

# Übungszettel: Regression

M.Psy.205, Dozent: Dr. Peter Zezula

Kai Schneider ([k.schneider01@stud.uni-goettingen.de](mailto:k.schneider01@stud.uni-goettingen.de))

20 Mai, 2021 09:06

## Deutsche Version

### Links

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

### Ziel des Übungszettels

1. Datawrangling wiederholen
2. ggplot wiederholen
3. Regressionsanalyse wiederholen
4. Zusammenhang zwischen t.test für unabhängige Stichproben und der einfachen linearen Regression nominalen binären Prädiktor verstehen.

### Hinweise zur Bearbeitung

1. 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.
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, Kapitel 7 (7.1 - 7.5, 7.9)	Buchkapitel, das Schritt für Schritt erklärt, worum es geht, und wie man Regressionen in R durchführt. <b>Große Empfehlung!</b>
<a href="#">R for Data Science</a>	Einsteiger-Buch von R-Gott Hadley Wickham. Hier wird topaktuell in die Arbeit mit R, insbesondere zur Datenaufbereitung und Visualisierung, eingeführt.
<a href="#">R Tutorial</a>	Schritt-für-Schritt Einführung in das Arbeiten mit R von Christian Treffenstädt. Nützlich, falls Sie grundlegende Dinge noch einmal nachschlagen möchten

## 1) Daten einlesen

1. Setzen Sie ein sinnvolles Arbeitsverzeichnis für den Übungszettel (in der Regel der Ordner, in dem Ihre .Rmd liegt). Aber Vorsicht: Beim Rendern (Knit) geht RStudio davon aus, dass das Working-Directory das ist, in dem auch die .Rmd-Datei liegt. Dies ist besonders wichtig, wenn es um relative Links geht.
2. Laden Sie den Datensatz [starwars.csv](#) herunter (*rechtsklick > Ziel speichern unter* oder *rechtsklick > Verknüpfte Datei laden*) und speichern Sie ihn in Ihrem Arbeitsverzeichnis (idealerweise haben Sie noch den Ordner vom letzten Übungszettel - speichern Sie den Datensatz im Unterordner /data).
3. Laden Sie die Pakete des **tidyverse** und fügen Sie eine entsprechende Code-Zeile an den Beginn Ihres Dokuments ein.
4. Lesen Sie den Datensatz `starwars.csv` unter dem Namen `sw_data` in R ein.

## 2) Data Wrangling

1. Erstellen Sie einen neuen Datensatz `sw_data2` welcher keine fehlenden Werte auf den Variablen `gender` und `height` enthält.
2. Erstellen Sie einen Barplot, welcher auf der x-Achse das Geschlecht und auf der y-Achse die mittlere Größe zeigt.
3. Ergänzen sie Fehlerbalken, färben Sie die Balken unterschiedlich ein und geben Sie ihrer Grafik einen passenden Titel. Was fällt ihnen bezüglich der Fehlerbalken auf? Warum werden für Hemaphroditen und Geschlechtslose keine Fehlerbalken gezeichnet?
4. Lassen Sie sich anzeigen wieviele Beobachtungseinheiten es je Geschlecht gibt. Lassen Sie sich auch den Mittelwert, Median und Varianz der Variable `height` für jedes Geschlecht angeben.
5. Erstellen Sie einen neuen Datensatz `sw_data3` welcher keine fehlenden Werte auf den Variablen `gender` und `height` enthält und in welchem nur noch die Geschlechter "male" und "female" vorkommen

## 3) t-test für unabhängige Stichproben

1. Nutzen Sie die Funktion `t.test()` um zu überprüfen, ob die Mittelwerte der Körpergröße zwischen Männern und Frauen signifikant voneinander abweichen. Nehmen Sie dafür (einfachhalber) Varianzhomogenität und Unabhängigkeit der Stichproben an. Testen Sie zweiseitig.
2. Treffen Sie eine Testentscheidung. Unterscheiden sich beide Gruppen signifikant voneinander? Genau genommen: Was testet der eigentlich?
3. Wie groß ist die Mittelwertsdifferenz?

## 4) Einfache Regression: Zusammenhang Lineares Modell und t-Test

1. Rechnen Sie eine einfache lineare Regression mit abhängiger Variable `height` und Prädiktor `gender`.

2. Vergleichen Sie die Mittelwerte beider Gruppen mit dem Intercept und Slope der Regression. Erinnern Sie sich an die Mittelwertsdifferenz. Was fällt ihnen auf?
3. Schauen Sie den t-Wert und p-Wert für den Signifikanztest ob der Slope signifikant von Null verschieden ist an. Vergleichen Sie diesen mit Ergebnis des t-Tests aus Aufgabe 3.

## 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 09:06