TMA4255 Applied Statistics
Solution to Exercise 8

Problem 1

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:

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

- TTEST: 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:

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

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

- P: 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)

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

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

Assumptions:

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

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

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

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

Model:

$$X_{ij} = \mu + \alpha_i + \epsilon_{ij}, \text{ der } \epsilon_{ij} \text{ er tilfeldige feil.}$$

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
```

KILDE	Frihetsgrader	Kvadratsum	"Mean-square"	F_{obs}	$p-$ verdi
faktor	$r-1$	$SSA = \sum_{i=1}^{r} n_i(\bar{X}_i - \bar{X})^2$	$SSA/(r-1)$	$\frac{\frac{SSA}{r-1}}{\frac{SSE}{N-r}}$	$P(F_{r-1,N-1} \geqslant F_{\text{obs}})$
feil	$N-r$	$SSE = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{X}_{ij} - \bar{X}_i)^2$	$SSE/(N-r)$		
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:

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

p-verdi:

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

): 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

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

Problem 2

a)

```
Results for: lympho.MTW

One-way ANOVA: count versus drug

Source  DF      SS     MS     F      P
drug     4   5,703  1,426  7,38  0,000
Error   33   6,372  0,193
Total   37  12,075

S = 0,4394   R-Sq = 47,23%   R-Sq(adj) = 40,83%


                          Individual 95% CIs For Mean Based on
```

```
                                Pooled StDev
Level  N   Mean   StDev  ----+---------+---------+---------+-----
A      6  6,6833  0,4355                     (-------*------)
B      8  6,3000  0,4140               (-----*-----)
C      9  5,6000  0,4093 (-----*-----)
D      7  6,1857  0,4488            (------*-----)
E      8  5,7250  0,4892   (------*-----)
                         ----+---------+---------+---------+-----
                           5,50      6,00      6,50      7,00

Pooled StDev = 0,4394
```

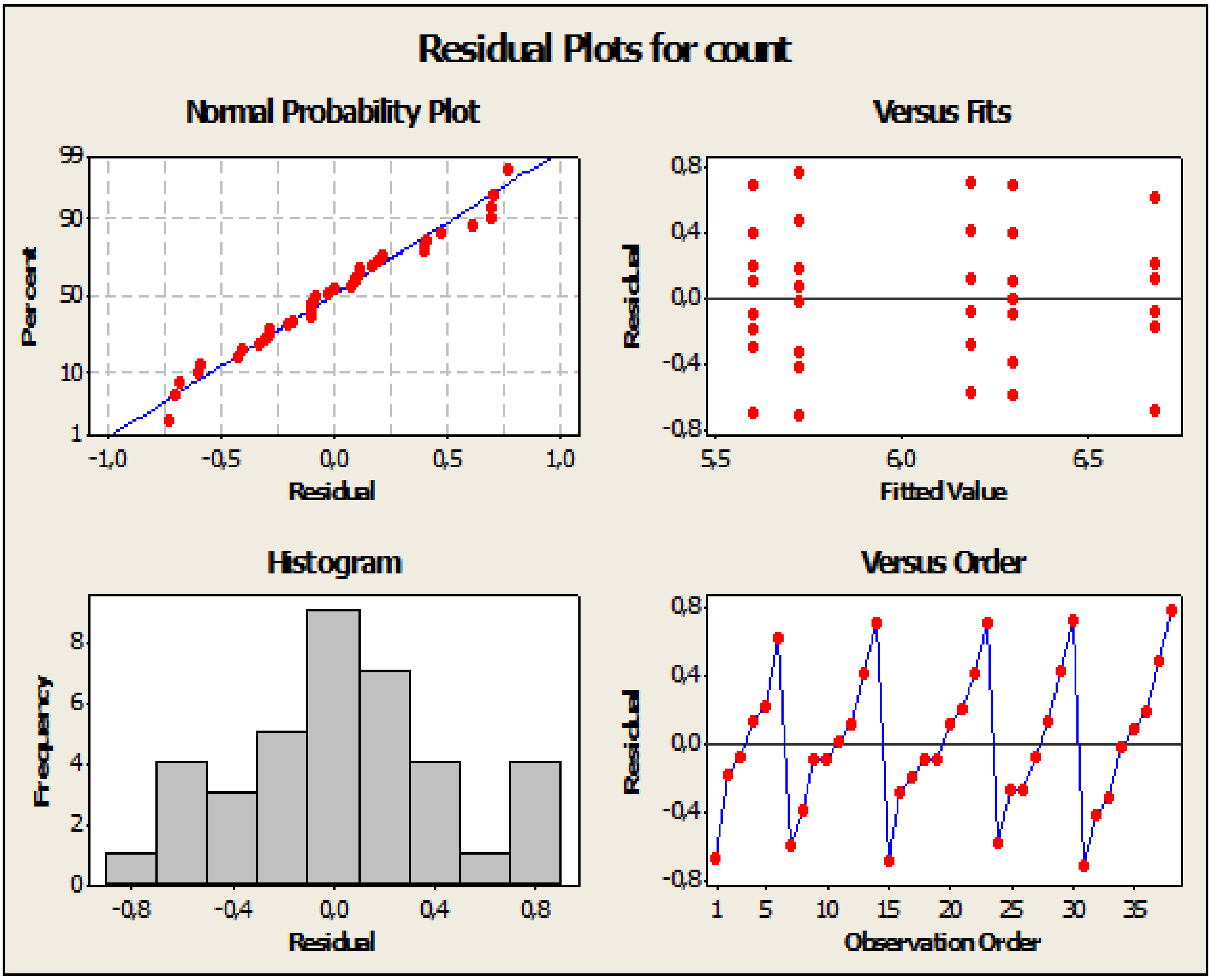
The effect of drug is significant.

```
Bartlett's Test (Normal Distribution)
Test statistic = 0,29; p-value = 0,991


Levene's Test (Any Continuous Distribution)
Test statistic = 0,11; p-value = 0,978
```

The variances for the different groups are not found to be different.

Residual plots show an adequate model fit wrt normality of error terms.



b) Using the method of Bonferroni to perform four given comparisons we will use significance level 0.05/4 = 0.125. This was given in the call to Fisher method in MINITAB (meaning that significance level 0.125 is used below).

```
Grouping Information Using Fisher Method

drug  N    Mean  Grouping
A     6  6,6833  A
B     8  6,3000  A B
D     7  6,1857  A B
E     8  5,7250    B C
C     9  5,6000      C

Means that do not share a letter are significantly different.
```

This means that when we ONLY compare A vs B, B vs C, C vs D and D vs E, we find that

- A and B does not differ,
- B and C differs
- C and D differs
- D and E does not differ.

c) We now study all pairwise comparisions with the method of Tukey.

```
Grouping Information Using Tukey Method

drug  N    Mean  Grouping
A     6  6,6833  A
B     8  6,3000  A B
D     7  6,1857  A B C
E     8  5,7250    B C
C     9  5,6000      C

Means that do not share a letter are significantly different.


Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons among Levels of drug

Individual confidence level = 99,32%
```

Using Tukeys method we conclude that A is different from both C and E, and B is different from C, but the finding from b) above (C and D differ) is not now significan when more tests are performed.

Problem 3

$$A_1, \ldots, A_4 = \text{workers (added as } 1, \ldots, 4 \text{ in C2)}$$
$$M_1, \ldots, M_4 = \text{machines (added as } 1, \ldots, 4 \text{ in C3)}$$

a) We assume that the skills of the workers do not influence the production units. This means we have one-way grouping, and we assume the model

$$Y_{ij} = \mu + \alpha_j + \epsilon_{ij}, \quad \sum_j \alpha_j = 0$$

Here:

- Y_{ij}: number of produced units by machine j and worker i.
- $E(Y_{ij}) = \mu + \alpha_j$.
- ϵ_{ij} assumed independent and $\sim N(0, \sigma^2)\ \forall i, j$.
- α_j is a factor which is special for machine j.
- μ: "average effect"

Wish to test wether the machines have different capacities:

$$\begin{aligned} H_0 &: \quad \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0 \\ H_1 &: \quad \text{at least one not equal.} \end{aligned}$$

The total variation in the data $SS_{\text{tot}} = \sum_{j=1}^4 \sum_{i=1}^4 (Y_{ij} - \bar{Y}_{\cdot\cdot})^2$, can be written as a sum of two sums of squares:[Theorem. 13.1]

$$SS_{\text{tot}} = SS_A + SS_E = \sum_{j=1}^4 4(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2 + \sum_{j=1}^4 \sum_{i=1}^4 (Y_{ij} - \bar{Y}_{\cdot j})^2$$

It can be shown that [Teorem 13.2]

$$\begin{aligned} E(SS_A) &= (4-1)\sigma^2 + \sum_{i=1}^4 4\alpha_i^2 = 3\sigma^2 + 4\sum \alpha_i^2 \\ E(SS_E) &= (16-4)\sigma^2 \\ F &= \frac{MS_A}{MS_E} = \frac{SS_A/(4-1)}{SS_E/(16-4)} \sim F_{(4-1),(16-4)} = F_{3,12} \end{aligned}$$

We see that if H_0 is correct, we can expect an $F_{0\text{ obs}}$ of about 1. If H_0 is wrong, we can expect a big value of $F_{0\text{ obs}}$.
Minitab gives us:

```
One-way Analysis of Variance
Analysis of Variance for Data

Source     DF        SS        MS        F        P
M           3      72,0      24,0     1,58    0,245
Error      12     182,0      15,2
Total      15     254,0
                                   Individual 95% CIs For Mean
                                   Based on Pooled StDev
Level       N      Mean     StDev  -+---------+---------+---------+-----
1           4    72,000     2,944  (----------*----------)
2           4    75,000     3,162          (---------*----------)
3           4    77,000     4,243               (---------*----------)
4           4    72,000     4,899  (----------*----------)
                                   -+---------+---------+---------+-----
Pooled StDev =    3,894          68,0      72,0      76,0      80,0
```

Here we have that:

$$F_{0\text{ obs}} = \frac{SS_A/3}{SS_E/12} = \frac{24.0}{15.2} = 1.58$$

the p-value:

$$p = P(F_{3,12} > F_{0\text{ obs}}) = P(F_{3,12} > 1.58) = 0.245$$

): p is larger than any reasonable significance level α, which means we can not reject H_0, and claim that there is a difference between the machines.

b) Now we assume that skills of the workers have an influence. Model:

$$X_{ij} = \mu + \alpha_j + \beta_i + \epsilon_{ij}, \quad \sum_j \alpha_j = \sum_i \beta_i = 0$$

We have:

- X_{ij}: The number of produced units with machine j and worker i.
- ϵ_{ij} assumed independent and $\sim N(0, \sigma^2)$ $\forall i, j$.
- α_j is a factor which is special for machine j.
- β_i is a factor which is special for worker i.
- μ: "average effect"

We have the same hypothesis test as in **a)**: H_0: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ against H_1: at least one is different.

We split the total variation into three sums of squares

$$SS_{\text{tot}} = SS_{\text{mask}} + SS_{\text{arb}} + SS_E$$

$$\Updownarrow$$

$$\sum_{j=1}^{4}\sum_{i=1}^{4}(X_{ij} - \bar{X}_{\cdot\cdot})^2 = 4\sum_{j=1}^{4}(\bar{X}_{\cdot j} - \bar{X}_{\cdot\cdot})^2 + \sum_{i=1}^{4}(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot})^2$$

$$+ \sum_{j=1}^{4}\sum_{i=1}^{4}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}_{\cdot\cdot})^2$$

The same type of argument as in **a)** tells us that we can expect a big value of $F_{0 \text{ obs}}$ if H_0 is wrong.

Here

$$F = \frac{SS_{\text{mask}}/(4-1)}{SS_E/((4-1)(4-1))} \sim F_{4-1,(4-1)(4-1)} = F_{3,9}$$

Minitab gives:

```
Two-way Analysis of Variance

Analysis of Variance for Data
Source        DF        SS         MS         F        P
A              3    160,00      53,33     21,82    0,000
M              3     72,00      24,00      9,82    0,003
Error          9     22,00       2,44
Total         15    254,00
```

And we see that

$$F_{0 \text{ obs}} = \frac{72.0/3}{22.0/9} = 9.82,$$

and this gives p-value: $p = P(F_{3,9} > 9.82) = 0.003$.

): We have a small p and we reject H_0.

c) Expected number of produced units from machine M_2:

$$\mu_{\cdot 2} = E(X_{\cdot 2}) = \mu + \alpha_2$$

Estimator: $\hat{\mu}_{\cdot 2} = \frac{1}{4}\sum_{i=1}^{4} X_{i2}$. This gives the point estimate: $\hat{\mu}_{\cdot 2} = \frac{1}{4}(77 + 71 + 78 + 74) = 75$.

We have:

$$E(\hat{\mu}_{\cdot 2}) = \frac{1}{4}\sum_{i=1}^{4} E(X_{i2}) = \frac{1}{4}\sum_{i=1}^{4}(\mu + \alpha_2 + \beta_i) = \mu + \alpha_2 + \frac{1}{4}\sum_{i=1}^{4}\beta_i = \mu + \alpha_2$$

and

$$\begin{aligned}
\text{Var}(\hat{\mu}_{\cdot 2}) &= E[((\hat{\mu}_{\cdot 2}) - E(\hat{\mu}_{\cdot 2}))^2] = E\left[\left(\frac{1}{4}\sum(Y_{i2} - E(Y_{i2}))\right)^2\right] \\
&= E\left[\left(\frac{1}{4}\sum_{i=1}^{4}\epsilon_{i2}\right)^2\right] = \frac{1}{16}\sum_{i=1}^{4} E(\epsilon_{i2})^2 = \left(\frac{1}{16}\right)^2 \sum_{i=1}^{4}\text{Var}(\epsilon_{i2}) \\
&= \frac{1}{4}\sigma^2
\end{aligned}$$

Therefore we get $\hat{\mu}_{\cdot 2} \sim N(\mu + \alpha_2, \frac{1}{4}\sigma^2) \Rightarrow \frac{\hat{\mu}_{\cdot 2} - (\mu + \alpha_2)}{\sigma/2} \sim N(0,1)$.

σ^2 is estimated in **b)** as $S^2 = \frac{1}{9}SS_E$.

Now we have:

$$\frac{\hat{\mu}_{\cdot 2} - (\mu + \alpha_2)}{S/2} \sim T_9$$

(same number of degrees of freedom as SS_E).

$(1-\alpha)\cdot 100$ % confidence interval:

$$P\left(-t_{\alpha/2,9} \leqslant \frac{\hat{\mu}_{\cdot 2} - (\mu+\alpha_2)}{S/2} \leqslant t_{\alpha/2,9}\right) = 1-\alpha$$

$$\Updownarrow$$

$$P(\hat{\mu}_{\cdot 2} - t_{\alpha/2,9}S/2 \leqslant \mu+\alpha_2 \leqslant \hat{\mu}_{\cdot 2} + t_{\alpha/2,9}S/2) = 1-\alpha$$

With numbers: $\hat{\mu}_{\cdot 2} = 75$, $\alpha = 0.1$, $t_{0.05,4} = 1.83$, $S^2 = 2.444$.

): 90 % confidence interval $\mu + \alpha_2$: $[73.6, 76.4]$.

Problem 4

Assume Y_{ijk} to be independent and normally distributed $\sim N(\mu_{ij}, \sigma^2)$.

- i indicates cottontype, $i = 1, 2, 3$.
- j indicates silktype, $j = 1, 2, 3, 4$.
- k indicates trial. k with combination ij, $k = 1, 2$.

$E(Y_{ijk}) = \mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$, $\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$.

a)

- μ is an "average effect".
- α_i is a factor that is special for cottontype i.
- β_j is a factor that is special for silktype j.
- γ_{ij} is a factor that is special for the interaction between cottontype i and silktype j.

Estimators for these four parameters: This is done using maximum likelihood here (optional), can also be done based on intuition.

Intuition will give us that:

- μ can be estimated using the overall mean, $\hat{\mu} = \bar{Y}_{\dots}$,
- α_i by the difference between the mean for cotton group i and the overall mean, $\hat{\alpha}_i = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\dots}$,
- β_j by the difference between the mean for silk group j, $\hat{\beta}_j = \bar{Y}_{\cdot j \cdot} - \bar{Y}_{\dots}$,
- γ_{ij} by the difference between the mean for the combined cotton and slik group and the mean over cotton, silk and the overall mean, $\hat{\gamma}_{ij} = \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\dots}$.

The following is optional, but should be possible to follow:
Probability density of one Y_{ijk}:

$$f_Y(Y_{ijk}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(Y_{ijk}-\mu_{ij})^2}$$

Find the joint density:

$$
\begin{aligned}
l(\mu, \alpha_i, \beta_j, \gamma_{ij}, \sigma^2 | Y_{ijk}) &= \prod_{i,j,k} f_Y(Y_{ijk}) \\
&= \left[\frac{1}{\sqrt{2\pi}\sigma}\right]^{3+4+2} e^{-\frac{1}{2\sigma^2}\sum_i \sum_j \sum_k (Y_{ijk}-\mu-\alpha_i-\beta_j-\gamma_{ij})^2}
\end{aligned}
$$

ln is a strictly increasing function, and $\ln l$ therefore has the same maximum points as l.

$$
L = \ln l = 9 \cdot \ln \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{2\sigma^2} \sum_i \sum_j \sum_k (Y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2
$$

Maximizing L wrt. the unknown parameters:

$$
\begin{aligned}
\frac{\partial L}{\partial \mu} &= -\frac{1}{2\sigma^2} 2 \sum_i \sum_j \sum_k (Y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})(-1) \\
&= \frac{1}{2\sigma^2} 2 \sum_i \sum_j \sum_k (Y_{ijk} - \mu) \\
\frac{\partial L}{\partial \mu} &= 0 \\
&\Rightarrow \sum_i \sum_j \sum_k (Y_{ijk} - \hat{\mu}) = 0 \\
&\Rightarrow \hat{\mu} = \bar{Y}_{\cdots}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial \alpha_i} &= \frac{1}{\sigma^2} \sum_j \sum_k (Y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij}) \\
&= \frac{1}{\sigma^2} \sum_j \sum_k (Y_{ijk} - \mu - \alpha_i) \\
\frac{\partial L}{\partial \alpha_i} &= 0 \\
&\Rightarrow \frac{1}{\sigma^2} \sum_j \sum_k (Y_{ijk} - \hat{\mu} - \hat{\alpha}_i) = 0 \\
&\Rightarrow \hat{\alpha}_i = \bar{Y}_{i\cdot\cdot} - \hat{\mu} = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots}
\end{aligned}
$$

$$
\frac{\partial L}{\partial \beta_j} = 0 \Rightarrow \hat{\beta}_j = \bar{Y}_{\cdot j \cdot} - \hat{\mu} = \bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdots} \tag{5}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial \gamma_{ij}} &= \frac{1}{\sigma^2}\sum_k (Y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij}) \\
\frac{\partial L}{\partial \gamma_{ij}} &= 0 \\
&\Rightarrow \sum_k Y_{ijk} - k\hat{\mu} - k\hat{\alpha}_i - k\hat{\beta}_j - k\hat{\gamma}_{ij} = 0 \\
&\Rightarrow \hat{\gamma}_{ij} = \frac{1}{k}\sum_k Y_{ijk} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j \\
&\qquad = \bar{Y}_{ij\cdot} - \bar{Y}_{\cdots} - (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdots}) - (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdots}) \\
&\qquad = \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots}
\end{aligned}
$$

(maksimum because $\frac{\partial^2 L}{\partial \gamma_{ij}^2} < 0$.)

b) Analysis of variance table:

Kilde	DF (DF)	Sum of squares (SS)
Rows (A)	$a-1$	$bn\sum_i(\bar{X}_{i\cdot\cdot} - \bar{X}_{\cdots})^2$
Columns (B)	$b-1$	$an\sum_j(\bar{X}_{\cdot j\cdot} - \bar{X}_{\cdots})^2$
Interaction (AB)	$(a-1)(b-1)$	$n\sum_i\sum_j(\bar{X}_{ij\cdot} - \bar{X}_{i\cdot\cdot} - \bar{X}_{\cdot j\cdot} + \bar{X}_{\cdots})^2$
Error	$ab(n-1)$	$\sum_i\sum_j\sum_k(X_{ijk} - \bar{X}_{ij\cdot})^2$
Total	$abn-1$	$\sum_i\sum_j\sum_k(X_{ijk} - \bar{X}_{\cdots})^2$

Here:

- a = number of cottontypes = 3.
- b = number of silktypes = 4.
- n = number of replicates for each combination(ij) = 2.
- X_{ijk} = k-th observation with cotton type i and silktype j.
- $\bar{X}_{ij\cdot}$ = average value of the n observations in "cell (i,j)".
- $\bar{X}_{i\cdot\cdot}$ = average value of the $b \cdot n$ observations of cotton type i.
- $\bar{X}_{\cdot j\cdot}$ = average value of the $a \cdot n$ observations of silk type j.
- $\bar{X}_{\cdots}$ = average value of all observations.

In general we have that

$$MS = SS/DF = \frac{\text{sum of squares}}{\text{DF}}$$

Hypothesis test:
H_0 : No interaction
H_1 : Interaction

Under H_0

$$F = \frac{MS_{\text{interaction}}}{MS_{\text{error}}} \sim F_{6,12}.$$

From the Minitab results in **c)** we have a p-value: $P(F_{6,12} > 2.31) = 0.103$.

): We use $\alpha = 0.05$ and can not reject H_0. We conclude that there is not evidence to believe that an interaction term is present.

c) Hypothesis test for A:

H_0 : A has no effect, i.e. $\alpha_1 = \alpha_2 = \alpha_3 = 0$

H_1 : A has effect

Under H_0

$$F_{\text{obs}} = \frac{MS_A}{MS_{\text{error}}} = \frac{SS_A/2}{SS_{\text{error}}/12} \sim F_{2,12}$$

Minitab gives:

```
Two-way Analysis of Variance
Analysis of Variance for Response

Source        DF       SS        MS        F        P
A              2    434,2     217,1    15,70    0,000
B              3    200,3      66,8     4,83    0,020
Interaction    6    191,4      31,9     2,31    0,103
Error         12    166,0      13,8
Total         23    992,0
```

p-verdi:

$$p = P(F_{2,12} > F_{\text{obs}}) = P(F_{2,12} > 15.70) = 0.000$$

): We reject H_0 and claim that A har effect.

Hypothesis test for B:

H_0 : B has no effekt, i.e. $\beta_1 = \beta_2 = \beta_3 = 0$

H_1 : B has effect

Just like above we get:

$$F_{\text{obs}} = \frac{MS_B}{MS_{\text{error}}} \sim F_{3,12}$$

p-verdi:

$$p = P(F_{3,12} > F_{\text{obs}}) = P(F_{3,12} > 4.83) = 0.02$$

): We use $\alpha = 0.05$ and we can therefore reject H_0 and say that B has effect.

Conclusion: Both cotton type and silk type influence the quality.