

# Übungszettel Anova

M.Psy.205 Multivariate Statistik. Dozent: Peter Zezula

Joshua Driesen ([joshua.driesen@stud.uni-goettingen.de](mailto:joshua.driesen@stud.uni-goettingen.de))

## Deutsch

### Links

[Übungszettel als PDF-Datei zum Drucken](#)

### Hinweise zur Bearbeitung

1. Bitte Bitte beantworten Sie die Fragen in einer .Rmd Datei. Sie können Sie über `Datei > Neue Datei > R Markdown...` eine neue R Markdown Datei erstellen. Den Text unter dem *Setup Chunk* (ab Zeile 11) können Sie löschen. [Unter diesem Link](#) können Sie auch unsere Vorlage-Datei herunterladen (Rechtsklick > Speichern unter...).
2. Informationen, die Sie für die Bearbeitung benötigen, finden Sie auf der [Website der Veranstaltung](#)
3. Zögern Sie nicht, im Internet nach Lösungen zu suchen. Das effektive Suchen nach Lösungen für R-Probleme im Internet ist tatsächlich eine sehr nützliche Fähigkeit, auch Profis arbeiten auf diese Weise. Die beste Anlaufstelle dafür ist der [R-Bereich der Programmiererplattform Stackoverflow](#)
4. Auf der Website von R Studio finden Sie sehr [hilfreiche Übersichtszettel](#) zu vielen verschiedenen R-bezogenen Themen. Ein guter Anfang ist der [Base R Cheat Sheet](#)

### Ressourcen

Da es sich um eine praktische Übung handelt, können wir Ihnen nicht alle neuen Befehle einzeln vorstellen. Stattdessen finden Sie hier Verweise auf sinnvolle Ressourcen, in denen Sie für die Bearbeitung unserer Aufgaben nachschlagen können.

Ressource	Beschreibung
Field, Kap. 10	Eine ausführliche, sehr gute Einführung in die einfaktorielle ANOVA (und Kontraste)
Field, Kap. 12	Für mehrfaktorielle Designs und Varianzanalysen

### Tipp der Woche

#### Befehle zu Paketen

Die in unseren Skripten benutzten Befehle `require()` und `library()` akzeptieren den Namen der Pakete, auf die sie sich beziehen, einfach so, also z.B. `require(dplyr)`. Wenn Sie auf Ihrem eigenen Rechner Pakete zum ersten Mal installieren wollen, erwartet Sie jedoch ein Stolperstein: Der `install.packages()`-Befehl verlangt den Paketnamen in Anführungszeichen (also als sogenannter String), z.B. `install.packages("dplyr")`.

## 1) Einfaktorielle ANOVA

1. Setzen Sie, wie Sie es bereits gewohnt sind, den von Ihnen gewählten Ordner als Arbeitsverzeichnis, und laden Sie die tidyverse-Pakete.
2. Lesen Sie den Datensatz `Superhero.dat` mit dem Befehl `read_delim()` direkt aus der URL <https://pzezula.pages.gwdg.de/data/Superhero.dat> in R ein. Dieser enthält Daten von Kindern, die in einem Superheldenkostüm in einer Notaufnahme vorstellig wurden. Dabei ist sowohl die Schwere der Verletzung vermerkt (“injury”), als auch, welchen Superheld die Kostüme nachbildeten (“hero”). Die Kodierung der Helden erfolgte nach folgendem Schema:
  - 1 - Spiderman
  - 2 - Superman
  - 3 - Hulk
  - 4 - Teenage Mutant Ninja TurtlesWir wollen nun mittels ANOVA herausfinden, ob verschiedene Superheldenkostüme mit unterschiedlich schweren Verletzungen assoziiert sind.
3. Ein FALSCHER Ansatz: Nutzen Sie den `lm()`-Befehl, um den Verletzungsgrad aus der Heldenkategorie vorherzusagen, und betrachten Sie das Ergebnis. Warum ist diese Art der Analyse NICHT valide?
4. Nutzen Sie den `factor()`-Befehl, um die Heldenvariable zu faktorisieren. Nutzen Sie auch `labels()`, um mit der Tabelle oben die Helden ihren jeweiligen Codes zuzuordnen.
5. Erzeugen Sie einen Bar-Plot mit Fehlerbalken.
6. Die unterschiedlich langen Fehlerbalken könnten auf eine Verletzung der Varianzhomogenität hinweisen. Prüfen Sie diese Vermutung mit dem Levene-Test. (Nutzen Sie hierzu mit “`??levene`” R’s Suchfunktion, um den entsprechenden Befehl zu finden.)
7. Führen Sie mit dem `aov()`-Befehl eine ANOVA durch, und speichern Sie das Ergebnis als `ANOVA1`. Unterscheiden sich die verschiedenen Superhelden hinsichtlich der Verletzungsschwere?
8. In der Vorlesung wurde die Äquivalenz von ANOVA und Regression besprochen. Führen Sie dementsprechend die gleiche ANOVA noch einmal durch, diesmal jedoch mit dem `lm()`-Befehl. Speichern Sie das Ergebnis als `ANOVA2`. Vergleichen Sie die beiden Ergebnisobjekte miteinander. Welche Werte entsprechen sich? Welche Analyseform beinhaltet mehr Information? Bonus: Was genau bedeutet `heroHulk = -6.250`?
9. Führen Sie eine weitere Analyse durch, in der Sie untersuchen, ob sich die beiden bekannten Comic-Universen Marvel und DC hinsichtlich der assoziierten Verletzungsschwere unterscheiden. Definieren Sie hierzu mit `contrasts(Superhero$hero) <- IhrKontrast` einen geplanten Kontrast. Spiderman und Hulk gehören zu Marvel, Superman entstammt dem DC-Universum, und die Ninja Turtles gehen nicht in diese Analyse ein.

## 2) Zweifaktorielle ANOVA

1. Lesen Sie den Datensatz `ChickFlick.dat` mit dem Befehl `read_delim()` direkt aus der URL <https://pzezula.pages.gwdg.de/data/ChickFlick.dat> in R ein. In diesem finden Sie die physiologischen Erregungsmessungen von Männern und Frauen, die im Labor entweder den klassischen “Chick-Flick” Bridget Jones’ Diary, oder den Thriller Memento zu sehen bekamen.
2. Erzeugen Sie einen Bar-Plot mit Fehlerbalken.
3. Prüfen Sie die Validität des Konzeptes “Chick-Flick”, indem Sie mittels ANOVA beantworten, ob die

beiden Geschlechter unterschiedlich auf verschiedene Filme reagieren.

4. Wieviel Vertrauen haben Sie in die Generalisierbarkeit Ihrer Schlussfolgerung aus der vorherigen Aufgabe? Welche Alternativerklärungen bleiben offen? Wie liesse sich die Aussagekraft erhöhen?

### 3) Rendern

Rendern, bzw. knitten Sie nun das Dokument über die Tastenkombination `strg + shift + k` (Windows) oder `cmd + shift + k`. Wenn das funktioniert: Top gemacht! Wenn nicht: Schauen Sie sich die Fehlermeldung an, und betrachten Sie insbesondere die Zeilen Ihrer Syntax, die in der Fehlermeldung auftauchen. Suchen Sie nach dem Fehler und probieren Sie es erneut!

## Literatur

*Anmerkung:* Diese Übungszettel basieren zum Teil auf Aufgaben aus dem Lehrbuch *Discovering Statistics Using R* (Field, Miles & Field, 2012). Sie wurden für den Zweck dieser Übung modifiziert, und der verwendete R-Code wurde aktualisiert.

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. London: SAGE Publications Ltd.

## English

### Links

[Exercise sheet in PDF](#)

### Some hints

1. Please give your answers in a .Rmd file. You may generate one from scratch using the file menu: ‘File > new file > R Markdown ...’ Delete the text below *Setup Chunk* (starting from line 11). Alternatively you may use this [sample Rmd](#) by donloading it.
2. You may find the informations useful that you can find on the [start page of this course](#).
3. Don’t hesitate to google for solutions. Effective web searches to find solutions for R-problems is a very useful ability, professionals to that too ... A really good starting point might be the R area of the programmers platform [Stackoverflow](#)
4. You can find very useful [cheat sheets](#) for various R-related topics. A good starting point is the [Base R Cheat Sheet](#).

## Ressources

This is a hands on course. We cannot present you all the useful commands in detail. Instead we give you links to useful ressources, where you might find hints to help you with the exercises.

---

Ressource	Description
Field, Chapter 10	A detailed introduction to single factor ANOVA and contrasts
Field, Chapter 12	This chapter deals with multi factorial designs and variance analyses

---

## Tip of the week

### Dealing with package names

We use two commands to load packages in our scripts, `require()` and `library()`, both of which understand quoted names of packages, f. e. `require("psych")`. To make life a bit easier, we can also use unquoted package names, f. e. `require(psych)` or `library(dplyr)`. But take care: We still need to use quoted package names when we want to install using `install.packages()` f. e. `install.packages("psych")`.

## 1) Single factor ANOVA (oneway)

1. Set your working directory and load tidyverse as usual.
2. Read the data file 'Superhero.dat' using either `read_delim()` or `read_tsv()` from the URL <https://pzezula.pages.gwdg.de/data/Superhero.dat> and store it in a data object in R. We find data of injured children that were dressed up like a superhero when they were presented in an emergency room of a hospital. 'injury' gives us severity of the injury and 'hero' specifies, which superhero was being copied: 1 - Spiderman  
2 - Superman  
3 - Hulk  
4 - Teenage Mutant Ninja Turtles  
Using a ANOVA we want to find out, whether there is an association between type of superhero and severity of injuries.
3. A wrong approach: Use the command `lm()` and take a look at the results. Why is this analysis **not** valid?
4. Use the command `factor()` to factorize variable `superhero`. Also define `labels()` to connect type of hero and the referred code.
5. Generate a barplot with errorbars.
6. The varying errorbars might indicate inhomogeneous variances. Test this using the levene test. Try `levene` to let R help you to learn about this command.
7. Use the command `aov()` to get an ANOVA and store the result under the name 'ANOVA1'. Are there significant differences between the superheroes severity of injuries?
8. You heard of equivalence of ANOVA and regression in the lecture. Adapt the same ANOVA to the data using the command `lm()` and store the result under 'ANOVA2'. Compare the two result objects. Can you identify corresponding results? Which analysis provides more information? Bonus: what exactly is meant by "heroHulk = -6.250"?
9. Do further analyses where you find out, whether the two comic universes "Marvel" and "DC" differ in severity of injuries. Do this by applying planned contrasts and use `contrasts(Superhero$hero) <- <your contrast>` for this. Spiderman and Hulk are members of Marvel, Superman is in the DC universe and Ninja Turtles are not analysed here.

## 2) Two factorial ANOVA

1. Read the data `ChickFlick.dat` using either `read_delim()` or `read_tsv()` from the URL <https://pzezula.pages.gwdg.de/data/ChickFlick.dat> and store it in a data object in R. You find data of physiological activation of men and women, that saw the classical "Chick-Flick" movie "Bridget Jones' Diary" or the Thriller "Memento".
2. Generate a bar plot with error bars included.

3. Check the validity of the concept “chick-flick” by finding out whether we have different reactions to the films due to gender.
4. Do you trust in the generalizability of the results of the above computation? Do you see alternative explanations? How could we make our results more trustworthy?

### 3) Render

Render (or knit) your Rmd file using the shortkey `strg + shift + k` (Windows) or `cmd + shift + k` (Mac). If it works, well done! If not, check your error messages, inspect the lines of your code, where the error is supposed to occur. Correct the error and over again.

### Literature

*Annotation:* This exercise sheet bases in part on exercises, that you can find in the textbook *Discovering Statistics Using R* (Field, Miles & Field, 2012). They were modified for the purpose of this sheet and the R-code was actualized.

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. London: SAGE Publications Ltd.

Version: 20 Mai, 2021 08:25