Solutions

1. (10 Points)

a) **(1 Pt)** (residual df + brand df + 1) / 4 = 9 jeans were selected from each brand.

b) **(2 Pts)** Null hypothesis: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ **(1 Pt)**. Answers given in words are also acceptable.
Yes. We reject the null hypothesis at level 0.05 as the p-value of the global-F test is smaller than 0.05 **(1 Pt)**.
0.5 Pt for a reasonable partial answer. For example, "We reject the null hypothesis since the p-value is small".

c) **(1.5 Pts)** The following contrasts can be used, or any equivalent one:

1. $L1$: $(1, 0, 0, -1)$
2. $L2$: $(0, 1, -1, 0)$
3. $L3$: $(1, -1, -1, 1)$
   1 Pt if at least two of them are correct.

d) **(1 Pt)** **0 Pt** if wrong justification (even if correct answer).
1 Pt if the answer is coherent with mistake in the previous task, or minor error but correct reasoning.
Yes, they are all orthogonal, as $\sum_{i=1}^{4} c_{ji} \cdot c_{ki} = 0$ for all $j \neq k$ and $j, k = 1, 2, 3$

e) **(1 Pt)** No. There can be at most 3 orthogonal contrasts as there are 4 "treatments".

f) **(2 Pts)** $MS_E = 2.43$. Hence, $F_{L1} = 5.65/2.43 = 2.33$ **(0.5 Pt)**, and $F_{L2} = 20.68/2.43 = 8.51$ **(0.5 Pt)**. Since the three contrasts are orthogonal, $SS_{L3} = SS_{\text{brand}} - SS_{L1} - SS_{L2} = 47.9 - 5.65 - 20.68 = 21.54$. Therefore, $F_{L3} = 21.54/2.43 = 8.87$ **(1 Pt)**.

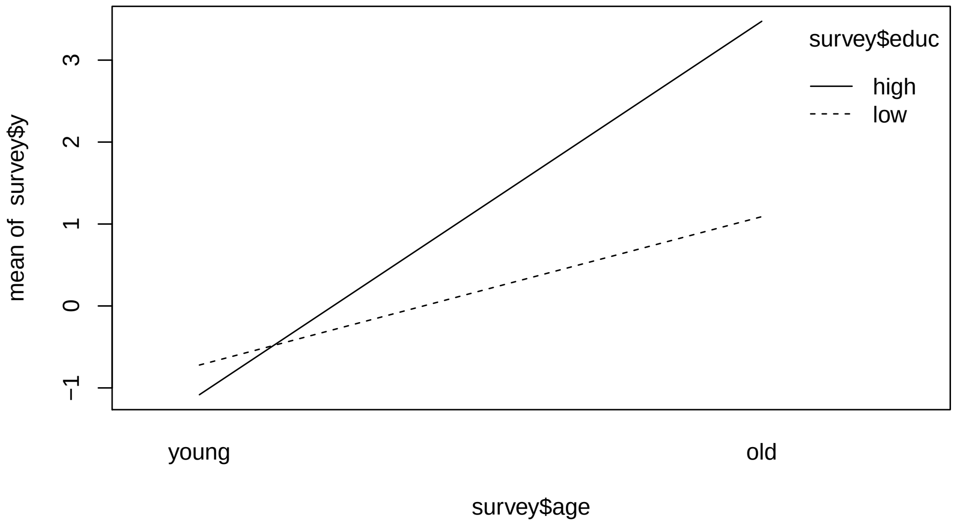
g) **(1.5 Pts)** L_1: cannot reject the null hypothesis since the value of the F-statistic is smaller than 4.15.
L_2: reject the null hypothesis, since the value of the F-statistics is larger than 4.15.
L_3: reject the null hypothesis, since the value of the F-statistics is larger than 4.15.
1 Pt if at least two conclusions are correct with appropriate reasoning.

2. (13 Points)

a) **(3 Pts)** Gian fits a two-way ANOVA model of y versus age and education with an interaction and obtains the following diagnostic plots. What assumptions can you check with each of these plots? Are these assumptions fulfilled in the present case?
1.5 Point for TA-plot and 1.5 Point for QQ-plot.
For each plot: 1 Point by correct assumption that can be tested. 0.5 Point by correct assessement of present case.

In more details: First plot (TA-plot): test that there is no systematic error(0.5) (however: here not relevant as we are fitting the full model), meaning that the model (variables chosen, transformations, etc) is correct. In addition, we check for equal variance (0.5) of the residuals. In the present case there is no visible problem. -0.5 if one mention the correct assumption but not the correct way to check it.
Second plot (QQ-plot): test that the residuals are normally distributed. In the present case the plot looks ok.

b) **(3 Pts)** Gian looks at the estimated coefficients and wants to understand his model. Draw the interaction plot corresponding to the estimated (dummy-) coefficients shown below. Focus on the qualitative aspects. **Remark: Use the (empty) plot on the next page as template but reproduce the plot on the answer sheet!**
The interaction plot should look qualitatively like that:



c) **(2 Pts)** Have a look at the following output. Complete the missing parts.
Remark: write your answer on the answer sheet!

	Df	Sum Sq	Mean Sq	F-value	Pr(>F)
age	1	240	240	20	1.31e-05
educ	1	2	2	0.16	0.689
age:educ	1	84	84	7	0.009
Residuals	196	2352	12		

The answers are found like that:

- $SS_{\text{res}} = 12 \cdot 196$
- $MS_{\text{age:educ}} = 7 \cdot 12$
- $SS_{\text{age:educ}} = 84 \cdot 1$
- $\text{df}_{\text{age:educ}} = 1$

d) **(1 Pt)** Given the ANOVA table above, what is the estimated standard deviation of the errors?
1 Point The estimated standard error is $\sqrt{12} = 3.46$.

e) **(2 Pts)** Given the ANOVA table above, what terms would you keep in the final model? Motivate your answer.
(all or nothing.)
One should keep the model as it is.
In particular, one should not remove the main effect educ as an interaction is present in the model!

f) **(2 Pts)** The previous ANOVA table was produced by the function call:

```
fit1 <- aov(y ~ age * educ, data = survey)
summary(fit1)
```

To check that he did not make any mistake, Gian asks a colleague to repeat the analysis. The latter uses the function call:

```
fit2 <- aov(y ~ educ * age, data = survey)
summary(fit2)
```

To his surprise he obtains different p-values for both education and age. What is the reason that their outputs are different and does the p-value of the interaction term differ too? Motivate your answer.
1 Point for each of element:

The reason why there is a difference is that the design must be *unbalanced* (1). In that case, the type I sum of squares used in `aov()` is dependent on the order of how the variables appear in the model formula.
The *interaction has the same p-value* (1) because it is adjusted for age and educ in both cases.

3. (11 Points)

a) **(2 Pts)** Specify the model assumptions on the random effect of `Subject` and on the error term.
Subject$_j \sim \mathcal{N}(0, \sigma^2_{Subject})$ i.i.d. and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ i.i.d. (2 points in total, for each small error deduct 0.5 point)

b) **(1 Pt)** For fixed j, are the Rating$_{ij}$ uncorrelated? Justify your answer.
No, the ratings are not uncorrelated as subjects like tea to a different extent. Formally this arises from the effect that Var(Subject$_i$) $\neq 0$.

c) **(1.5 Pts)** Specify the hypothesis that is tested in the following output and state the p-value.
This is a global test for Type, i.e. Type$_i$ = Type$_{i'}$ for all $i \neq i'$ (1 point). The p-value is approximately 10^{-8} (0.5 points).

d) **(1 Pt)** Calculate the point estimate for the expected rating of the second tea type.
$0.82 + 1.72 = 2.54$ (1 point)

e) **(1.5 Pts)** Give point estimates for the variance of the random effect of `Subject` and for the variance of `Rating`.
$\widehat{\text{Var}}(\text{Subject}_i) = 0.8151$ (0.5 point) and $\widehat{\text{Var}}(\text{Rating}_{ijk}) = 0.55 + 0.82 = 1.32$ (1 point).

f) **(1 Pt)** The vendor wants to make sure that the estimate for the expected rating of the second tea type is correct. A friend of his uses a fixed effect model and performs the analysis shown below. As above, he uses the encoding scheme `contr.treatment`. Based on the R-output below, he calculates $0.048 + 1.72 = 2.2$. What does this number refer to?
$0.048 + 1.72$ is an estimate how the reference subject (Subject1) rates the second tea type in expectation, not averaged over the subjects.

g) **(1.5 Pts)** Can we reject the hypothesis that the true variance of the random effect of `Subject` is equal to 4 on a significance level of 0.05? Give reasons for your answer.
We can reject this hypothesis if the 95% confidence interval for Var(Subject$_j$) does not cover 4. The confidence interval is $(0.52^2, 1.5^2)$ and does not cover 4, hence we can reject the hypothesis.

h) **(1.5 Pts)** The vendor wants to check the model assumptions. Which plots should he look at?
TA-plot (0.5 point), QQ-Plot of the estimated random effects (0.5 point), QQ-Plot of the residuals (0.5 point).

4. (7 Points)

1) **1 Point.** Solution d). Option a) is incorrect since the company is only interested in the effect of different treatments and it is not interested in the effects of the blocks. Option b) is incorrect - blocking does not affect whether the variance of errors is constant. Option c) is incorrect - there indeed would be a decrease in the error degrees of freedom, but this a negative consequence of blocking and thus not the reason for choosing randomized complete block design.

2) **1 Point.** Solution b). Option b) is correct because a block factor with r levels has $r - 1$ degrees of freedom. Thus after the addition of a block factor the degrees of freedom of the residuals drop by $r - 1$. Option a) is incorrect because a randomized complete block design (RCBD) tries to ensure homogeneity within blocks. Option c) is incorrect - the number of experimental units in each block of an RCBD cannot be less than the number of treatment, otherwise the block would be incomplete. Option d) is incorrect - blocking must be done before randomization, otherwise we are not guaranteed to have that each treatment appears in each block.

3) **1 Point.** Solution e) because a treatment difference is significant if the corresponding confidence interval does not cover 0.

4) **1 Point.** Solution e). All the confidence intervals would be shorter and thus some treatment differences that were not significant may become significant.

5) **1 Point.** Solution d). A randomized complete block design can only be used with one blocking factor. A Latin square must necessarily have the number of treatments and block levels (in both blocks) equal, which is not satisfied here. A Graeco-Latin square cannot be used for the same reason.

6) **1 Point.** Solution b) because it is a two-factor factorial experiment (with factors A,B), repeated in blocks (factor C). In options a) and d), the block factor is not additive, and in option c), the interaction factor between A and B is missing.

7) **1 Point.** Solution a). The incomplete block design D2 has more variance in its estimates of treatment differences than does the complete block design D1 with the same variance and number of units. Therefore they should prefer design D1.

5. (6 points)

1) **(1 Point)** Solution **d)**. Pairs of treatments appear the same number of times in each block (in this example each pair appears once), thus it is a balanced incomplete block design.

2) **(1 Point)** Solution **d)**. A Latin square of order 4 is an arrangement of 4 symbols in a 4x4 square array in such a way that each symbol occurs once in each row and once in each column. Here we take rows to be the workers and columns to be the machines.

3) **(1 Point)** Solution **c)**. This is a split plot design with temperature being the whole-plot factor and milk being the split-plot factor.

4) **(1 Point)** Solution **a)**. We can identify that other options are incorrect by looking at the magnitudes of the effects and mutual comparison of the magnitudes (note that the *absolute* effects are plotted in a half-normal plot).

5) **(1 Point)** Solution **c)**. In a two-series factorial design, the number of parameters to estimate equals the number of observations. Hence the residuals have zero degrees of freedom and thus we cannot easily construct confidence intervals and tests.

6) **(1 Point)** Solution **c)**. The correct answer is c) because there we add a random effect for the whole plots, as is required in the model equation for the split plot design.