

# 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](#)

### Übungszettel mit Lösungen

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

[Der gesamte Übungszettel als .Rmd-Datei](#) (Zum Downloaden: Rechtsklick > Speichern unter...)

### 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 Turtles  
Wir 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.

## Lösung

```
library(tidyverse)
```

### Unteraufgabe 1

```
Superhero <- read_delim("https://pzezula.pages.gwdg.de/data/Superhero.dat", delim = "\t")
```

### Unteraufgabe 2

```
##  
## -- Column specification -----  
## cols(  
##   hero = col_double(),  
##   injury = col_double()  
## )
```

```
Falsch <- lm(injury ~ hero, data = Superhero)  
summary(Falsch)
```

### Unteraufgabe 3

```
##  
## Call:  
## lm(formula = injury ~ hero, data = Superhero)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -29.841 -10.992  -0.323   8.427  41.838   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)   56.520      6.886   8.207 6.21e-09 ***   
## hero          -6.679      2.476  -2.697  0.0117 *     
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 15.56 on 28 degrees of freedom  
## Multiple R-squared:  0.2062, Adjusted R-squared:  0.1779   
## F-statistic: 7.275 on 1 and 28 DF,  p-value: 0.01171  
  
## die Heldenvariable wird als numerische Variable ausgewertet.
```

Diese Auswertung interpretiert die Zahlencodes der Helden als intervallskalierten Prädiktor. Solange Superman aber nicht um genausoviel heldiger als Spiderman ist, wie die Ninja Turtles als Hulk, ist dies aber Unsinn.

```

# we define a new variable hero.f with factorized categorial levels and label them also
Superhero$hero.f <- factor(Superhero$hero,
  levels = c(1:4), labels =
    c("Spiderman", "Superman", "Hulk", "Ninja Turtle")
)
Superhero$hero.f

```

#### Unteraufgabe 4

```

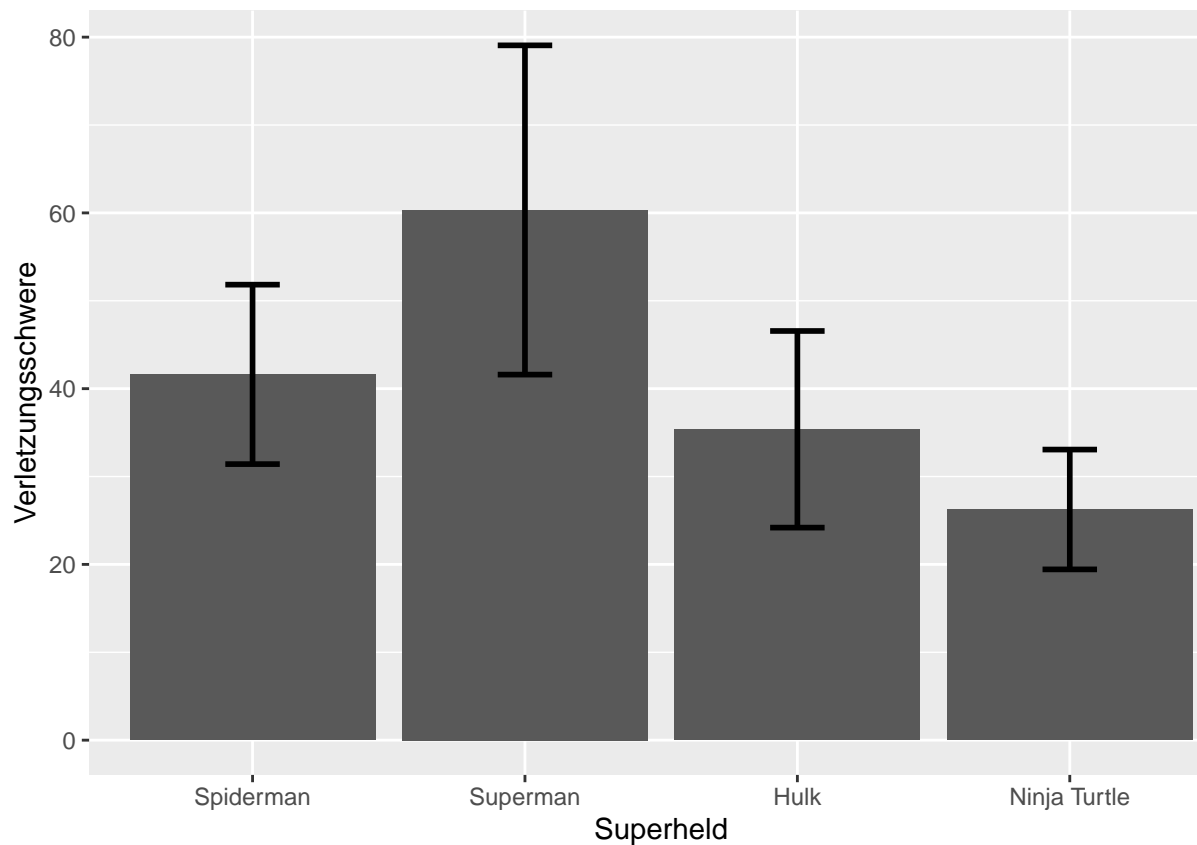
## [1] Spiderman Spiderman Spiderman Spiderman Spiderman Spiderman Spiderman Spide
## [9] Superman Superman Superman Superman Superman Superman Hulk Hulk
## [17] Hulk Hulk Hulk Hulk Hulk Hulk Ninja Turtle Ninja Turtle
## [25] Ninja Turtle Ninja Turtle Ninja Turtle Ninja Turtle Ninja Turtle
## Levels: Spiderman Superman Hulk Ninja Turtle

```

```

# library(Hmisc)
bar <- ggplot(Superhero, aes(hero.f, injury))
bar + stat_summary(fun = mean, geom = "bar") +
  stat_summary(fun.data = mean_cl_normal, geom = "errorbar",
    width = 0.2, size = 1) +
  labs(x = "Superheld", y = "Verletzungsschwere")

```



#### Unteraufgabe 5

```
??levene
car::leveneTest(Superhero$injury, Superhero$hero.f)
```

### Unteraufgabe 6

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 3   0.827  0.491
##      26
```

Nicht signifikant, auf Varianzhomogenität wird geschlossen

```
ANOVA1 <- aov(injury ~ hero.f, Superhero)
summary(ANOVA1)
```

### Unteraufgabe 7

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## hero.f      3   4181   1393.5    8.317 0.000483 ***
## Residuals  26   4357    167.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Signifikant, die Superhelden unterscheiden sich hinsichtlich der Verletzungsschwere

```
ANOVA2 <- lm(injury ~ hero.f, Superhero)
summary(ANOVA2)
```

### Unteraufgabe 8

```
##
## Call:
## lm(formula = injury ~ hero.f, data = Superhero)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.333  -8.250   1.688   7.562  24.667
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    41.625     4.577    9.095 1.47e-09 ***
## hero.fSuperman  18.708     6.991    2.676  0.0127 *
## hero.fHulk      -6.250     6.472   -0.966  0.3431
## hero.fNinja Turtle -15.375     6.472   -2.376  0.0252 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.94 on 26 degrees of freedom
## Multiple R-squared:  0.4897, Adjusted R-squared:  0.4308
## F-statistic: 8.317 on 3 and 26 DF,  p-value: 0.0004828
```

Die F-statistic in ANOVA2 entspricht ANOVA1. Zusätzlich enthält ANOVA2 Informationen über die verwendeten Dummy-Variablen, sodass gerichtete Aussagen über einzelne Mittelwerte möglich werden.

Bonus: heroHulk = -6.250 bedeutet, dass Kinder im Hulkkostüm im Durchschnitt um 6.25 Einheiten weniger schlimme Verletzungen hatten als Kinder in der Referenzgruppe unserer Dummykodierung, also Kinder im Spidermankostüm.

```
contrasts(Superhero$hero.f) <- c(1, -2, 1, 0)
MarvelVsDC <- lm(injury ~ hero.f, Superhero)
summary.lm(MarvelVsDC)
```

## Unteraufgabe 9

```
##
## Call:
## lm(formula = injury ~ hero.f, data = Superhero)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.333  -8.250   1.688   7.562  24.667
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  40.8958     2.3817  17.171 1.05e-15 ***
## hero.f1      -7.2778     2.0656  -3.523 0.001598 **
## hero.f2     -0.6036     4.5796  -0.132 0.896160
## hero.f3    -17.4690     4.6367  -3.768 0.000855 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.94 on 26 degrees of freedom
## Multiple R-squared:  0.4897, Adjusted R-squared:  0.4308
## F-statistic: 8.317 on 3 and 26 DF,  p-value: 0.0004828
```

Der erste Kontrast hero1 wird signifikant, das negative Vorzeichen zeigt, dass DC mit schwereren Verletzungen assoziiert ist. Da wir nur einen der drei möglichen orthogonalen Kontraste definiert haben, füllt R die beiden anderen mit zu unserem Kontrast und zueinander orthogonalen, anderen Kontrasten auf. Da unsere Hypothese aber schon vom ersten Kontrast erschöpfend geprüft wird, können wir die Kontraste hero2 und hero3 ignorieren.

## 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?

## Lösung

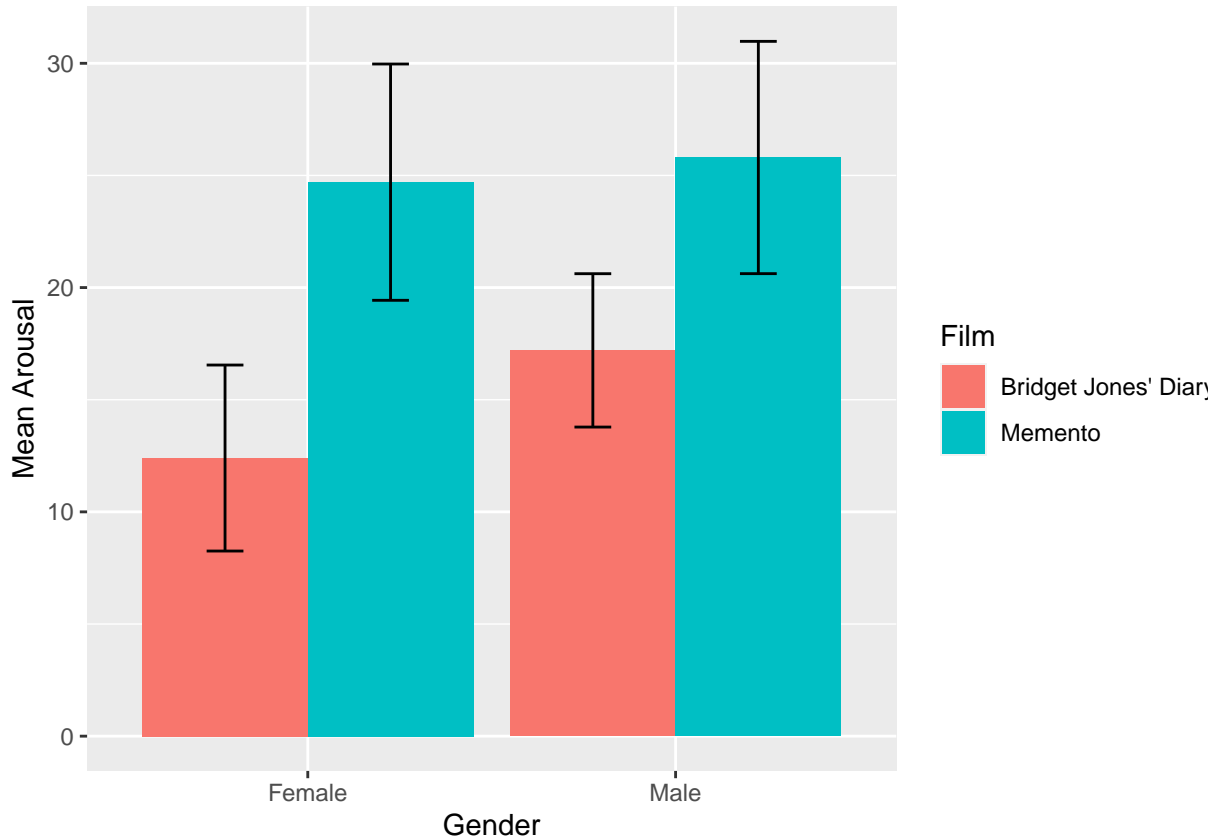
```
ChickFlick <- read_delim("https://pzezula.pages.gwdg.de/data/ChickFlick.dat", delim = "\t")
```

### Unteraufgabe 1

```
##  
## -- Column specification -----  
## cols(  
##   gender = col_character(),  
##   film = col_character(),  
##   arousal = col_double()  
## )
```

```
# we set factor  
ChickFlick$gender <- factor(ChickFlick$gender)  
ChickFlick$film <- factor(ChickFlick$film)
```

```
# library(Hmisc)  
bar <- ggplot(ChickFlick, aes(gender, arousal, group = film, fill = film))  
bar + stat_summary(fun = mean, geom = "bar", position = "dodge") +  
  stat_summary(fun.data = mean_cl_normal,  
              geom = "errorbar",  
              position = position_dodge(width = 0.90),  
              width = 0.2) +  
  labs(x = "Gender", y = "Mean Arousal", fill = "Film")
```



### Unteraufgabe 2

```
ChickFlickANOVA <- aov(arousal ~ gender * film, ChickFlick)
summary(ChickFlickANOVA)
```

### Unteraufgabe 3

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## gender      1   87.0    87.0    2.135    0.153
## film        1 1092.0  1092.0  26.785 8.78e-06 ***
## gender:film  1   34.2    34.2    0.839    0.366
## Residuals  36 1467.7    40.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Der Interaktionsterm wird nicht signifikant, unser Experiment kann das Konzept, dass Männer und Frauen unterschiedlich auf verschiedene Filme reagieren, nicht untermauern.

**Unteraufgabe 4** Für ein so schwammiges Konzept wie “Chick-Flick” benötigen wir valideres, und vor allem mehr Stimulusmaterial. Die beiden gewählten Filme scheinen zwar recht gute Prototypen ihrer Kategorien zu sein, jedoch ist nicht auszuschließen, dass etwa Bridget Jones’ Diary einfach insofern ein schlechter Chick-Flick ist, dass er es nicht schafft, die frauenspezifische Erregung dieses Genres zu erzeugen.



### 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](#)

**Exercise sheet with solutions included**

[Exercise sheet with solutions included as PDF](#)

The source code of this sheet as `.Rmd` (Right click and “store as” to download ...)

### 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 downloading 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.

### Solutions

```
library(tidyverse)
```

### Subtask 1

```
Superhero <- read_delim("https://pzezula.pages.gwdg.de/data/Superhero.dat", delim = "\t")
```

## Unteraufgabe 2

```
##  
## -- Column specification -----  
## cols(  
##   hero = col_double(),  
##   injury = col_double()  
## )
```

```
# or  
Superhero <- read_tsv("https://pzezula.pages.gwdg.de/data/Superhero.dat")
```

```
##  
## -- Column specification -----  
## cols(  
##   hero = col_double(),  
##   injury = col_double()  
## )
```

```
Wrong <- lm(injury ~ hero, data = Superhero)  
summary(Wrong)
```

## Subtask 3

```
##  
## Call:  
## lm(formula = injury ~ hero, data = Superhero)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -29.841 -10.992  -0.323   8.427  41.838   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)   56.520     6.886   8.207 6.21e-09 ***   
## hero          -6.679     2.476  -2.697  0.0117 *     
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 15.56 on 28 degrees of freedom  
## Multiple R-squared:  0.2062, Adjusted R-squared:  0.1779   
## F-statistic: 7.275 on 1 and 28 DF,  p-value: 0.01171
```

In this calculation we would interpret group codes as continuous variable (interval-scaled). This is nonsense.

```

Superhero$hero.f <- factor(Superhero$hero,
  levels = c(1:4), labels =
    c("Spiderman", "Superman", "Hulk", "Ninja Turtle")
)
Superhero$hero.f

```

#### Subtask 4

```

## [1] Spiderman Spiderman Spiderman Spiderman Spiderman Spiderman Spiderman Spide
## [9] Superman Superman Superman Superman Superman Superman Hulk Hulk
## [17] Hulk Hulk Hulk Hulk Hulk Hulk Ninja Turtle Ninja Turtle
## [25] Ninja Turtle Ninja Turtle Ninja Turtle Ninja Turtle Ninja Turtle
## Levels: Spiderman Superman Hulk Ninja Turtle

```

```

# library(Hmisc)
bar <- ggplot(Superhero, aes(hero.f, injury))
bar + stat_summary(fun.y = mean, geom = "bar") +
  stat_summary(fun.data = mean_cl_normal, geom = "errorbar",
    width = 0.2, size = 1) +
  labs(x = "Superhero", y = "Severity of Injuries")

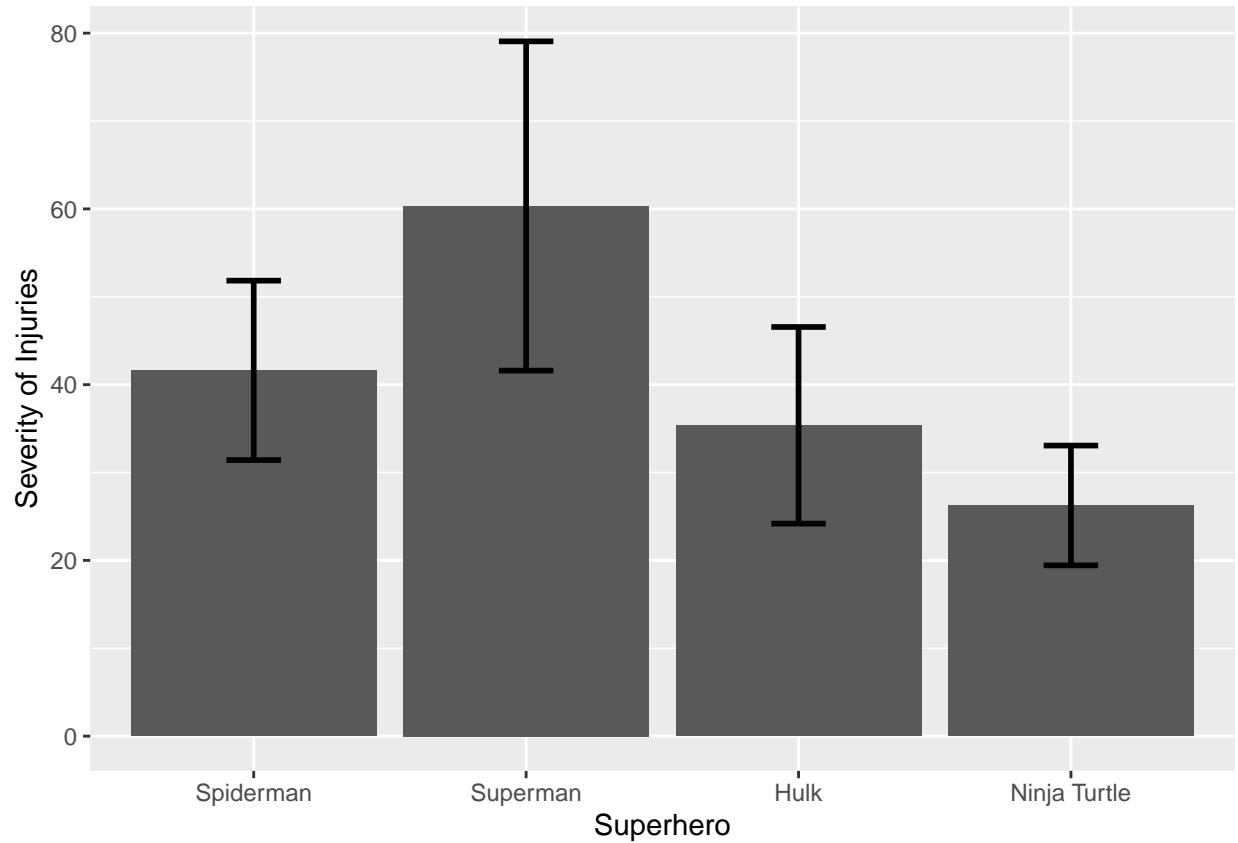
```

#### Subtask 5

```

## Warning: `fun.y` is deprecated. Use `fun` instead.

```



```
# we would enter
# ??levene
# but this doesnt make sense in a Rmd that will be rendered
# so we comment it out

# leveneTest() is part of package car
# we call it without loading it
car::leveneTest(Superhero$injury, Superhero$hero.f)
```

### Subtask 6

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 3  0.827  0.491
##      26
```

No significant differences of variances. So we assume homogeneity of variances.

```
ANOVA1 <- aov(injury ~ hero.f, Superhero)
summary(ANOVA1)
```

## Subtask 7

```
##           Df Sum Sq Mean Sq F value   Pr(>F)
## hero.f      3  4181  1393.5    8.317 0.000483 ***
## Residuals  26  4357   167.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We find significant differences of the severity of injuries between the types of superheroes.

```
ANOVA2 <- lm(injury ~ hero.f, Superhero)
summary(ANOVA2)
```

## Subtask 8

```
##
## Call:
## lm(formula = injury ~ hero.f, data = Superhero)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.333  -8.250   1.688   7.562  24.667
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      41.625     4.577   9.095 1.47e-09 ***
## hero.fSuperman    18.708     6.991   2.676  0.0127 *
## hero.fHulk        -6.250     6.472  -0.966  0.3431
## hero.fNinja Turtle -15.375     6.472  -2.376  0.0252 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.94 on 26 degrees of freedom
## Multiple R-squared:  0.4897, Adjusted R-squared:  0.4308
## F-statistic: 8.317 on 3 and 26 DF,  p-value: 0.0004828
```

The F-statistics correspond in ANOVA1 and ANOVA2. ANOVA2 gives us additional informations about the used dummy variables. So there are some specific tests or comparisons between pairs of means.

Bonus: `heroHulk = -6.250` tells us, that children dressed like Hulk had 6.25 units less severe injuries than the reference group of our dummy coding system. Here the reference group would be children dressed up like Spidermen.

```
contrasts(Superhero$hero.f) <- c(1, -2, 1, 0)
MarvelVsDC <- lm(injury ~ hero.f, Superhero)
summary.lm(MarvelVsDC)
```

## Subtask 9

```
##
## Call:
## lm(formula = injury ~ hero.f, data = Superhero)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -28.333  -8.250   1.688   7.562  24.667
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  40.8958     2.3817  17.171 1.05e-15 ***
## hero.f1      -7.2778     2.0656  -3.523 0.001598 **
## hero.f2      -0.6036     4.5796  -0.132 0.896160
## hero.f3     -17.4690     4.6367  -3.768 0.000855 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.94 on 26 degrees of freedom
## Multiple R-squared:  0.4897, Adjusted R-squared:  0.4308
## F-statistic: 8.317 on 3 and 26 DF,  p-value: 0.0004828
```

The first contrast `hero1` is highly significant. The ‘-’ shows, that DC is associated with more severe injuries. As we defined only one of the three possible orthogonal contrasts, R completes the two missing orthogonal contrasts. But our hypothesis is already tested by our first contrast, so we can ignore the results of the other ones (`hero2` and `hero3`).

## 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?

### Solutions

```
ChickFlick <- read_delim("https://pzezula.pages.gwdg.de/data/ChickFlick.dat", delim = "\t")
```

#### Subtask 1

```
##
## -- Column specification -----
## cols(
##   gender = col_character(),
##   film = col_character(),
##   arousal = col_double()
## )
```

```
# or
ChickFlick <- read_tsv("https://pzezula.pages.gwdg.de/data/ChickFlick.dat")
```

```
##
## -- Column specification -----
## cols(
##   gender = col_character(),
##   film = col_character(),
##   arousal = col_double()
## )
```

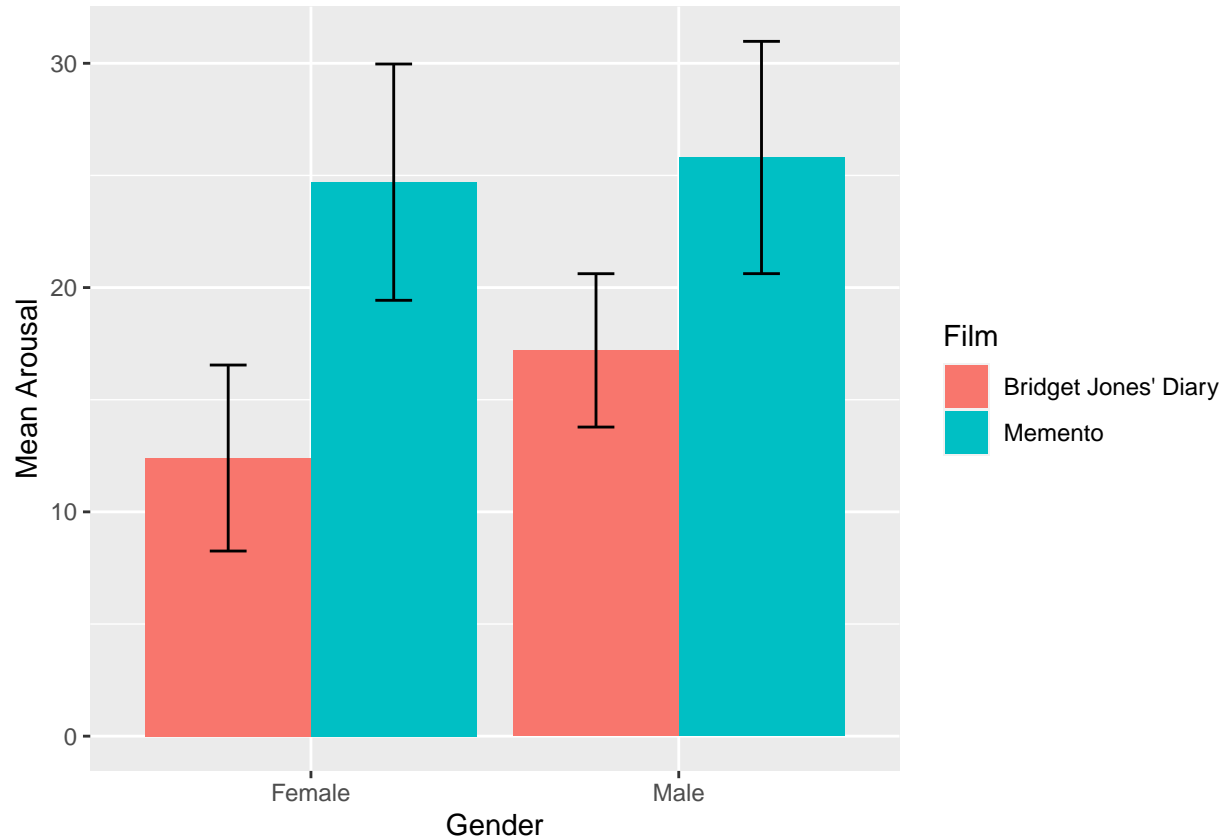
```
# we set factor
ChickFlick$gender <- factor(ChickFlick$gender)
ChickFlick$film <- factor(ChickFlick$film)
```

```
# library(Hmisc)
bar <- ggplot(ChickFlick, aes(gender, arousal, group = film, fill = film))
bar + stat_summary(fun.y = mean, geom = "bar", position = "dodge") +
  stat_summary(fun.data = mean_cl_normal,
              geom = "errorbar",
              position = position_dodge(width = 0.90),
              width = 0.2) +
  labs(x = "Gender", y = "Mean Arousal", fill = "Film")
```

## Subtask 2

```
## Warning: `fun.y` is deprecated. Use `fun` instead.
```





```
ChickFlickANOVA <- aov(arousal ~ gender * film, ChickFlick)
summary(ChickFlickANOVA)
```

**Subtask 3**

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## gender      1   87.0    87.0    2.135    0.153
## film        1 1092.0  1092.0   26.785 8.78e-06 ***
## gender:film 1   34.2    34.2    0.839    0.366
## Residuals  36 1467.7    40.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We cannot prove our interaction term be significant. So we don't have evidence in our data that men and women react different to the movies in question.

**Subtask 4** We really need more and especially more valid stimulus material to get clearer about a concept as nebulous as “Chick-Flick”. Although both movies seem to be good prototypes of their category, “Bridget Jones’ Diary” might not cause the specific female activation pattern that this genre is supposed to produce.

### 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 08:26