=3) = 0.025 \end{array}\right\} \underline{0.025 < p < 0.05}$

$p < \alpha \Rightarrow$ Forkaster H_0

Konkl: På et 0.05 sign.nivå kan vi påstå at observasjonene ikke stemmer overens med

4 (s.7) SIBI H-97

A) Barnehagebarn:

$$\bar{X}_b = \frac{\sum x_b}{n_b} = \frac{53.1}{13} = \underline{\underline{4.08}}$$

$$S_b^2 = \frac{\sum x_b^2 - \frac{(\sum x_b)^2}{n_b}}{n_b - 1} = 5.34 \Rightarrow S_b = \underline{\underline{2.31}}$$

Hjemme: $\bar{X}_H = \frac{69.5}{12} = \underline{5.79}$

$$S_H^2 = \frac{469.15 - (69.5)^2/12}{12-1} = 6.06 \Rightarrow S_H = \underline{\underline{2.46}}$$

B) μ: gj.sn. antall fraværsdager per år.

$H_0: \mu_B - \mu_H \geq 0$ mot $H_a: \mu_B - \mu_H < 0$

$\alpha = 0.05$

Antar at utvalgene er uavhengige og tilfeldig valgt fra tilnærmet normalfordelte populasjoner

Siste antagelse, tiln. normalfordelte variable, er litt tvilsom iom at variabelen er begrenset i 0, og vi har lave verdier ⇒ kan ha en fordeling som er skjev med hale mot høyre, og dermed en dårlig normaltilnærming.

Test obs;

$$t^* = \frac{(\bar{x}_B - \bar{x}_H) - (\mu_B - \mu_H)}{\sqrt{\left(\frac{S_B^2}{n_B}\right) + \left(\frac{S_H^2}{n_H}\right)}}, \quad df = \min(n_B - 1, n_H - 1)$$

$$\Rightarrow t^* = \frac{4.08 - 5.79}{\sqrt{\frac{2.31^2}{13} + \frac{2.46^2}{12}}} \approx -1.79, \quad df = 11$$

(s.8) SIBI H.97

P-verdi; $P = P(t < -1.79 \mid df = 11) = P(t > 1.79 \mid df = 11)$

(Tab 6) ⇒ $0.05 < P < 0.10$

(Tab 7) ⇒ $P < 0.060$ (runder av nedover både for df (→10) og t^* (→ 1.7))

$P > \alpha = 0.05$ ⇒ Forkaster ikke H_0

Konklusjon: På et 0.05 sign.nivå har vi ikke tilstrekkelig bevis for å påstå at $\mu_B - \mu_H < 0$, dvs at "barnehagebarn" er mindre syke enn "hjemmeværende" etter at de har begynt på skolen.

C) 95% KI for $\mu_B - \mu_H$:

$$(\bar{x}_B - \bar{x}_H) \pm t(df, \alpha/2)\sqrt{\frac{S_B^2}{n_B} + \frac{S_H^2}{n_H}}, \quad df = \min(n_B - 1, n_H - 1)$$

$$t(11, 0.025) = 2.20$$

$$\Rightarrow (4.08 - 5.79) \pm 2.20 \cdot \sqrt{\frac{2.31^2}{13} + \frac{2.46^2}{12}}$$

$$\Rightarrow 95\%\ KI: \underline{\underline{(-3.81, 0.39)}}$$

Forklar konfidensintervallet ....