Written Exam
(2 hours)

General remarks:

- 10 hand- or typewritten A4 pages (= 5 sheets written on both sides). Pocket calculator. No communication devices (mobile phones, WLAN, etc.).
- Switch off your mobile phone!
- Do not stay too long on a part where you experience a lot of difficulties.
- Use only the separate sheets for your answers.
- If not stated otherwise, all tests have to be done at the 5%-level.
- Exercises 1, 4, and 5 are multiple-choice exercises. In each sub-exercise, exactly one answer is correct. A correct answer adds 1 *plus*-point and a wrong answer $\frac{1}{2}$ *minus*-points. You get a minimum of 0 points for each multiple-choice exercise. Tick the correct answer to the multiple choice exercises in the separately added answer sheet.

Good Luck!

1. **(7 Points)** Loris wants to organize a party. However, he does not know which type of peanuts he should buy as a snack. Therefore, he sets up an experiment to investigate which type of peanuts is liked most. He considers 5 types of peanuts and randomly selects 12 different people from a pool of volunteers for each type. Each participant in the study has to taste the peanut snack (s)he has been assigned to and fill in a form. Based on this information Loris calculates a continuous taste score. At the end of the experiment Loris obtains a data set `dat` with a total of 60 rows and 2 columns. The columns (variables) are:

- `score`: This is a numerical variable indicating the score value of a participant for the peanut snack he has been assigned to.
- `snack`: This is a categorical variable with 5 levels indicating the type of peanut snack (levels in the order of which they are encoded in R : `chili`, `natural`, `paprika`, `salt`, `sugar`).

In particular we have for the variable `snack`:

60 − 5

```
> xtabs(~ snack, data = dat)

snack
  chili natural paprika    salt   sugar
     12      12      12      12      12
```

Loris fits a cell means model and obtains the following result:

```
> fit <- aov(score ~ snack, data = dat)
> summary(fit)
            Df Sum Sq Mean Sq F value   Pr(>F)
snack        4  68612   17153   12.22 3.53e-07 ***
Residuals    ?  77175    1403
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

d

1) Looking at the above table, how many degrees of freedom does the mean squared error have?

a) 1
b) 4
c) 5
d) 55
e) 56
f) 59
g) 60

e

2) Now we consider the estimated coefficients.

```
> options(contrasts = c("contr.sum", "contr.poly"))
> fit <- aov(score ~ snack, data = dat)
> round(coef(fit), 2)
```

```
(Intercept)      snack1      snack2      snack3      snack4
      82.82       30.61      -14.78      -29.92      -35.19
```

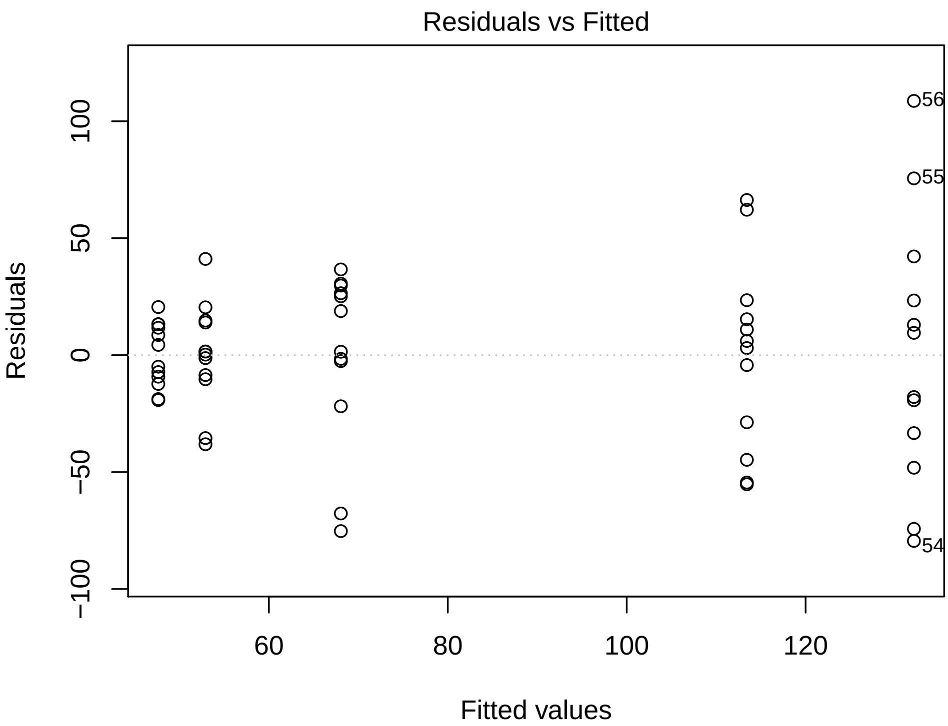
According to the fitted model, for which type of snack do we get the highest `score` estimate?

a) `chili`
b) `natural`
c) `paprika`
d) `salt`
e) `sugar`

d

3) The figure below displays the Tukey-Anscombe plot for the fitted model. Based on the figure below, which one of the following statements is true?

a) There is evidence that the assumption of zero expectation of the errors is violated.
b) There is evidence that the assumption of independent error terms is violated.
c) We do not see any violation of model assumptions.
d) There is evidence that the assumption of constant variance of the error terms is violated.
e) There is evidence that the assumption of normally distributed error terms is violated.



c

4) Loris is interested in investigating more specific questions, in particular whether there is a significant difference between the score of `salt` and `paprika`. Which of the following contrasts helps encoding the null hypothesis that the true difference is equal to zero?

   a) $(-1, 1, 0, 0, 0)$
   b) $(0, 0, 1, 1, 0)$
   c) $(0, 0, 1, -1, 0)$
   d) $(0, 1, -1, 1, -1)$

b

5) Suppose you want to perform 14 independent tests at significance level 0.05. Given that all the null hypotheses are true, what is the probability of not doing any false rejection?

   a) 30.00%
   b) 48.77%
   c) 51.23%
   d) 70.00%
   e) 77.38%
   f) 80.00%

0.95^{14}

6) For a contrast $c \in \mathbb{R}^5$ Loris calculates the test statistics

$$T = \frac{\mathrm{MS}_c}{\mathrm{MS}_E} \overset{H_0}{\sim} F_{df_1, df_2}$$

c

~~eff~~

where MS_c is the mean squares associated with the contrast and MS_E is the mean squared error. We know that the test statistics follows an F_{df_1, df_2} distribution under H_0. He obtains a value of $T = 3.51$. Which is the relevant 95%-quantile to look at in this case? Does the test reject the null hypothesis $H_0 : \sum_{i=1}^{5} c_i \mu_i = 0$ vs $H_A : \sum_{i=1}^{5} c_i \mu_i \neq 0$ at the 5% level?

   a) The relevant 95%-quantile is the one of the $\mathbf{F_{1,60}}$-distribution, which is 4.00. The test **does not reject** H_0 at the 5% level.
   b) The relevant 95%-quantile is the one of the $\mathbf{F_{1,60}}$-distribution, which is 4.00. The test **does reject** H_0 at the 5% level.
   c) The relevant 95%-quantile is the one of the $\mathbf{F_{1,55}}$-distribution, which is 4.02. The test **does not reject** H_0 at the 5% level.
   d) The relevant 95%-quantile is the one of the $\mathbf{F_{1,55}}$-distribution, which is 4.02. The test **does reject** H_0 at the 5% level.
   e) The relevant 95%-quantile is the one of the $\mathbf{F_{5,55}}$-distribution, which is 2.38. The test **does not reject** H_0 at the 5% level.
   f) The relevant 95%-quantile is the one of the $\mathbf{F_{5,55}}$-distribution, which is 2.38. The test **does reject** H_0 at the 5% level.

7) Assume we want to perform m independent tests controlling the family-wise error rate (FWER) at level α using the Bonferroni correction. Which one of the following statements is true?

a) When applying the Bonferroni correction, more tests will reject the corresponding null hypotheses compared to the situation with no correction.

b) The Bonferroni correction consists in testing the individual hypotheses at level $\alpha \cdot m$.

c) We can only apply the Bonferroni correction here because the tests are independent.

d) The Bonferroni correction can be performed by multiplying the p-values obtained from the individual tests by m and using the original α.

$\frac{\alpha}{m}$

2. **(11 Points)** A Swiss ski manufacturer plans to introduce new ski and new ski boot types to the market. The plan is to sell them in packages, i.e. skis combined with boots. For pricing the products they also want to consider customer satisfaction with the corresponding products. During the start of the season they give each combination of the 3 skis (variable `ski` with levels `S1`, `S2`, and `S3`) and boots (variable `boot` with levels `B1` and `B2`) to different (randomly assigned) people to test them for a week. Then their satisfaction scores (numerical variable `satisfaction` with values on a continuous scale of 0 to 100) are recorded. Overall, the company records 58 scores this way: 10 observations each for 5 of the ski-boot combinations, and 8 for the remaining combination (see table below). The manufacturer asks Gian (the statistics expert) to analyse the data. Below you can see an overview of the number of observations for the different combinations.

```
> xtabs(~ boot + ski, data = dat)

    ski
boot S1 S2 S3
  B1  8 10 10
  B2 10 10 10
```

a) **(1.5 Points)** Gian wants to fit a two-way ANOVA model of `satisfaction` versus `ski` and `boot` type with an interaction. What is the R code to fit this model, supposing the data set is saved in a data-frame `dat`?

b) **(1 Point)** After saving the fitted model in an object called `fit`, Gian intends to calculate the type III sum of squares. What is the R code he could use in general to calculate type III sum of squares for **all** variables in the model?

c) **(2 Points)** Gian finally obtains the following ANOVA table (with type III sum of squares). What are the values for MS_{boot} and MS_E? Moreover, we are interested in the parts marked with "...". Derive how you would calculate these values and fill in the obtained values in the ANOVA table on the **answer sheet**!

```
           Sum Sq Df    F value    Pr(>F)
ski           795  2    17.2790 1.762e-06
boot          429  1      ...   7.035e-05
ski:boot      147  2     3.1988   0.04896
Residuals    1196 ...
```

d) **(1 Point)** The company is particularly interested whether the satisfaction with the skis depends on the type of boot being worn. Is there such a significant association? Justify your answer!

e) **(2 Points)** The p-value in Gian's type III sum of square approach for the interaction term is 0.049. Another analyst (Peter) claims that Gian should have used type I sum of squares instead. Gian insists that this does not really matter, as the p-value would stay the same even in this case. Who is right, Gian or Peter? Justify your answer!

Below you can see the estimated coefficients.

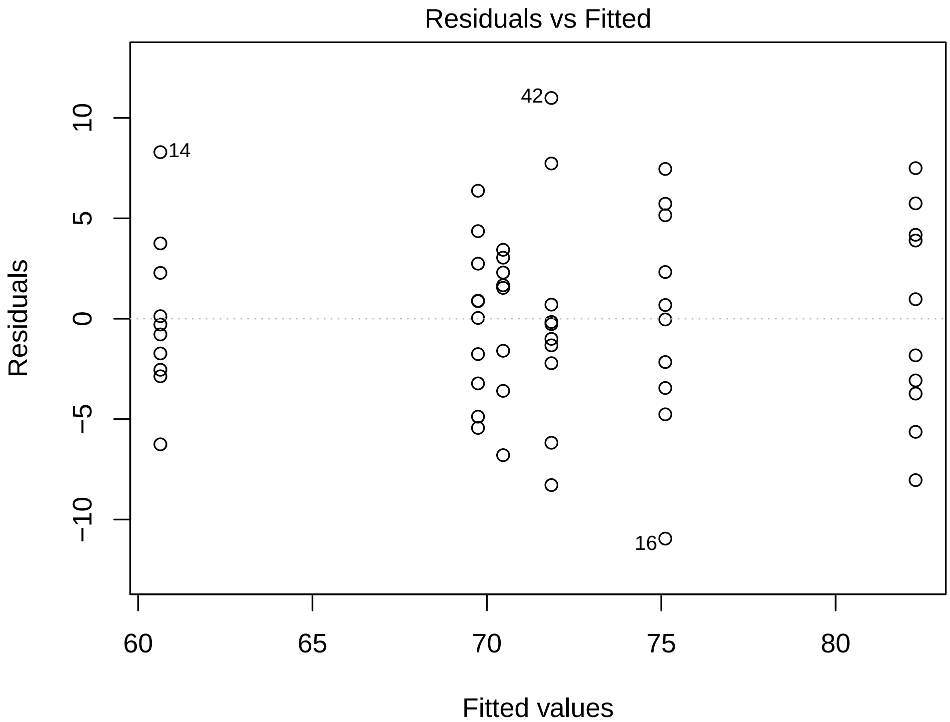
```
Full coefficients are

(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) What is the estimated satisfaction score of ski type `S1` combined with boot type `B2`? What is the estimated satisfaction score of ski type `S2` with boot type `B2`?

g) (1 Point) Gian claims that the sum-to-zero constraints were used for the estimation, while Peter is doubtful. Who is right, Gian or Peter? Justify your answer!

h) (1 Point) Below you find the Tukey-Anscombe plot corresponding to the fit.



Peter is worried about the fact that the points on the Tukey-Anscombe plot scatter on six vertical lines. Is this indeed worrisome? Justify your answer!

3. (8 Points) Andreas is an interior designer and wants to open a new co-working space where all sorts of people can work and share the offices. He bought an old factory hall and built offices inside. Now he has to decide which interior concept he will realize.

Andreas picked his three favorite interior concepts (treatment) and wants to find out with an experiment which one of the interior concepts provides the largest satisfaction (response). The response can be assumed to be continuous. Andreas found 9 participants who agree to come for two working sessions of two hours each. In each session, the participant will be assigned to an office with a corresponding interior concept. The participant works alone in that office for two hours before the response is being measured. Thus, Andreas will get one measurement per participant and session. You can assume the order of the sessions to be irrelevant. Your task is to plan the experiment.

Please note that you can solve nearly all tasks below independently of each other. Note that you can only reach the maximum points for a given task below if you picked the design which best fits the corresponding task.

a) **(2 Points)** How do you assign the treatments? What is the name of the design? Is it possible to find a balanced design?

b) **(1 Point)** Would you tell a participant what the most promising concept according to Andreas is? Justify your answer!

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.

d) **(1 Point)** General question: Why do we generally randomize treatments to experimental units?

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?

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!

4. (7 Points) Biologist Loris wants to investigate the diversity of six classes of insects that can be found in the forests of the canton of Zurich. These classes of insects are encoded in the factor `type`. The diversity (response) can be measured as a continuous quantity which is related to the quantity of insects captured in a trap at some randomly selected locations. Loris is specifically interested in differences between these six classes of insects and wants to generalize his findings to the whole canton of Zurich. He empties the traps at three different time-points (~~at all locations~~). For the sake of simplicity, we assume that time has no influence on the response, i.e. we treat the different time-points as independent replicates. Loris decides to use the following model

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk},$$

where α_i is the effect of the class of insects, β_j is the effect of the location, $(\alpha\beta)_{ij}$ is the effect defined by the combination of the class of insects and the location, and ϵ_{ijk} is the error term. Below you see parts of the R output of Loris' analysis. He used the encoding scheme `contr.treatment`, i.e. `typeA` as reference level.

```
> 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
```

c

1) Which of the following statements do you have to use for the formula argument of `lmer` to obtain the R output from above?

   a) `diversity` $\sim$ `type + location + type:location`
   b) `diversity` $\sim$ `type + location + (1 | type:location)`
   c) `diversity` $\sim$ `type + (1 | location) + (1 | type:location)`
   d) `diversity` $\sim$ `type + (type | location)`
   e) `diversity` $\sim$ `type * location + (1 | type:location)`

c

2) Which of the following statements is correct with respect to the choice of the model implied by the description of the exam question (not necessarily the R output)?

   a) The variable `location` should be a fixed effect because Loris wants to compare these specific locations.
   b) The variable `type` has six levels which means that we should include it as a random effect.
   c) Loris wants to generalize his findings to the whole canton of Zurich. This is why he includes the variable `location` as a random effect.
   d) There are too few levels of the variable `location`. We should not include it as a random effect.
   e) None of the above answers is correct.

d

3) How many locations did Loris pick?

   a) 6
   b) 8
   c) 9
   d) 10
   e) 12

d

4) Is there a systematic effect of the variable `type`?

   a) No, we cannot reject the null hypothesis.
   b) Yes, the p-value is close to 0.05.
   c) Yes, it is highly significant and therefore we know that the effect is of practical relevance.
   d) Yes, it is highly significant. For practical relevance, we should check the confidence interval and have the corresponding background knowledge.

b

5) Loris is interested in the class of insects (variable `type`) with the highest expected value. What is its expected value (rounded to the first decimal place)?

   a) 42.0
   b) 36.0
   c) 32.4
   d) 27.4
   e) 8.6

C

6) Are the variances of both random effects significantly larger than zero?

   a) No, they are not.
   b) Just one of them is significant.
   c) Yes, both are significant and the estimate of the variance of the random effect defined only by the variable `location` is larger.
   d) Yes, both are significant and the estimate of the variance of the random effect defined only by the variable `location` is smaller.

C

7) Loris thinks about changing the experimental design. Which answer is **wrong**?

   a) If Loris is interested in specific locations and wants to compare them, he would include all the variables as fixed effects (incl. the interaction term).
   b) Loris thinks that there is a lot of variation because of the height of the forest (factor with three levels). He wants to account for height although he has no specific interest in height. In that case, one would typically include height as an additive block factor.
   c) Loris can compare specific locations and do corresponding tests with the model that was fitted in R.
   d) Loris can use the function `glht` (of the R package `multcomp`) to test contrasts of fixed effects.

5. (6 points)

1) Which of the following statements is **wrong**?
   a) Blocking is used to reduce variance by choosing blocks such that each has homogeneous experimental units.
   b) In an RCB (randomized complete block design), we can never test interactions between block and treatment.
   c) In an RCB with a factorial treatment structure, we can test interactions between treatment factors even if we only observe every treatment combination once in each block.
   d) Blocking can increase precision, even if the p-value corresponding to the block factor is not significant.

2) Which of the following statements about multiple testing is true?
   a) The more different tests we perform (each at a fixed level α), the less likely we are committing a type I error.
   b) Tukey Honest Significant Difference is less powerful than the Bonferroni correction.
   c) The family-wise error rate (the probability of committing a type I error) can always be controlled by using the Bonferroni procedure.

3) Consider the following experimental design with a treatment factor having levels A, B, C, D, E, and F.

Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
A	A	B	D	D	B
B	C	C	E	F	F

Which of the following statements is true?
   a) It is not possible to estimate all treatment differences.
   b) The design is balanced.
   c) This is an unreduced BIBD.
   d) If estimating each of the treatment differences $A-B$, $F-B$, and $D-F$ has variance $2\sigma^2$, then estimating the treatment difference $A-D$ via the decomposition $A-D=(A-B)-(F-B)-(D-F)$ has variance $6\sigma^2$.

4) A car company wants to investigate the effect of 4 different tire types and two motor types on tire wear. The car company conducts this test on 8 cars. It samples 4 motors of each motor type and 8 tires of each tire type. The motor types are randomly assigned to each car. Furthermore, for each car one tire of each tire type is randomly assigned to one of the four wheels. The response is tire wear for each tire. This is an example of a split-plot design with ...
   a) cars as the whole-plot factor.
   b) tires as whole-plots.
   c) motor types as the whole-plot factor.
   d) wheels as the whole-plot factor.

5) Which of the following statements about power is true?

a) Power is the probability that H_A holds, given that we reject H_0.

b) The power of a test and its significance level α always sum up to one.

c) Generally speaking, if the significance level α decreases, then the power increases.

d) Power is the probability that H_0 is rejected, given that H_A holds.

6) We conduct a test, where the probability of making a type I error is 0.2 and the probability of making a type II error is 0.7. What is the power of this test?

a) 0.1

b) 0.8

c) 0.3

d) 0.7

e) 0.2