TMA4255 Applied Statistics
Solution to Exercise 8

Problem 1

a) Main effect of z_1 = expected average response when z_1 is on the high level minus the expected average response when z_1 is on the low level.

$$\begin{aligned}\frac{y_4+y_2}{2}-\frac{y_3+y_1}{2} &= \frac{(\beta_0+\beta_1+\beta_2+\beta_{12})+(\beta_0+\beta_1-\beta_2-\beta_{12})}{2} \\ &- \frac{(\beta_0-\beta_1+\beta_2-\beta_{12})+(\beta_0-\beta_1-\beta_2+\beta_{12})}{2} = 2\beta_1\end{aligned}$$

Main effect of z_2 = expected average response when z_2 is on the high level minus the expected average response when z_2 is on the low level.

$$\begin{aligned}\frac{y_4+y_3}{2}-\frac{y_2+y_1}{2} &= \frac{(\beta_0+\beta_1+\beta_2+\beta_{12})+(\beta_0-\beta_1+\beta_2-\beta_{12})}{2} \\ &- \frac{(\beta_0+\beta_1-\beta_2-\beta_{12})+(\beta_0-\beta_1-\beta_2+\beta_{12})}{2} = 2\beta_2\end{aligned}$$

The interaction between z_1 and z_2: 1) half the main effect of z_1 when z_2 is on the high level minus 2) half the main effect of z_1 when z_2 is on the low level.

1) the main effect of z_1 when z_2 is on the high level:

$$\frac{(\beta_0+\beta_1+\beta_2+\beta_{12})-(\beta_0-\beta_1+\beta_2-\beta_{12})}{1} = 2\beta_1+2\beta_{12}$$

2) the main effect of z_1 when z_2 is on the low level:

$$\frac{(\beta_0+\beta_1-\beta_2-\beta_{12})-(\beta_0-\beta_1-\beta_2+\beta_{12})}{1} = 2\beta_1-2\beta_{12}$$

Then, the interaction between z_1 and z_2:

$$\frac{2\beta_1+2\beta_{12}}{2}-\frac{2\beta_1-2\beta_{12}}{2} = 2\beta_{12}$$

b) Main effect of z_1 while keeping z_2 at low level (we have already calculated above):

$$\frac{(\beta_0+\beta_1-\beta_2-\beta_{12})-(\beta_0-\beta_1-\beta_2+\beta_{12})}{1} = 2\beta_1-2\beta_{12}$$

Main effect of z_1 while keeping z_2 at high level (we have already calculated above):

$$\frac{(\beta_0+\beta_1+\beta_2+\beta_{12})-(\beta_0-\beta_1+\beta_2-\beta_{12})}{1} = 2\beta_1+2\beta_{12}$$

c) Based on the results in **b** we see that main effect of z_1 when z_2 is at its low level is $2\beta_1 - 2\beta_{12}$, and main effect of z_1 when z_2 is at its high level is $2\beta_1 + 2\beta_{12}$. If the interaction between z_1 and z_2 is zero we can find the main effect of z_1 by fixing the other factors at a given level. But, when there are interactions present, just fixing one factor at a given level will not give us estimate of the main effect, but the main effect and the interaction effect (as shown above). Therefore, we do not fix one factor at one level and vary the other factor in DOE!

Problem 2

a) **Two-sample T-test:**

We assume that

- $X_1, \ldots, X_n$, $Y_1, \ldots, Y_m$, $X_j \sim N(\mu_X, \sigma_X^2)$, $Y_j \sim N(\mu_Y, \sigma_Y^2)$, $n = 10$, $m = 8$.
- $\sigma_X^2 = \sigma_Y^2$

```
Two-Sample T-Test and CI: X_i; Y_i

Two-sample T for X_i vs Y_i

N     Mean  StDev  SE Mean
X_i  10  5201,3   10,2      3,2
Y_i   8  5182,0   19,6      6,9


Difference = mu (X_i) - mu (Y_i)
Estimate for difference:  19,3000
95% CI for difference:  (4,1579; 34,4421)
T-Test of difference = 0 (vs not =): T-Value = 2,70  P-Value = 0,016  DF = 16
Both use Pooled StDev = 15,0584
```

Explanation of the result from Minitab:

- N: The number of observations in each column.
- MEAN: average$= \frac{1}{N}\sum_{j=1}^{N} X_j = \bar{X}$.
- STDEV: $S = \sqrt{\frac{1}{n-1}\sum_{j=1}^{N}(X_j - \bar{X})^2}$.
- SE MEAN: standard deviation for $\bar{X}$, this is equall to$\frac{S}{\sqrt{N}}$. (correspondingly for Y.)
- 95 PCT CI: 95 % confidence interval for $(\mu_X - \mu_Y)$.

  The T-statistic is given by

$$T = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{S^2}{n} + \frac{S^2}{m}}} \sim T_{n+m-2} = T_{16}$$

  (Student-T-distributed with 16 degrees of freedom.) Here S^2 is pooled-stdev (see page 308) i.e. estimated variance under the assumption that the two samples have the same variance:

$$S = \sqrt{\frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}$$

$$= \sqrt{\frac{\sum_{j=1}^{n}(X_j - \bar{X})^2 + \sum_{j=1}^{n}(Y_j - \bar{Y})^2}{n+m-2}} = 15.1$$

To find a 95% confidence interval we set ut:

$$P\left(-t_{0.025,16} \leqslant \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\frac{1}{n} + \frac{1}{m}}S} \leqslant t_{0.025,16}\right) = 1 - 0.05 = 0.95$$

$$\Updownarrow$$

$$P\left(\bar{X} - \bar{Y} - t_{0.025,16}\sqrt{\tfrac{1}{n} + \tfrac{1}{m}}S \leqslant \mu_X - \mu_Y \leqslant \bar{X} - \bar{Y} + t_{0.025,16}\sqrt{\tfrac{1}{n} + \tfrac{1}{m}}S\right) = 0.95$$

The confidence interval is therefore given by:

$$\begin{aligned} \bar{X} - \bar{Y} \pm t_{0.025,16}\sqrt{\tfrac{1}{n} + \tfrac{1}{m}}S &= 5201.3 - 5182.0 \pm 2.12\sqrt{\tfrac{1}{10} + \tfrac{1}{8}}15.1 \\ &= [4.2, 34.4] \end{aligned}$$

- <u>T-TEST</u>: Here we test H_0: $\mu_X = \mu_Y$ against H_1: $\mu_X \neq \mu_Y$ The test is based on the same T-statistic:

  $$T_{0\ \mathrm{obs}} = \frac{5201.3 - 5182.0}{\sqrt{\frac{1}{10} + \frac{1}{8}}15.1} = 2.7$$

  (We write $T_{0\ \mathrm{obs}}$ to indicate that we observe T under H_0, i.e. $\mu_X - \mu_Y = 0$.)

- <u>P</u>: p-value,

  $$p = P(T_{16} \geqslant 2.7) + P(T_{16} \leqslant -2.7) = 2P(T_{16} \geqslant 2.7) = 0.016$$

  (Two sided test and symmetric T-distribution.)

  With significance level $\alpha = 0.01$ we can *not* reject the hypothesis because $p > \alpha$, i.e. we can not assume unequal strength in the copper wires.

b) Variance analysis of one-way grouping:

Rename the variable (to get the same notation as in the book)

$$X_1, X_2, \dots, X_n \to X_{11}, X_{12}, \dots, X_{1n_1}$$
$$Y_1, Y_2, \dots, Y_n \to X_{21}, X_{22}, \dots, X_{2n_2}$$

and we have that $n_1 = 10$ and $n_2 = 8$. $N = n_1 + n_2 = 18$. (Total number of observations)

Assumptions:

$$E(X_{1j}) = \mu_1, \quad j = 1, \dots, n_1$$
$$E(X_{2j}) = \mu_2, \quad j = 1, \dots, n_2$$
$$\mathrm{Var}(X_{ij}) = \sigma^2, \quad i = 1, 2$$

(i.e. the number of groups=2). We follow the notation from the book

$$\mu_i = \mu + \alpha_i,$$

and $\mu = \frac{n_1\mu_1 + n_2\mu_2}{N}$ is "grand mean". We call α_i *the effect* of an observation coming from group i.

Model:

$$X_{ij} = \mu + \alpha_i + \epsilon_{ij}, \text{ where } \epsilon_{ij} \text{ is the random error.}$$

Variance table:

```
One-way ANOVA: X_i; Y_i

Source  DF    SS    MS     F      P
Factor   1  1656  1656  7,30  0,016
Error   16  3628   227
Total   17  5284

S = 15,06   R-Sq = 31,33%   R-Sq(adj) = 27,04%


Individual 95% CIs For Mean Based on
Pooled StDev
Level   N    Mean  StDev    -+---------+---------+---------+--------
X_i    10  5201,3   10,2                      (-------*--------)
Y_i     8  5182,0   19,6    (--------*---------)
-+---------+---------+---------+--------
5172      5184      5196      5208

Pooled StDev = 15,1
```

Source	Degrees of freedom	Sum of squares	"Mean-square"	F_{obs}	p
factor	$r-1$	$SSA = \sum_{i=1}^{r} n_i(\bar{X}_i - \bar{X})^2$	$SSA/(r-1)$	$\frac{\frac{SSA}{r-1}}{\frac{SSE}{N-r}}$	$P(F_{r-1,N-1} \geqslant F_{\text{obs}})$
error	$N-r$	$SSE = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{X}_{ij} - \bar{X}_i)^2$	$SSE/(N-r)$		
total	$N-1$	$SS_{\text{tot}} = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(X_{ij} - \bar{X})^2$			

The test that has been done is:

$$H_0 \quad : \quad \mu_1 = \mu_2 \tag{1}$$
$$H_1 \quad : \quad \mu_1 \neq \mu_2. \tag{2}$$

Under H_0 $\mu_1 = \mu_2 = \mu$ so that an equivalent test is:

$$H_0 \quad : \quad \alpha_1 = \alpha_2 \tag{3}$$
$$H_1 \quad : \quad \alpha_1 \neq 0 \text{ or } \alpha_2 \neq 0. \tag{4}$$

p:

$$p = P(F_{r-1,N-r} \geqslant F_{\text{obs}}) = 1 - P(F_{1,16} \leqslant 7.30) = 0.016$$

We have $p = 0.016 > \alpha = 0.01$, i.e. we do not reject H_0.

The p-value is the same as for the test in **a)** because

$$T_\nu^2 = F_{1,\nu}$$