

Data Analysis — Final Exam

Summer Term 2023

Maik Bieleke & Johanna Stähler

Table of contents

Instructions	1
Problem 1: Data preprocessing (10 points)	3
Problem 2: Analysis workflow (10 points)	4
Problem 3: Demographic variables (15 points)	4
Problem 4: Cronbach's α (15 points)	5
Problem 5: Visualizing the User Experience Questionnaire (UEQ) (25 points)	5
Problem 6: Analysis of the Flow State Scale-2 Short (FSS-2S) (15 points)	7

Instructions

Article

This exam is based on Weigl, K., Schuster, S., & Riener, A. (2021). Investigation of UX and flow experience in sports activities during the covid-19 pandemic: A comparative analysis of cycling apps. In *Proceedings of the 9th International Conference on Sport Sciences Research and Technology Support*, p. 61-68. <http://dx.doi.org/10.5220/0010688200003059>

Both the article ([Weigl2021.pdf](#)) and the data ([Weigl2021.csv](#)) are available on Ilias. You should read the paper carefully before working on the exam. Make sure that you understand the methods and analyses, as well as the dataset.

Scoring

The maximum possible score for the exam is 100 points. Up to 90 points can be earned by solving the six problems, each of which comprises several tasks. Another 10 points can be earned by adhering to the coding and scripting style introduced in the seminar.

Problem	Points
Problem 1	10
Problem 2	10
Problem 3	15
Problem 4	15
Problem 5	25
Problem 6	15
Coding & Scripting Style	10

Using packages, functions, and features

If not explicitly stated otherwise in the problem, only R packages, R functions, and RStudio features introduced in the seminar are allowed for solving each task. You will receive 0 points for a task that uses other packages, functions, or features.

Referencing seminar content

Whenever you use or mention R packages, R functions, or RStudio features from the seminar, provide a brief reference to the relevant lecture slide (e.g., “L4, Slide 5” is short for Slide 5 in Lecture 4 on Operations). You will receive 0 points for a task that has no proper references.

Example: Calculate the mean and standard deviation of vector A.

Answer:

```
# L4, Slide 5
mean(vectorA)
sd(vectorA)
```

Submitting Files

You submit three files:

- The `.r` script for analyzing the dataset.
- A text document (e.g., `.docx`) with detailed explanations and descriptions and the screenshot in Problem 1.
- The figure in Problem 5 as `.png` file.

Problem 1: Data preprocessing (10 points)

Note that this problem deals with the original dataset of the article, not with the dataset you received for the exam. Accordingly, all answers can be provided in the text document even if they involve R code.

Figure 1 shows a screenshot that has been generated based on the original dataset.

```
$ ID : int 1 2 3 4 5 6 7 8 9 10 ...
$ Adidas : logi NA NA NA NA NA NA ...
$ A_FSS2S_MeanScore : chr "4,5555555555555556" "3,666666666666667" "3,888888888888889" "3,888888888888889" ...
$ A_UEQ_Factor1_Attractiveness_MeanScore : chr "3" "1" "1,5" "3" ...
$ A_UEQ_Factor2_Perspicuity_MeanScore : chr "3" "1,25" "2,25" "3" ...
$ A_UEQ_Factor3_Efficiency_MeanScore : chr "2,25" "1,5" "2,25" "3" ...
$ A_UEQ_Factor4_Dependability_MeanScore : chr "2" "1" "1,5" "3" ...
$ A_UEQ_Factor5_Stimulation_MeanScore : chr "3" "1" "1,5" "2,5" ...
$ A_UEQ_Factor6_Novelty_MeanScore : chr "3" "2,5" "1,25" "7,5" ...
$ Komoot : logi NA NA NA NA NA NA ...
$ K_FSS2S_MeanScore : chr "4,111111111111111" "3,888888888888889" "3,666666666666667" "4,111111111111111" ...
$ K_UEQ_Factor1_Attractiveness_MeanScore : chr "3" "2,333333333333333" "1,666666666666667" "2,333333333333333" ...
$ K_UEQ_Factor2_Perspicuity_MeanScore : chr "3" "2,5" "1,5" "1,5" ...
$ K_UEQ_Factor3_Efficiency_MeanScore : chr "2,25" "2,25" "1,5" "1,75" ...
$ K_UEQ_Factor4_Dependability_MeanScore : chr "3" "2,25" "1,5" "1" ...
$ K_UEQ_Factor5_Stimulation_MeanScore : chr "3" "2" "2,25" "2,25" ...
$ K_UEQ_Factor6_Novelty_MeanScore : chr "2,25" "1,5" "2" "7,5" ...
$ Strava : logi NA NA NA NA NA NA ...
$ S_FSS2S_MeanScore : chr "4,222222222222222" "4" "3,555555555555556" "4,555555555555556" ...
$ S_UEQ_Factor1_Attractiveness_MeanScore : chr "3" "2,166666666666667" "1,5" "8,333333333333333" ...
$ S_UEQ_Factor2_Perspicuity_MeanScore : chr "3" "2,5" "1,5" "3" ...
$ S_UEQ_Factor3_Efficiency_MeanScore : chr "3" "2,5" "1,5" "1,25" ...
$ S_UEQ_Factor4_Dependability_MeanScore : chr "1,25" "2,5" "7,5" "2" ...
$ S_UEQ_Factor5_Stimulation_MeanScore : chr "1" "2" "1,25" "5" ...
$ S_UEQ_Factor6_Novelty_MeanScore : chr "2,5" "7,5" "1" "-2,5" ...
$ Geschlecht : int 1 2 1 1 2 1 2 1 2 ...
$ Alter : int 39 53 24 22 57 19 48 55 25 54 ...
$ Staatsangehörigkeit_Deutsch : int 1 1 1 1 1 1 1 1 1 ...
$ Familienstand : int 3 3 2 1 3 1 3 3 3 ...
$ Ausbildungsabschluss : int 8 5 8 3 9 3 4 4 8 ...
$ Wohnort : int 1 1 2 3 4 3 1 1 3 ...
$ Führerschein_Besitz : int 1 1 1 1 1 1 1 1 1 ...
$ Führerschein_Besitz_Jahre : int 21 35 7 4 40 1 26 32 8 31 ...
$ Pendeln_Ä-PNV : int 2 2 1 1 2 2 2 2 2 ...
$ Fahrrad_Verkehrsmittel_Freizeit : int 1 1 1 1 1 1 1 1 1 ...
$ Ä-PNV_Verkehrsmittel_Freizeit : int 2 2 2 1 2 2 2 2 2 ...
$ Regelmäßiges_Laufen : int 0 3 0 0 0 3 5 1 0 ...
$ Regelmäßiges_Walken : int 0 0 2 0 2 0 0 0 0 ...
$ Teilnahme_Wettkämpfen : int 0 0 0 0 0 0 0 0 0 ...
$ Fahrradfahren_proWoche : int 7 5 3 4 2 4 6 5 1 ...
$ Fahrradfahren_Hobbysport_1Jahr : int 2 2 1 2 2 2 2 2 1 ...
$ AppRanking_Rank1 : int 1 1 2 3 2 1 2 2 1 ...
$ AppRanking_Rank2 : int 2 2 3 2 3 2 1 1 3 ...
$ AppRanking_Rank3 : int 3 3 1 1 1 3 3 3 2 ...
$ Bekannte_Apps_Spezifikation : chr "Strava und Komoot" "alle drei" "" "" ...
```

Figure 1: Screenshot based on the original dataset

1. Explain in writing what output the screenshot displays, and what information it provides about the data.
2. Identify 5 potential issues with the data shown in the screenshot. Describe in writing how each issue can be addressed, and specify the R functions involved.
3. One value in the screenshot is highlighted in yellow. Suggest 6 ways of extracting it with operations introduced in the seminar. Explain the R code for each approach, and explain which approach(es) you would favor and why.

Problem 2: Analysis workflow (10 points)

Note that no R coding is required for solving the following tasks. Accordingly, all answers (including the screenshot) can be provided in the text document.

1. Establish a data analysis workflow along the guidelines provided in the seminar, using RStudio features only. Your workflow should involve creating a directory with all the files and folders needed for the analysis of the exam data.
2. Describe in writing how you established the directory in RStudio. Produce a screenshot (or take a picture) of the appropriate RStudio pane or tab that shows your directory.
3. Explain in writing how your workflow handles the working directory and the workspace for your analysis.

Note

For solving the following problems, it will be necessary to take steps that are not directly required in the tasks, such as importing the data, loading packages, or preprocessing variables. Carry out these additional steps as necessary for solving the problems.

Problem 3: Demographic variables (15 points)

In this problem, you will analyze some demographic variables. Make sure that your output closely matches the output shown below, which might require preprocessing the variables. Note that character functions (L4, Slide 9ff) are neither necessary nor allowed.

1. Gender: Absolute frequency of women and men

```
women  men
      20   14
```

2. Gender: Relative frequency of women and men.

```
women  men
58.82 41.18
```

3. Age: Means, standard deviations, and ranges for women and men.

```
# A tibble: 2 x 5
  sex      Min    Max      M    SD
  <fct> <int> <int> <dbl> <dbl>
1 women     19     57  32.5  13.3
2 men       19     56  39.5  13.3
```

4. The authors reported how many of their participants “regularly go running (12/34)” (p. 62). Find the relevant variable in the dataset and reconstruct how the authors arrive at these numbers. Then modify and analyze the variable to produce the following output.

```
no regular running    regular running
                22                12
```

5. The authors claim that 18 participants knew at least one of the three apps. Find the relevant variable in the dataset and use R code to demonstrate that this statement is incorrect.

Problem 4: Cronbach’s α (15 points)

Several Cronbach’s α values are reported in the article. Cronbach’s α is a measure of the internal consistency of a scale. One way to compute it in R is the `misty::item.alpha()` function. Solve the following tasks based on the function’s documentation.

1. Show that the Cronbach’s α values reported in the article cannot be calculated with the available dataset according to the documentation.
2. What additional data would allow to calculate the values according to the documentation? Describe two possibilities.
3. Write code that could be used to calculate the standardized Cronbach’s α if this additional data were available.
4. Change your code to compute the standardized McDonald’s ω using the `sampe` package. McDonald’s ω is an alternative measure of the internal consistency of a scale that is not reported in the paper.
5. Compute Cronbach’s α and McDonald’s ω using the example data in the documentation of `misty::item.alpha()`. Extract the two coefficients to produce the following console output, without typing the numeric values of the coefficients manually. Remember to provide references to the seminar contents as necessary.

```
Coefficient alpha is 0.88 and coefficient omega is 0.88 as well.
```

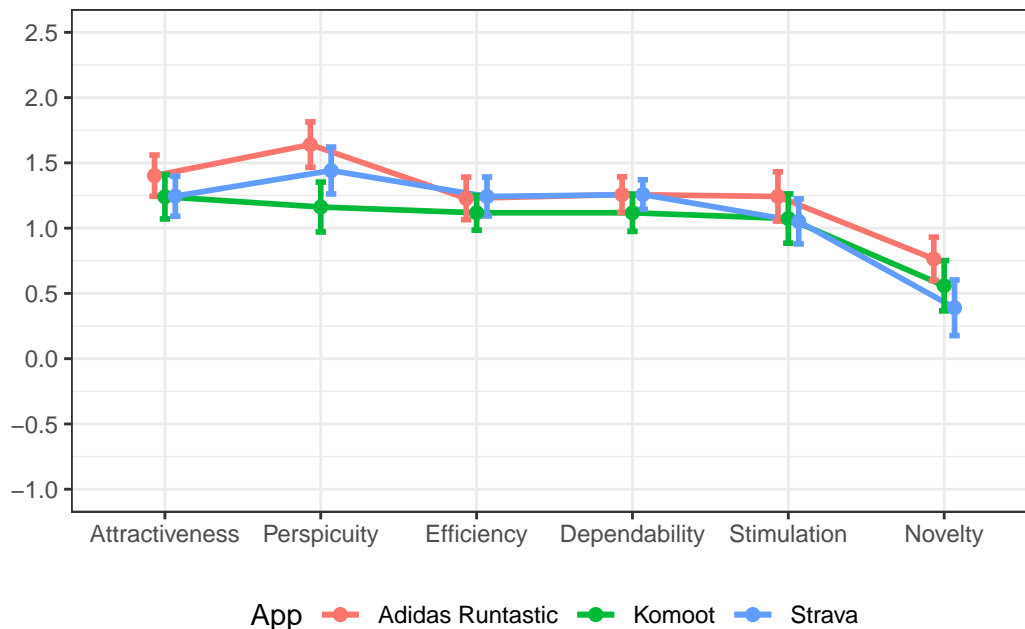
Problem 5: Visualizing the User Experience Questionnaire (UEQ) (25 points)

The authors visualized the user experience with different apps in Figure 1 of the article. The figure comprises (a) the means for the different plots and (b) the benchmarks from Schrepp et al. (2014). In this problem, your task is to produce a simplified version of their graph that shows only (a).

1. Explain in writing why the UEQ data in the dataset is not tidy in the sense discussed in the seminar.
2. Write code that modifies the dataset in order to make it suitable for creating the plot in Task 4. Explain in writing how your code makes the dataset tidy.
3. Based on the modified dataset from Task 2, compute the summary statistics required for producing the graph: means and standard errors of the mean. Note that the standard error of the mean is defined as

$$SD_x/\sqrt{N}$$

4. Create a plot that resembles the following one as closely as possible (pay attention to details, such as correct labelling of the x-axis).



5. Save the plot in `.png` format in the designated location of your directory. Make sure that the proportions of the figure are similar to the plot depicted in Task 4.
6. Looking at Figure 1 in the article, what kind of geom(s) could be used to add the benchmark data to the graph produced in Task 4? Identify three potential obstacles for adding the benchmark data to the plot created above.

Problem 6: Analysis of the Flow State Scale-2 Short (FSS-2S) (15 points)

1. Explain in writing why the data for the FSS-2S cannot be submitted to an ANOVA with the dataset in its current format (assuming packages and functions introduced in the seminar).
2. Modify the dataset in order to run the analysis.
3. Based on the modified dataset, compute the ANOVA. The output should look like this:

Anova Table (Type 3 tests)

Response: Score

	Effect	df	MSE	F	ges	p.value
1	App 1.89, 62.31	0.11	0.04	<.001		.952

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '+' 0.1 ' ' 1

Sphericity correction method: GG

4. Using the documentation, explain in writing how and why the output above deviates from the results reported in the paper.
5. Change your ANOVA call such that you exactly reproduce the results in the paper.