Solutions

1. (7 points)

1) (**1 Point**) Solution **d)**. The degrees of freedom of the RSS are $N - g$, where N is the total number of observations and g is the number of treatments.

2) (**1 Point**) Solution **e)**. The highest mean estimate for `score` is obtained for the `snack` type `sugar`. The estimate is given by the intercept summed with all the other coefficient estimates, since we are using the option `contr.sum` and `sugar` is the last factor level. Its coefficient estimate is thus $82.82 + 49.28 = 132.10$, where 49.28 is the sum of the estimates of the coefficients for the other snacks.

3) (**1 Point**) Solution **d)**. There is evidence for increasing error variance with increasing fitted values, therefore we have evidence for non-constant variance of the error terms (since we clearly observe that the scatter is increasing with the fitted values).

4) (**1 Point**) Solution **c)**.

5) (**1 Point**) Solution **b)**. Since the 14 tests are independent and we assume the null hypotheses hold, the probability of not doing any false rejection is given by

$$(1 - \alpha)^{14} = 0.4877, \quad \alpha = 0.05.$$

6) (**1 Point**) Solution **c)**. The relevant 95%-quantile is the one of the $F_{1,55}$-distribution, which is 4.02. The test does not reject H_0 at level 0.05, because T is smaller than the relevant 95%-quantile.

7) (**1 Point**) Solution **d)**. The Bonferroni correction consists in dividing the significance level α by m, the number of tests performed, and to test the individual hypotheses at level $\frac{\alpha}{m}$ to achieve control over the family-wise error rate at level α. Equivalently, one can multiply the p-values by m and keep the original significance level α. Thus answer d) is the correct option.

2. (11 Points)

a) (**1.5 Points**)

```
> fit <- aov(satisfaction ~ ski * boot, data = dat)
```

0.5 Points for aov; **0.5 Points** for ski * boot and **0.5 Points** for data = dat.

b) (**1 Point**)
Both of the following answers give a total of **1 Point**. Mentioning both does not yield extra points.
```
drop1(fit, scope = ~., test = "F")
```
0.5 Points for drop1 and **0.5 Points** for scope. Alternatively,
```
library(car)
Anova(fit, type = "III")
```
0.5 Points for Anova and **0.5 Points** for type="III". The package car for Anova does not have to be named.

c) (**2 Points**)
The degrees of freedom of the Residuals is $58 - 1 - 2 - 1 - 2 = 52$ (**0.5 Points**). $MS_{boot} = 429/1 = 429$ (**0.5 Points**) and $MS_{Residuals} = 1196/52 = 23$ (**0.5 Points**) such that the F value for boot is $429/23 = 18.65$ (**0.5 Points**).

d) (**1 Point**)
Acceptable answers would be, for example: Yes. The p-value 0.049 of the **interaction term** is smaller than 0.05, the interaction term is thus significant at the 5% level (**1 Point**). Or: Yes, as the p-value 0.049 of the **interaction term** is smaller than 0.05, there is evidence that the satisfaction with the skis depends on the boot worn (**1 Point**).

e) (**2 Points**)
Gian is right. The test of the interaction term in both the Type I and the Type III approach is based on the reduction of sum of squares once the other variables, i.e. `ski` and `boot`, are already in the model (**2 Points**). Note the fact that the data are unbalanced and thus argumentation mentioning this do not yield points.

Below you can see the estimated coefficients.

```
Full coefficients are

(Intercept):    70.47
ski:            S1      S2      S3
                0.00    1.38    11.83
boot:           B1      B2
                0.00    -9.83
ski:boot:       S1:B1   S2:B1   S3:B1   S1:B2   S2:B2   S3:B2
                0.00    0.00    0.00    0.00    7.72    2.65
```

f) (**1.5 Points**)
The estimated satisfaction score of ski type S1 combined with boot type B2 is $70.47 - 9.83 = 60.64$ (**0.5 Points**). The estimated satisfaction score of ski type S2 combined with boot type B2 is $70.47 + 1.38 - 9.83 + 7.72 = 69.74$ (**1 Point**).

g) (**1 Point**)
Gian is wrong. One of the treatments, namely, S1 combined with B1 was set as the reference level (**1 Point**).

h) (**1 Point**)
There is no reason to be worried. The six lines correspond to the six groups, respective their estimated group means (**1 Point**).

3. (**8 Points**)

a) (**2 Points**) How do you assign the treatments? What is the name of the design? Is it possible to find a balanced design?
- IBD / BIBD (**0.5 Points**)
- Randomly assign each participant two of the three treatments (**0.5 Points**) (if possible, then one would do this in a balanced fashion). It is ok to draw the design instead of describing it.
- Draw the design and count. It is balanced. (**1 Point**)
- If the student did only look at the necessary condition (but not sufficient), then we only give 0.5 Points instead of 1 Point for the balancedness. The condition is $\lambda = \frac{r \cdot (k-1)}{g-1}$ is a whole number, where r is the number of replicates per treatment, k number of units per block, and g is the number of treatments. In this case, $\lambda = \frac{6 \cdot 1}{2} = 3$.
- If design allows for the comparisons at hand but is not optimal, then we only give 0.5 Points.

b) (**1 Point**) Would you tell a participant what the most promising concept according to Andreas is? Justify your answer! No, one would not tell the participant our expectation of the outcome or the most promising concept (**0.5 Points**) because we do not want to see any other influence. One tries to exclude the influence of other factors onto the response which "depend" on the treatment (**0.5 Points**).

c) (**2 Points**) Which model do you use to analyze the data collected from the design proposed in a)? Please write down the model using mathematical notation, name all the terms, and write down the R syntax that you would use to fit the model.
- Model with additive block per subject: $Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$ where α_i is the treatment factor (concept), $i = 1, 2, 3$, and β_j is the block factor (subject), $j = 1, \ldots, 9$ (**1.5 Points**).
- R syntax using `lm` or `aov`, e.g. `lm(y ~ concept + block)` (**0.5 Points**).
- It would be possible to add a second block factor into the model which would be a factor encoding the order of the seen concept per participant but we excluded this (see task).
- It is ok to add a random intercept instead of a block factor (subject) into the model.

d) **(1 Point)** General question: Why do we generally randomize treatments to experimental units? One eliminates potential confounders **(1 Point)** if we randomize the treatments to the experimental units. The only systematic difference is the treatment level.

e) **(1 Point)** General question: Why would one include a block factor when designing an experiment (and for the later analysis)? Do we have a block factor in the above design?

- They generally reduce the variance of the residuals which helps for the inference of the treatment **(0.5 Points)**, i.e. p-values get smaller / more significant.
- Yes, there is a block in the above design **(0.5 Points)**.

f) **(1 Point)** Assume the same setup as described in the exam question above with the difference that Andreas has enough money for each participant to come for three sessions. Which design would you use? Justify your answer!

- RCB Design **(0.5 Points)**
- Randomize the concepts to the three sessions per participant. This means that each participant rates each concept once **(0.5 Points)**.

4. **(7 Points)**

```
> fit <- lmer(formula = ..., data = insect.data)


Random effects:
 Groups         Name        Variance Std.Dev.
 type:location (Intercept) 2.704    1.644
 location      (Intercept) 7.858    2.803
 Residual                  8.764    2.960
Number of obs: 180


Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   27.390     1.161  20.013  23.588 4.46e-16
typeB         -2.870     1.061  45.000  -2.706  0.00959
typeC         -7.697     1.061  45.000  -7.256 4.25e-09
typeD        -14.571     1.061  45.000 -13.737  < 2e-16
typeE          4.998     1.061  45.000   4.712 2.38e-05
typeF          8.639     1.061  45.000   8.145 2.12e-10


> anova(fit)


Type III Analysis of Variance Table with Satterthwaite's method
     Sum Sq Mean Sq NumDF DenDF F value     Pr(>F)
type 5567.3  1113.5     5    45  127.06 < 2.2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


> confint(fit, oldNames = FALSE)


                                  2.5 %       97.5 %
sd_(Intercept)|type:location   0.586802    2.2029306
sd_(Intercept)|location        1.677933    4.6078221
sigma                          2.621919    3.3783450
(Intercept)                   25.123891   29.6551783
typeB                         -4.880781   -0.8593273
typeC                         -9.707557   -5.6861031
typeD                        -16.581519  -12.5600654
typeE                          2.987516    7.0089702
typeF                          6.628565   10.6500192
```

1) **(1 Point)** Solution **c)**.

2) **(1 Point)** Solution **c)**. Loris picks the location randomly in order to be able to generalize his findings.

3) **(1 Point)** Solution **d)**. The total number of observations is 180 as one can read off the output. Loris chose to study 6 different groups / types of insects and to make 3 replicates. Solve the following equation for x: $180 = 3 \cdot 6 \cdot x$.

4) **(1 Point)** Solution **d)**. The effect is highly significant but this does not mean that it is relevant! Only solution d) is correct.

5) (**1 Point**) Solution **b)**. The level F of the variable type has the highest expected value, namely 36.0.

6) (**1 Point**) Solution **c)**. Look at the confidence intervals and compare the estimates of the random effects.

7) (**1 Point**) Solution **c)**. We cannot compare specific levels of a factor which we used to define a random effect.

5. (6 points)

1) (**1 Point**) Solution **b)** In an RCB, we can test interactions between block and treatment if we have replicates.

2) (**1 Point**) Solution **c)** The Bonferroni procedure controls the family-wise error rate.

3) (**1 Point**) Solution **d)** The sum of three uncorrelated random variables with variance $2\sigma^2$ has variance $3 \cdot 2\sigma^2$.

4) (**1 Point**) Solution **c)** Motor types are assigned to cars. Hence motor type is the whole-plot factor.

5) (**1 Point**) Solution **d)** Solution d) is exactly the definition of power.

6) (**1 Point**) Solution **c)** The power of a test equals one minus the probability of making a type II error.