Fasit eksamen STK1000 2. desember 2008

Oppgave 1

a). Har brukt parret t-prosedyre. Eneggede tvillinger har samme genetiske utgangspunkt. Vi sammenligner effekten av to 'behandlinger' (oppvekst med og uten biologiske foreldre) på ellers like individer.

b). $H_0 : \mu = 0$

$H_a : \mu \neq 0$

$$t = \frac{\bar{x}}{s/\sqrt{n}} = \frac{-3.26}{8.81/\sqrt{31}} = -2.06$$

Her er $\bar{x}$ og s gjennomsnitt og standardavvik for differansene.

c). $P = 2P(T \geq 2.06)$ der $T \sim t(30)$. Fra tabell ser vi at $P < 2 \cdot 0.025 = 0.05$. Fra Minitab får vi mer nøyaktig at P=0.048. Med signifikansnivå $\alpha = 0.05$ forkaster vi H_0 og konkluderer med signifikant forskjell i forventet IQ mellom de som har vokst opp hos biologiske foreldre og de som ikke har det.

d). Tilstrekkelig å svare at de som har vokst opp hos andre enn biologisk familie har høyere forventet IQ enn de som har vokst opp hos biologisk familie. Evt. skrive at dette stemmer med at vi forkastet H_0 mot tosidig alternativ på nivå 0.05 og at $\bar{x}$ er negativ.

Bruker $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$ der $t^* = 2.75$ finnes fra tabell (99%, 30 frihetsgrader). Innsatt får vi intervallet (-7.61, 1.085). Dette intervallet er lenger enn 90%-intervallet fordi vi har høyere konfidenskoeffisient. Kan kommentere at intervallet nå inneholder 0 (så ingen forkastning av nullhypotesen på nivå $\alpha = 0.01$).

Oppgave 2

a). Uten belegg: $(\mu_1, \sigma_1, n_1, \bar{x}_1, s_1)$. Med belegg $(\mu_2, \sigma_2, n_2, \bar{x}_2, s_2)$.

$H_0 : \mu_1 = \mu_2$ 'Ingen forskjell'

$H_a : \mu_1 > \mu_2$ 'Lavere forventet konsentrasjon med belegg'

Bruker en to-utvalgs-t-test fordi vi sammenligner to uavhengige populasjoner. Må også spesifisere om man velger å anta lik varians ($\sigma_1 = \sigma_2$) eller ikke. Begge godtas. Merk at de to empiriske standardavvikene er nesten like.

Med antakelsen $\sigma_1 = \sigma_2$ (Pooled 2-Sample-t):

Finner pooled standardavvik $s_P = 6.27$.

$$t = \frac{\bar{x_1} - \bar{x_2}}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 2.52$$

$P = P(T \geq 2.52)$ der $T \sim t(45)$. Antall frihetsgrader kommer fra $n_1 + n_2 - 2$. Fra tabell ser vi at $P < 0.01$ når antall frihetsgrader er 40. Med signifikansnivå $\alpha = 0.05$ forkaster vi H_0 og konkluderer med at belegget gir lavere forventet konsentrasjon av virkemiddelet etter 12 timer.

Uten antakelse om $\sigma_1 = \sigma_2$ (2-Sample-t):

$$t = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 2.52$$

$P = P(T \geq 2.52)$ der $T \sim t(22)$. Antall frihetsgrader kommer fra den minste av $n_1 - 1$ og $n_2 - 1$. Fra tabell ser vi at $P < 0.01$ når antall frihetsgrader er 22. Samme konklusjon.

b). 2-utvalgs-t-prosedyrer (og pooled-t) robust mot ikkenormalitet spesielt når n_1 og n_2 er tilnærmet like store. Se s. 493 i boken.

Oppgave 3

a). Y_i armspenn, responsvariabel, x_i høyde, forklaringsvariabel, $i = 1, ..., 8$. Enkel lineær regresjons-modell:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

der vi antar at ϵ_i er uavhengige og normalfordelte $N(0, \sigma)$ for alle i.

Parametre i modellen er β_0, β_1 og σ, estimatene for disse finner vi fra utskriften som $b_0 = -6.53$, $b_1 = 1.0976$, og $S = 1.26513$.

For $x = 66$ får vi forventet armspenn $\hat{\mu}_y = 65.91$ tommer (evt. 66.1 ved innsetting i armspan = -6.5 + 1.10 height).

b). C% konfidensintervall for β_1: $b_1 \pm t^* SE_{b_1}$. For 95% og 6 frihetsgrader $(n-2)$ finner vi $t^* = 2.447$. $SE_{b_1} = 0.1536$ fra utskriften. Gir intervallet (0.7217, 1.4735). Inneholder 1, så ingen grunn til å påstå at Leonardo tok feil (Ingen forkastning av hypotesen $H_0 : \beta_1 = 1$ på nivå $\alpha = 0.05$).

c). $r^2 = 0.895$ stor. Må også kommentere residualplott og normalfordelingsplott. r^2 hadde blitt den samme, korrelasjonen er ikke avhengig av hva som er responsvariabel og hva som er forklaringsvariabel.