Løsningsforslag (ST1201/ST6201 høst 2015)

1.

Fra “Tabeller og formler ...” er sannsynlighetstettheten (ST) til en χ^2-fordelt stokastisk variabel X med $2n$ frihetsgrader (n=1,2,...) gitt som

$$f_X(x) = c_n x^{n-1} e^{-x/2}$$

for $x \geq 0$ og 0 ellers, hvor $c_n = 1/(2^n(n-1)!)$. Siden $f_X(x)$ er en ST, må det gjelde at

$$c_n \int_0^\infty x^{n-1} e^{-x/2} dx = 1, \quad n = 1, 2, ...$$

Dermed er

$$\mathrm{E}(X^{-1}) = c_n \int_0^\infty x^{-1} x^{n-1} e^{-x/2} dx = c_n \int_0^\infty x^{(n-1)-1} e^{-x/2} dx =$$

$$= \frac{c_n}{c_{n-1}} c_{n-1} \int_0^\infty x^{(n-1)-1} e^{-x/2} dx = \frac{c_n}{c_{n-1}} = \frac{1}{2(n-1)}, \quad n = 2, 3, ...$$

Tisvarende,

$$\mathrm{E}(X^{-2}) = c_n \int_0^\infty x^{-2} x^{n-1} e^{-x/2} dx = c_n \int_0^\infty x^{(n-2)-1} e^{-x/2} dx =$$

$$= \frac{c_n}{c_{n-2}} c_{n-2} \int_0^\infty x^{(n-2)-1} e^{-x/2} dx = \frac{c_n}{c_{n-2}} = \frac{1}{4(n-1)(n-2)}, \quad n = 3, 4, ...$$

2.

a) Testes

$H_0 : \mu = 30 = \mu_0$

$H_1 : \mu > 30$

Testobservator

$$T = \sqrt{n} \frac{\bar{Y} - \mu_0}{S} \sim t_{n-1}$$

(under H_0). I vårt tilfelle

$$t = \sqrt{10} \frac{32.16 - 30}{\sqrt{233.324/9}} = 1.34$$

$$t_{0.05,9} = 1.833$$

H_0 forkastes ikke.

b) $(1-\alpha)$-konfidensintervall er

$$\left[\bar{Y} - t_{\alpha/2,n-1} \frac{S}{\sqrt{n}}, \bar{Y} + t_{\alpha/2,n-1} \frac{S}{\sqrt{n}} \right].$$

I vårt tilfelle er det $[28.5, 35.8]$.

c)

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = -1.05$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = 99.66$$

Testes

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 < 0$

Testobservator

$$T = \sqrt{\sum (x_i - \bar{x})^2} \frac{\hat{\beta}_1}{S}$$

der

$$S^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2.$$

Under H_0, $T \sim t_{n-2}$. I vrt tilfelle $T = -2.71$. Kritisk verdi $-t_{0.05,8} = -1.86$. H_0 forkastes.

3.

a) $\bar{X} - \bar{Y}$ er en lineærkombinasjon av X-ene og Y-ene som er uavhengige og normalfordelt. Dermed blir også $\bar{X} - \bar{Y}$ normalfordelt.

$$\mathrm{E}(\bar{X} - \bar{Y}) = \mathrm{E}(\bar{X}) - \mathrm{E}(\bar{Y}) = \mu_x - \mu_y$$

$$\mathrm{Var}(\bar{X} - \bar{Y}) = \mathrm{Var}(\bar{X}) + \mathrm{Var}(\bar{Y}) = \frac{\sigma^2}{n} + \frac{k\sigma^2}{m} = \sigma^2 \left(\frac{1}{n} + \frac{k}{m} \right)$$

b)

$$\bar{X} - \bar{Y} \sim N(\mu_x - \mu_y, \sigma^2(1/n + k/m))$$

derfor

$$\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sigma\sqrt{1/n + k/m}} \sim N(0,1)$$

og

$$\mathrm{P}\left(-z_{\alpha/2} \leq \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sigma\sqrt{1/n + k/m}} \leq z_{\alpha/2} \right) = 1 - \alpha.$$

Intervallet blir

$$\left[\bar{X} - \bar{Y} - z_{\alpha/2}\sigma\sqrt{\frac{1}{n} + \frac{k}{m}}, \bar{X} - \bar{Y} + z_{\alpha/2}\sigma\sqrt{\frac{1}{n} + \frac{k}{m}} \right].$$

c)

$$\mathrm{E}(S_p^2) = \frac{\sigma^2}{n+m-2} \mathrm{E}\left(\frac{(n-1)S_x^2}{\sigma^2} \right) + \frac{\sigma^2}{n+m-2} \mathrm{E}\left(\frac{(m-1)S_y^2}{k\sigma^2} \right) =$$

$$= \frac{\sigma^2}{n+m-2}(n-1) + \frac{\sigma^2}{n+m-2}(m-1) = \sigma^2.$$

$(n+m-2)S_p^2/\sigma^2$ er en sum av to uavhengige χ^2-fordelte variabler:

$$\frac{(n+m-2)S_p^2}{\sigma^2} = \frac{(n-1)S_x^2}{\sigma^2} + \frac{(m-1)S_y^2}{k\sigma^2}$$

derfor er den også χ^2-fordelt.

T kan skrives på formen

$$T = \frac{Z}{\sqrt{V/(n+m-2)}}$$

der

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sigma\sqrt{1/n + k/m}} \sim N(0,1)$$

og

$$V = \frac{(n+m-2)S_p^2}{\sigma^2} \sim \chi^2_{n+m-2}.$$

Siden Z er kun en funksjon av $\bar{X}$ og $\bar{Y}$, mens S_p^2 kun er en funksjon av S_x^2 og S_y^2, er Z og V uavhengige. Per definisjon av t-fordeling blir dermed T t-fordelt med $n+m-2$ frihetsgrader.

d) Vi har at

$$\mathrm{P}(-t_{\alpha/2,n+m-2} \leq T \leq t_{\alpha/2,n+m-2}) = 1-\alpha$$

eller

$$\mathrm{P}\left(\bar{X} - \bar{Y} - t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n} + \frac{k}{m}} \leq \mu_x - \mu_y \leq \bar{X} - \bar{Y} - t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n} + \frac{k}{m}}\right) =$$

$$= 1-\alpha$$

og intervallet er

$$\left[\bar{X} - \bar{Y} - t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n} + \frac{k}{m}}, \bar{X} - \bar{Y} - t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n} + \frac{k}{m}}\right].$$

Lengden av intervallet blir

$$L = 2t_{\alpha/2,n+m-2}S_p\sqrt{\frac{1}{n} + \frac{k}{m}} =$$

$$= 2t_{\alpha/2,n+m-2}\sqrt{\frac{(n+m-2)S_p^2}{\sigma^2}}\sqrt{\frac{\sigma^2}{n+m-2}\left(\frac{1}{n} + \frac{k}{m}\right)}.$$

Vi får dermed at

$$\mathrm{E}(L) = 2t_{\alpha/2,n+m-2}\mathrm{E}\left(\sqrt{\frac{(n+m-2)S_p^2}{\sigma^2}}\right)\sqrt{\frac{\sigma^2}{n+m-2}\left(\frac{1}{n} + \frac{k}{m}\right)}.$$

Siden $(n+m-2)S_p^2/\sigma^2 \sim \chi^2_{n+m-2}$ må $\mathrm{E}((n+m-2)S_p^2/\sigma^2)$ kun være en funksjon av summen $n+m$ og ikke n og m hver for seg. For å minimere lengden gitt at $n+m=N$ er fiksert må vi dermed minimere

$$g(t) = \frac{1}{t} + \frac{k}{N-t}$$

med hensyn på t. Deriverer og setter lik null. Finner $t = \frac{N}{1+\sqrt{k}}$. Hvis dette er et heltall, er svar

$$n = \frac{N}{1+\sqrt{k}}, \quad m = N - \frac{N}{1+\sqrt{k}}.$$

Hvis t ikke er et heltall må man forsøke frem med avrunding nedover og oppover for å bestemme hvilke av disse to som gir det korteste intervallet.

4. We consider the null hypothesis $H_0 : \mu_a - \mu_m = 5$ and the alternative hypothesis $H_1 : \mu_a - \mu_m < 5$.

The observations come naturally in pairs, and there is reason to believe that the March and the April temperatures of a year are dependent. Therefore we choose a paret test, which is performed as a single sample test using the differences $d_i = y_i - x_i$. Under the null hypothesis, the test statistic

$$T = \sqrt{12}\frac{\bar{D} - 5}{S_D}$$

has the t distribution with 11 degrees of freedom. A small value of T is indicative of H_1, and the critical value is $-t_{0.05} = -1.769$.

With our data, $\bar{d} = \bar{y} - \bar{x} = 4.86$, $S_D^2 = 7.39$. So, $t = -0.18$, which is not in the critical region, and we do not reject H_0. At the 0.05 significance level there is not evidence to state that $H_1 : \mu_a - \mu_m < 5$.

5.

$$\mathrm{Cov}(2X + Y, X - Y) = 2\mathrm{Var}(X) - \mathrm{Cov}(X, Y) - \mathrm{Var}(Y) = 0$$