Løsningsforslag (TMA4255 August 2017)

1.
a) Evidently $n = 20$.

$$\sum_{i=1}^{n} Y_i = \text{Mean(C2)} \cdot n = 0.341 \cdot 20 = 6.82.$$

Since

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2,$$

we obtain

$$\sum_{i=1}^{n} Y_i^2 = (n-1) \cdot \text{StDev(C2)}^2 + n \cdot \text{Mean(C2)}^2 = 19 \cdot 1.025^2 + 20 \cdot 0.341^2 = 22.288.$$

[Pooled StDev = 1,016] is the average of numbers in the column StDev.

The hypothesis $H_0 : \mu_X = \mu_Y = \mu_Z$ is rejected because the P-value is less than the significance level.

b) The test statistic is

$$T = \sqrt{n}\frac{\bar{X}}{S_X} = \sqrt{n}\frac{\text{Mean(C1)}}{\text{StDev(C1)}} = -1.35.$$

The test: if $T < -t_{0.05,19}$, then H_0 is rejected. Since $-t_{0.05,19} = -1.729$, H_0 is not rejected.

c) We use the Tukey test for a multiple comparison. The three intervals, given in the output, are $[L1, U1]$, $[L2, U2]$, $[L3, U3]$, with

$$L1 = \text{Mean(C2)} - \text{Mean(C1)} - q(0.05, 3, 57)\frac{\text{Pooled StDev}}{\sqrt{n}}$$

$$U1 = \text{Mean(C2)} - \text{Mean(C1)} + q(0.05, 3, 57)\frac{\text{Pooled StDev}}{\sqrt{n}}$$

$$L2 = \text{Mean(C3)} - \text{Mean(C1)} - q(0.05, 3, 57)\frac{\text{Pooled StDev}}{\sqrt{n}}$$

$$U2 = \text{Mean(C3)} - \text{Mean(C1)} + q(0.05, 3, 57)\frac{\text{Pooled StDev}}{\sqrt{n}}$$

$$L3 = \text{Mean(C3)} - \text{Mean(C2)} - q(0.05, 3, 57)\frac{\text{Pooled StDev}}{\sqrt{n}}$$

$$U3 = \text{Mean(C3)} - \text{Mean(C2)} + q(0.05, 3, 57)\frac{\text{Pooled StDev}}{\sqrt{n}},$$

where $q(\alpha, k, N)$ is such a value that

$$P(Q_{k,N} \geq q(\alpha, k, N)) = \alpha,$$

where the random variable $Q_{k,N}$ has the studentized range distribution with k and N degrees of freedom. According to the Tukey test, H_0 is rejected if the corresponding interval does not contain zero. Thus, $H_0 : \mu_X = \mu_Z$ and $H_0 : \mu_Y = \mu_Z$ are rejected while $H_0 : \mu_X = \mu_Y$ is not regected.

2.

a) Let X be a random variable having the normal distribution with the expectation μ and the variance σ^2/n. Denote

$$p = P(X < \mu_0 - 3\sigma/\sqrt{n}) + P(X > \mu_0 + 3\sigma/\sqrt{n}).$$

Let N be the number of the sample (following the shift) where the out-of-control situation is detected for the first time. Then

$$P(N = 1) = p, P(N = 2) = (1 - p)p, ..., P(N = k) = (1 - p)^{k-1}p, ...$$

In other words, N has the geometric distribution with parameter p. Its expectation, which is the average number of samples required (following the shift) to detect the out-of-control situation, is

$$EN = \frac{1}{p}.$$

We have

$$P(X > \mu_0 + 3\sigma/\sqrt{n}) = P\left(\sqrt{n}\frac{X - \mu}{\sigma} > \sqrt{n}\frac{\mu_0 - \mu}{\sigma} + 3\right) = P(Z > 1),$$

where Z has the standard normal distribution. Similarly

$$P(X < \mu_0 - 3\sigma/\sqrt{n}) = P(Z < -5).$$

$P(Z < -5)$ is negligible with respect to $P(Z > 1)$, therefore

$$p \approx P(Z > 1) = 0.1587,$$

and

$$EN = 6.3.$$

3.

a) $\bar{X} - \bar{Y}$ is a linear combination of X-s and Y-s which are independent and normally distributed. Hence $\bar{X} - \bar{Y}$ is normally distributed.

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

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

So,

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

therefore

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

and

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

The interval is

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

b)

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

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

$(n+m-2)S_p^2/\sigma^2$ is a sum of two independent χ^2-distributed variables:

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

therefore it is also χ^2-distributed.

T can be written in the form

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

where

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

and

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

Since Z is a function of $\bar{X}$ and $\bar{Y}$, while S_p^2 is a function of S_x^2 and S_y^2, variables Z and V are independent. Therefore T has t-distribution with $n+m-2$ degrees of freedom.

c) We have

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

or

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

$$= 1 - \alpha$$

therefore the interval is

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

4.

a) Generally, the symmetric $100(1-\alpha)\%$ confidence interval is

$$\left[\hat{\beta}_1 - t_{\frac{\alpha}{2},n-2}\frac{S}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}, \hat{\beta}_1 + t_{\frac{\alpha}{2},n-2}\frac{S}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}}\right].$$

Since distribution of $(n-2)S^2/\sigma^2$ has 18 degrees of freedom, $n = 20$. Therefore $t_{0.025,18} = 2.101$, $t_{0.005,18} = 2.878$, and the symmetric 99% confidence interval is

$$\left[1 - \frac{2.878}{2.101}0.5, 1 + \frac{2.878}{2.101}0.5\right] = [0.315, 1.685].$$