

Übungszettel MANOVA

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

Johannes Brachem (johannes.brachem@stud.uni-goettingen.de)

Deutsche Version

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 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, Kapitel 16	Buchkapitel, das Schritt für Schritt erklärt, worum es geht, und wie man Multivariate Varianzanalysen in R durchführt. Große Empfehlung!

Tipp der Woche

Hinweis: Der Tipp ist diese Woche etwas länger.

1) Befehle nutzen, ohne Pakete zu laden

Normalerweise müssen Sie ein Paket mit `library()` laden, damit Sie Funktionen aus diesem Paket nutzen können. Wenn Sie eine Funktion aber nur sehr selten benötigen, können Sie einen Trick verwenden, nämlich den doppelten Doppelpunkt `::`. Das funktioniert nach dem Schema `package::function()` und wird von uns auch benutzt, um Funktionen mit häufigen Namen richtig einzusetzen.

Wenn Sie diese Schreibweise verwenden, muss das jeweilige Paket nicht geladen sein. Beispiel: `psych::describe()`. Das Paket muss allerdings installiert sein, damit diese Schreibweise funktioniert.

2) Wie genau funktioniert eigentlich die Pipe? %>%

Die Pipe kommt aus dem Paket `magrittr` und wird im `tidyverse` häufig verwendet. Sie kennen die Pipe vor allem aus `dplyr`, aber wenn Sie `magrittr` oder `dplyr` geladen haben (d.h. mit `library()` aktiviert), dann können Sie die Pipe für fast alle Befehle in R benutzen, wenn Sie möchten. Das funktioniert so:

Der Code, der auf der *linken* Seite der Pipe steht, wird von R "unter der Haube" auf der *rechten* Seite der Pipe eingesetzt, und zwar standardmäßig immer als erstes *Argument* einer *Funktion*, so dass andere Argumente der Funktion mit einem Komma dahinter gestellt werden. Wenn Sie den Code auf der linken Seite der Pipe an eine andere Stelle auf der rechten Seite der Funktion einsetzen möchten, dann können Sie R mit einem Punkt `.` sagen, wo der Code der linken Seite eingesetzt werden soll (siehe Beispiel 4).

Beispiel 1

Nehmen wir mal an, wir wollen nur VP über 18 auswählen.

```
# ohne Pipe
filter(example_data, age > 18)

# mit Pipe
example_data %>% filter(age > 18)
```

Beispiel 2

Das ist besonders nützlich, wenn Befehle verschachtelt sind. Nehmen wir mal an, wir wollen nur VP über 18 auswählen und uns nur die Gruppe und unsere abhängige Variable anzeigen lassen.

```
# ohne Pipe
select(filter(example_data, age > 18), group, dependent_variable)

# mit Pipe
example_data %>%
  filter(age > 18) %>%
  select(group, dependent_variable)
```

Beispiel 3

Das funktioniert für fast alle Funktionen. Hinweis: `na.rm = TRUE` sorgt dafür, dass Missing Values bei der Berechnung des Mittelwerts weggelassen werden.

```
# ohne Pipe
mean(age, na.rm = TRUE)

# mit Pipe
age %>% mean(na.rm = TRUE)
```

Beispiel 4

```
# ohne Pipe
lm(dependent_variable ~ group, data = example_data)

# mit Pipe
example_data %>% lm(dependent_variable ~ group, data = .)
```

Vorgehen

Hinweis: Screenshot aus Field, Miles & Field (2012). Für fast alle Verfahren finden Sie in diesem Lehrbuch ähnliche Übersichten. Diese sind sehr wertvoll für die Prüfungsvorbereitung und darüber hinaus.

16.6.2. General procedure for MANOVA ^①

To conduct factorial MANOVA you should follow this general procedure:

- 1 *Enter data.*
- 2 *Explore your data:* begin by graphing the data and computing descriptive statistics. You should check multivariate normality and take a look at the variance–covariance matrices for each group.
- 3 *Set contrasts for all predictor variables:* you need to decide what contrasts to do and to specify them appropriately for all of the independent variables in your analysis.
- 4 *Compute the MANOVA:* you can then run the main multivariate analysis of variance. Depending on what you found in the previous step, you might need to run a robust version of the test.
- 5 *Run univariate ANOVAs:* having conducted the MANOVA, you can follow it up with separate ANOVAs for each dependent variable.
- 6 *Discriminant function analysis:* better than the option above, consider running a discriminant function analysis.

1) Daten einlesen

1. Laden Sie die nötigen Pakete und setzen Sie ein sinnvolles Arbeitsverzeichnis.
2. Laden Sie den Datensatz `ocd_data.dat` über den Link https://pzezula.pages.gwdg.de/data/ocd_data.dat in R ein.
3. Kodieren Sie die Variable `group` in `ocd_data` als Faktor mit sinnvoller Baseline. Geben Sie der Gruppe "No Treatment Control" das Label "NT".

Lösung

Unteraufgabe 1

```
library(tidyverse)

setwd("~/ownCloud/_Arbeit/Hiwi Peter/gitlab_sheets")
```

Unteraufgabe 2

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

## Parsed with column specification:
## cols(
##   group = col_character(),
##   actions = col_double(),
##   thoughts = col_double()
## )
```

Unteraufgabe 3

```
ocd_data$group <- ocd_data$group %>%
  factor(levels = c("No Treatment Control", "BT", "CBT"),
        labels = c("NT", "BT", "CBT"))
```

2) Überblick über die Daten

Bedeutung der Variablen

Unser Beispieldatensatz enthält hypothetische Daten zur Evaluation von Therapieprogrammen bei Zwangsstörungen.

Variable	Beschreibung
group	Faktor, der angibt, welche Art von Therapie die Versuchsperson erhalten hat. NT = Keine Therapie, BT = Verhaltenstherapie, CBT = Kognitive Verhaltenstherapie
actions	Häufigkeit von Zwangshandlungen nach der Behandlung
thoughts	Häufigkeit von Zwangsgedanken nach der Behandlung

1. Erstellen Sie zunächst einen einfachen Scatterplot, der den Zusammenhang zwischen zwanghaftem Verhalten auf der x-Achse und zwanghaften Gedanken auf der y-Achse darstellt.
2. Fügen Sie dem Plot eine Regressionslinie hinzu.
3. Nutzen Sie den Befehl `facet_wrap()`, um den Plot aus Aufgabe 2.2 nach Gruppen getrennt darzustellen.
4. Nutzen Sie den Code `ocd_data %>% select(actions, thoughts) %>% by(ocd_data$group, cov)`, um sich die Varianz-Kovarianz-Matrizen für jede Gruppe anzeigen zu lassen.
5. Nutzen Sie den Code `ocd_data %>% by(ocd_data$group, psych::describe)`, um sich für jede Gruppe deskriptive Daten ausgeben zu lassen. Können Sie den Befehl verstehen? *Falls Sie eine Fehlermeldung bekommen, installieren Sie das Paket `psych`.*
6. Führen Sie den unten stehenden Code aus, um die Annahme der multivariaten Normalverteilung an Ihren Daten zu überprüfen. Können Sie den Code verstehen? Sind die Daten in jeder Gruppe multivariat Normalverteilt? *Falls Sie eine Fehlermeldung bekommen, installieren Sie das Paket `mvnortest`.*
7. Der Box's M-Test prüft die Gleichheit der Varianz-Covarianz-Matrizen. Führen Sie den Befehl `boxM()` aus der `library(heplots)` durch, um diese Voraussetzung zu überprüfen. Können wir von Gleichheit der Varianz-Covarianz-Matrizen ausgehen? **Falls Sie eine Fehlermeldung bekommen, installieren Sie das Paket `heplots`.**

```
# Daten vorbereiten
nt <- ocd_data %>% filter(group == "NT") %>% select(2:3) %>% t()
bt <- ocd_data %>% filter(group == "BT") %>% select(2:3) %>% t()
cbt <- ocd_data %>% filter(group == "CBT") %>% select(2:3) %>% t()

# Tests durchführen
mvnortest::mshapiro.test(nt)
mvnortest::mshapiro.test(bt)
mvnortest::mshapiro.test(cbt)
```

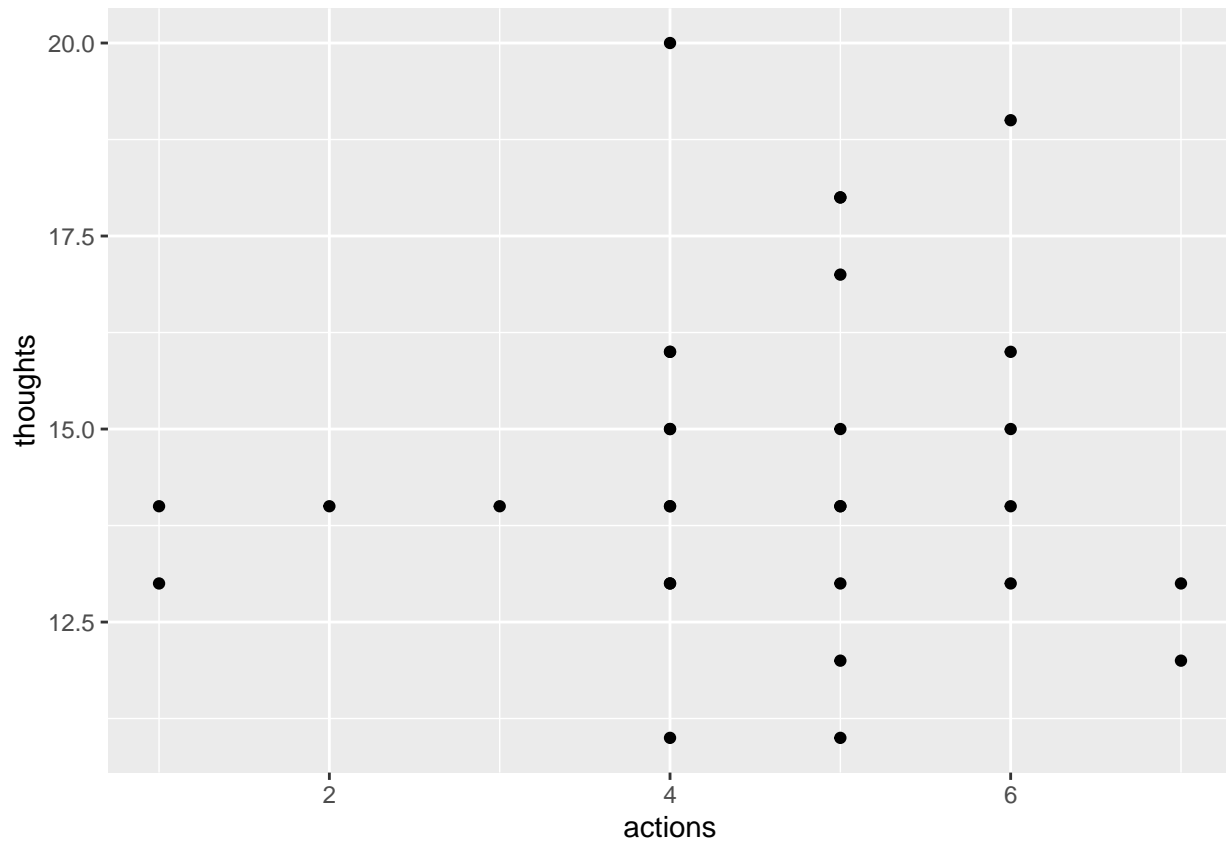
Lösung

Unteraufgabe 1

```
# Objekt erstellen
baseplot <- ggplot(ocd_data, aes(x = actions, y = thoughts))

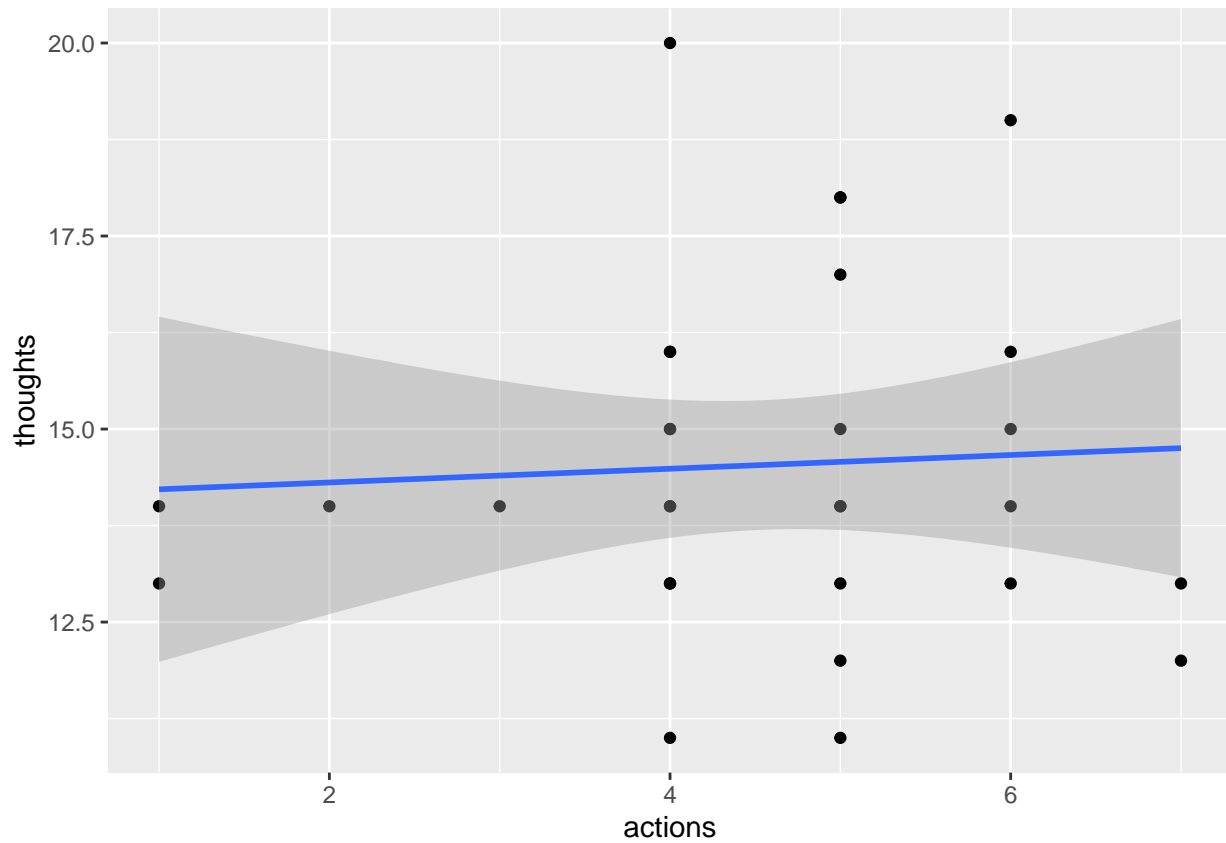
# Punkte hinzufügen
scatterplot <- baseplot + geom_point()

# Plot anzeigen
scatterplot
```



Unteraufgabe 2

```
# Regressionslinie hinzufügen  
lineplot <- scatterplot + geom_smooth(method = "lm")  
  
# Plot anzeigen  
lineplot  
  
## `geom_smooth()` using formula 'y ~ x'
```



Unteraufgabe 3

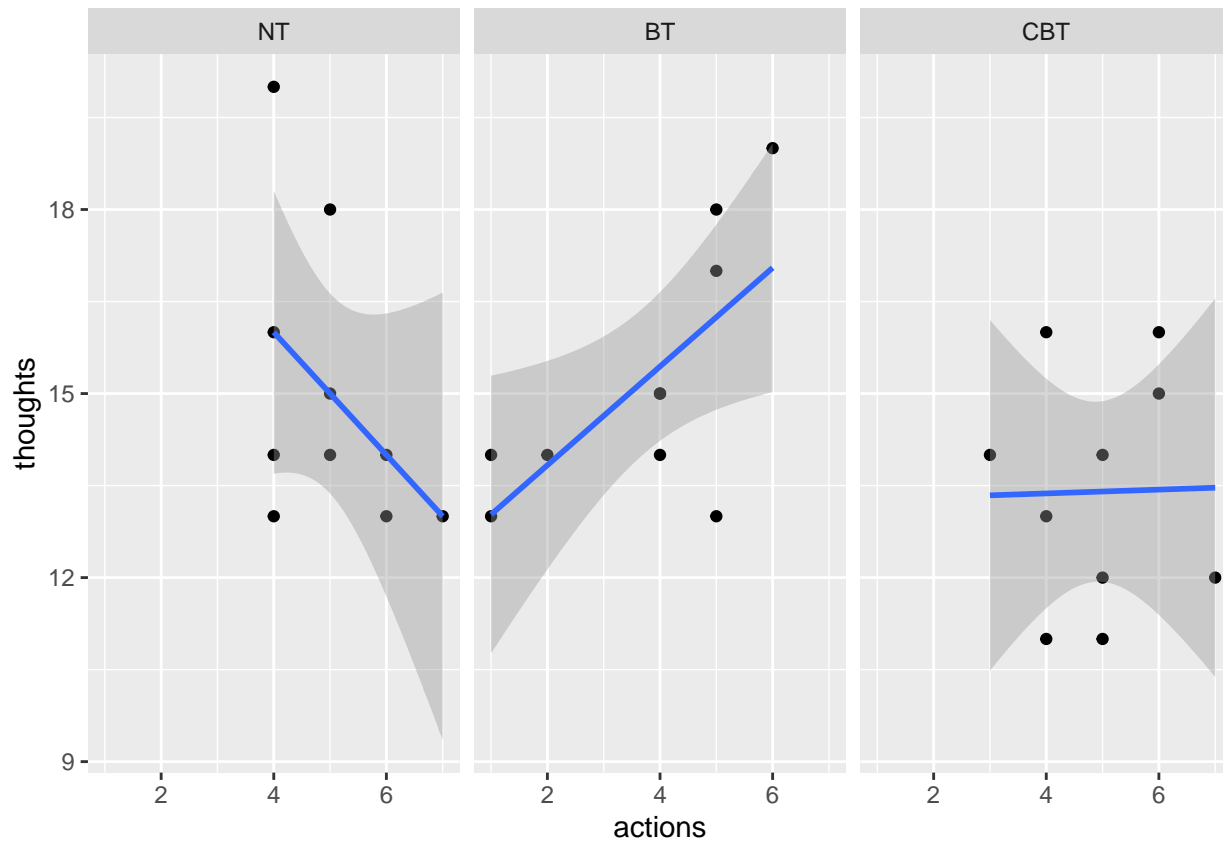
```

# Plot aufteilen
groupplot <- lineplot + facet_wrap(~ group)

# Plot anzeigen
groupplot

## `geom_smooth()` using formula 'y ~ x'

```



Unteraufgabe 4

```
ocd_data %>% select(actions, thoughts) %>% by(ocd_data$group, cov)
```

```
## ocd_data$group: NT
##      actions thoughts
## actions  1.111111 -1.111111
## thoughts -1.111111  5.555556
## -----
## ocd_data$group: BT
##      actions thoughts
## actions  3.122222  2.511111
## thoughts 2.511111  4.400000
## -----
## ocd_data$group: CBT
##      actions thoughts
## actions  1.43333333 0.04444444
## thoughts 0.04444444 3.60000000
```

Unteraufgabe 5

```
ocd_data %>% by(ocd_data$group, psych::describe)
```

```
## ocd_data$group: NT
##      vars  n mean  sd median trimmed mad min max range skew kurtosis  se
## group*   1 10   1 0.00   1   1.00 0.00  1  1    0 NaN      NaN 0.00
## actions  2 10   5 1.05   5   4.88 1.48  4  7    3 0.51   -1.22 0.33
## thoughts 3 10  15 2.36  14  14.62 1.48 13 20    7 0.96   -0.54 0.75
```

```
## -----
## ocd_data$group: BT
##      vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
## group*    1 10  2.0 0.00   2.0   2.00 0.00   2  2    0  NaN      NaN 0.00
## actions   2 10  3.7 1.77   4.0   3.75 1.48   1  6    5 -0.46  -1.45 0.56
## thoughts  3 10 15.2 2.10  14.5  15.00 1.48  13 19    6  0.61  -1.28 0.66
## -----
## ocd_data$group: CBT
##      vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
## group*    1 10  3.0 0.0   3.0   3.00 0.00   3  3    0  NaN      NaN 0.00
## actions   2 10  4.9 1.2   5.0   4.88 1.48   3  7    4  0.17  -1.18 0.38
## thoughts  3 10 13.4 1.9  13.5  13.38 2.22  11 16    5  0.09  -1.67 0.60
```

Unteraufgabe 6

Mit `t()` transponieren wir hier die Matrizen der Daten pro Gruppe. Vorher waren die Informationen von oben nach unten gespeichert (in Spalten), jetzt sind sie von links nach rechts gespeichert (in Zeilen). Das ist ungewöhnlich, aber nötig für den Test.

```
# Daten vorbereiten
nt <- ocd_data %>% filter(group == "NT") %>% select(2:3) %>% t()
bt <- ocd_data %>% filter(group == "BT") %>% select(2:3) %>% t()
cbt <- ocd_data %>% filter(group == "CBT") %>% select(2:3) %>% t()

# Tests durchführen
mvnormtest::mshapiro.test(nt)
```

```
##
## Shapiro-Wilk normality test
##
## data: Z
## W = 0.82605, p-value = 0.02998
mvnormtest::mshapiro.test(bt)
```

```
##
## Shapiro-Wilk normality test
##
## data: Z
## W = 0.89122, p-value = 0.175
mvnormtest::mshapiro.test(cbt)
```

```
##
## Shapiro-Wilk normality test
##
## data: Z
## W = 0.9592, p-value = 0.7767
```

Für die erste Gruppe haben wir ein signifikantes Ergebnis, d.h. dort sind die Daten nicht multivariat normalverteilt.

Für den Zweck dieses Übungszettels führen wir die Analyse trotzdem fort.

Unteraufgabe 7

```
res.boxm <- heplots::boxM(ocd_data[,c('actions', 'thoughts')], group=ocd_data$group)
res.boxm

##
```



```
## Box's M-test for Homogeneity of Covariance Matrices
##
## data:  ocd_data[, c("actions", "thoughts")]
## Chi-Sq (approx.) = 8.8932, df = 6, p-value = 0.1797
# summary(res.boxm) # for details
```

Der p-Wert des BoxM-Test ist nicht unter 0.05, also halten wir noch an der H0 fest, dass sich die Varianz-Covarianz-Matrizen nicht signifikant unterscheiden.

3) MANOVA durchführen

Vorab: Erklärung

MANOVAS können in R mit dem Befehl `manova()` durchgeführt werden. Dieser Befehl funktioniert genau so wie `lm()` und `aov()` in der Form: `manova(outcome ~ predictor, data = data)`. Der Unterschied ist, dass im Vorfeld alle verwendeten Outcome-Variablen mit dem `cbind()`-Befehl zu einem Objekt "zusammengeschnürt" werden.

1. Setzen Sie Kontraste für die Interpretation der Analyse. Das funktioniert genau so wie bei ANOVAs und Regressionen. Schauen Sie zur Not in Ihren Aufzeichnung von früheren Übungszetteln nach.
 - a) Erster Kontrast: Vergleich von BT und NT
 - b) Zweiter Kontrast: Vergleich von CBT und NT *Hinweis: Hier handelt es sich um nicht-orthogonale Kontraste. Das ist an dieser Stelle in Ordnung, weil wir nur eine Prädiktorvariable haben. (siehe Field, Kap. 16.6.6: Setting Contrasts)*
Bemerkung: Wir müssen keine Kontraste setzen. Bei den Default-Einstellungen würde BT zur Referenzgruppe und wir würden den Unterschied, auch der Gruppe CBT als Effekt prüfen.
2. Erstellen Sie das Outcome-Objekt, indem Sie `ocd_data$thoughts` und `ocd_data$actions` mit Hilfe von `cbind()` verbinden.
3. Nutzen Sie den Befehl `manova()`, um die Analyse durchzuführen. Speichern Sie das Ergebnis in einem Objekt.
4. Wenden Sie auf das Objekt aus 3.3 den Befehl `summary()` mit dem zusätzlichen Argument `intercept = TRUE` an.
5. Welchen Schluss können Sie aus dem Output ziehen?

Lösung

Unteraufgabe 1

```
.bt.vs.nt <- c(-1,1,0)
.cbt.vs.nt <- c(-1,0,1)

# Kontraste an den Faktor binden
contrasts(ocd_data$group) <- cbind(.bt.vs.nt, .cbt.vs.nt)

mean(ocd_data$actions)
```

```
## [1] 4.533333
```

```
ocd_data %>% group_by(group) %>% summarize(mean(actions))
```

```
## # A tibble: 3 x 2
##   group `mean(actions)`
##   <fct>      <dbl>
## 1 NT          5
## 2 BT         3.7
## 3 CBT        4.9
```

Unteraufgabe 2

```
outcome <- cbind(ocd_data$actions, ocd_data$thoughts)
```

Unteraufgabe 3

```
modell1 <- manova(outcome ~ group, data = ocd_data)
```

Unteraufgabe 4

```
summary(modell1, intercept = TRUE)
```

```
##              Df  Pillai approx F num Df den Df  Pr(>F)
## (Intercept)  1 0.98285   745.23     2    26 < 2e-16 ***
## group        2 0.31845     2.56     4    54 0.04904 *
## Residuals    27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Unteraufgabe 5

Der F-Test für die Gruppen wird signifikant ($F(4,54) = 2.56, p = .049$). Das bedeutet, dass die Art der Therapie die Zwanghaftigkeit der untersuchten Personen, gemessen durch Handlungen und Gedanken, beeinflusst hat. Mehr Schlüsse können wir erst einmal nicht ziehen.

4) MANOVA interpretieren

1. Wenden Sie den Befehl `summary.aov()` auf ihr MANOVA-Modell an.
2. Interpretieren Sie die Ergebnisse kurz.
 - a) Dürfen wir auf Grundlage dieser Ergebnisse die geplanten Kontraste untersuchen?
3. Erstellen Sie für jedes einzelne Outcome ein eigenes ANOVA-Modell und betrachten Sie den Output, um Ihre Kontraste zu interpretieren.
4. Was folgern Sie aus den Ergebnissen?

Lösung

Unteraufgabe 1

```
summary.aov(modell1)
```

```
## Response 1 :
##              Df Sum Sq Mean Sq F value Pr(>F)
## group        2  10.467   5.2333  2.7706 0.08046 .
## Residuals    27  51.000   1.8889
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response 2 :
##              Df Sum Sq Mean Sq F value Pr(>F)
## group        2  19.467   9.7333  2.1541 0.1355
## Residuals    27 122.000   4.5185
```

Unteraufgabe 2

Beide einzelnen ANOVAs liefern keine signifikanten Ergebnisse, d.h. wir haben keinen Hinweis darauf, dass besonders die Zwangsgedanken oder besonders die Zwangshandlungen von der Art der Therapie beeinflusst werden. Das ist sehr interessant, da die Therapie die Kombination beider Komponenten ja scheinbar durchaus beeinflusst.

Auf dieser Grundlage haben wir eigentlich keinen Anlass, die geplanten Kontraste zu untersuchen. Wir tun es dennoch, zu Demonstrationszwecken.

Unteraufgabe 3

```
out1 <- lm(actions ~ group, data = ocd_data)
out2 <- lm(thoughts ~ group, data = ocd_data)
```

```
summary(out1)
```

```
##
## Call:
## lm(formula = actions ~ group, data = ocd_data)
##
## Residuals:
##   Min     1Q  Median     3Q    Max
## -2.700 -0.975  0.100  1.075  2.300
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.5333     0.2509  18.067 <2e-16 ***
## group.bt.vs.nt  -0.8333     0.3549  -2.348  0.0264 *
## group.cbt.vs.nt  0.3667     0.3549   1.033  0.3106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.374 on 27 degrees of freedom
## Multiple R-squared:  0.1703, Adjusted R-squared:  0.1088
## F-statistic: 2.771 on 2 and 27 DF,  p-value: 0.08046
```

```
summary(out2)
```

```
##
## Call:
## lm(formula = thoughts ~ group, data = ocd_data)
##
## Residuals:
##   Min     1Q  Median     3Q    Max
##  -2.40  -1.40  -0.70   1.45   5.00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     14.5333     0.3881  37.448 <2e-16 ***
## group.bt.vs.nt   0.6667     0.5488   1.215  0.2350
## group.cbt.vs.nt -1.1333     0.5488  -2.065  0.0487 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.126 on 27 degrees of freedom
## Multiple R-squared:  0.1376, Adjusted R-squared:  0.07372
## F-statistic: 2.154 on 2 and 27 DF,  p-value: 0.1355
```

Unteraufgabe 4

Für Zwangshandlungen erweist sich die Verhaltenstherapie als signifikant besser als die Kontrollgruppe, für Zwangsgedanken dagegen die kognitive Verhaltenstherapie.

5) Rendern (knit)

Lassen Sie die Datei mit `Strg + Shift + K` (Windows) oder `Cmd + Shift + K` (Mac) rendern. Sie sollten nun im "Viewer" unten rechts eine "schön aufpolierte" Version ihrer Datei sehen. Falls das klappt: Herzlichen Glückwunsch! Ihr Code kann vollständig ohne Fehlermeldung gerendert werden. Falls nicht: Nur mut, das wird schon noch! Gehen Sie auf Fehlersuche! Ansonsten schaffen wir es ja in der Übung vielleicht gemeinsam.

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 Version

Links

[Exercise sheet as PDF](#)

Exercise sheet with solutions included

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

[The source code of this sheet as .Rmd](#) (Right click and "save as" to download ...)

Some hints

1. Please try to solve this sheet in an .Rmd file. You can create one from scratch using `File > New file > R Markdown . . .`. You can delete the text beneath *Setup Chunk* (starting from line 11). Alternatively, you can download our template file unter [this link](#) (right click > save as...).
2. You'll find a lot of the important information on the [website of this course](#)
3. Please don't hesitate to search the web for help with this sheet. In fact, being able to effectively search the web for problem solutions is a very useful skill, even R pros work this way all the time! The best starting point for this is the [R section on the programming site Stackoverflow](#)
4. On the R Studio website, you'll find highly helpful [cheat sheets](#) for many of R topics. The [base R cheat sheet](#) might be a good starting point.

Ressources

Since this is a hands-on seminar, we won't be able to present each and every new command to you explicitly. Instead, you'll find here references to helpful ressources that you can use for completing this sheets.

Ressource	Description
Field, chapter 16	Book chapter explaining step by step the why and how of multivariate analyses of variance in R. Highly recommended!

Hint of the week

Note: This week's hint will be a bit longer than usual.

1) Using commands without loading packages

Usually, you'll have to load a package using `library()` in order to use functions from that package. However, there's a trick for when you'll only need a function sporadically: Adding the package name in front of the function using a double colon like this: `package::function()`. This also helps in using a function with a name that occurs in multiple packages. E.g., you might

have noticed the warning `dplyr::filter()` masks `stats::filter()` when loading the tidyverse. Here, `filter()` alone will use the dplyr-command. If you want to use the filter-function from the stats-commands, you can do so by typing `stats::filter()`.

The relevant packages don't need to be loaded for this to work. However, they have to be installed on your computer.

2) How exactly does the pipe work? %>%

Originally, the pipe stems from the package `magrittr` and is used frequently in the tidyverse. You mostly know piping from `dplyr`, but as long as you have loaded either `magrittr` or `dplyr` (using the `library()`-function), you'll be able to use the pipe for almost all R commands - if that is something you want to do. It works like this:

R inserts the code on the *left hand side* of the pipe "under the hood" into the *right hand side* of the pipe, the default being using it as the first argument of a function. This way you can add additional arguments with a comma behind it. If you want the left code to be inserted at a different location, you can use the dot `.` to tell R where to put it (see example no. 4).

Example no. 1

Say you only want to include subjects older than 18.

```
# without pipe
filter(example_data, age > 18)

# with pipe
example_data %>% filter(age > 18)
```

Example no. 2

This is most useful when you have nested commands. Here, we only want to include subjects older than 18, and only look at the group variable and our dependent variable.

```
# without pipe
select(filter(example_data, age > 18), group, dependent_variable)

# with pipe
example_data %>%
  filter(age > 18) %>%
  select(group, dependent_variable)
```

Example no. 3

This works for almost all functions. Note: `na.rm = TRUE` excludes missing values when calculating the mean.

```
# without pipe
mean(age, na.rm = TRUE)

# with pipe
age %>% mean(na.rm = TRUE)
```

Example no. 4

```
# without pipe
lm(dependent_variable ~ group, data = example_data)

# with pipe
example_data %>% lm(dependent_variable ~ group, data = .)
```

Procedure

Note: Screenshot from Field, Miles & Field (2012). There are summaries like this for almost all statistical procedures covered in this seminar. They're invaluable for preparing for our exam and for life beyond this seminar.

16.6.2. General procedure for MANOVA ^①

To conduct factorial MANOVA you should follow this general procedure:

- 1 *Enter data.*
- 2 *Explore your data:* begin by graphing the data and computing descriptive statistics. You should check multivariate normality and take a look at the variance–covariance matrices for each group.
- 3 *Set contrasts for all predictor variables:* you need to decide what contrasts to do and to specify them appropriately for all of the independent variables in your analysis.
- 4 *Compute the MANOVA:* you can then run the main multivariate analysis of variance. Depending on what you found in the previous step, you might need to run a robust version of the test.
- 5 *Run univariate ANOVAs:* having conducted the MANOVA, you can follow it up with separate ANOVAs for each dependent variable.
- 6 *Discriminant function analysis:* better than the option above, consider running a discriminant function analysis.

1) Load data

1. Load the necessary packages and set a reasonable working directory.
2. Load the dataframe `ocd_data.dat` from the link https://pzezula.pages.gwdg.de/data/ocd_data.dat into R
3. Code the group variable as a factor with a reasonable baseline. Give the “No Treatment Control” group the level “NT”.

Solution

Subtask 1

```
library(tidyverse)
```

```
setwd("~/ownCloud/_Arbeit/Hiwi Peter/gitlab_sheets")
```

Subtask 2

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

```
## Parsed with column specification:  
## cols(  
##   group = col_character(),  
##   actions = col_double(),  
##   thoughts = col_double()  
## )
```

Subtask 3

```

ocd_data$group <- ocd_data$group %>%
  factor(levels = c("No Treatment Control", "BT", "CBT"),
        labels = c("NT", "BT", "CBT"))

```

2) Data overview

Meaning of the variables

Our data example contains hypothetical data of an evaluation study on different therapies for obsessive compulsive disorder

Variable	Meaning
group	Factor specifying the therapy method: NT = No Treatment, BT = behavioral therapy, CBT = cognitive behavioral therapy
actions	Frequency of obsessive actions after the therapy
thoughts	Frequency of obsessive thoughts after the therapy

1. First, create a simple scatterplot showing the relationship between obsessive actions on the x-axis and obsessive thoughts on the y-axis.
2. Add a regression line to the plot.
3. Use the command `facet_wrap()` to show the plot from 2.2 separately for each group.
4. Use the command `ocd_data %>% select(actions, thoughts) %>% by(ocd_data$group, cov)` to get separate variance-covariance-matrices for each group.
5. Use the command `ocd_data %>% by(ocd_data$group, psych::describe)` to get descriptive statistics for each group. Are you able to understand this command? **If you get an error message, you might have to install the package psych.**
6. Use the code below to investigate whether the data satisfy the assumption of being normally distributed. Can you understand the code? Is the data multivariately normally distributed for each group? **If you get an error message, you might have to install the package mvnrmtest.**
7. The Box's M-test checks for equal variance-covariance-matrices. Use the command `boxM()` of `library(heplots)` to check that. Do we violate this assumption? **If you get an error message, you might have to install the package heplots.**

```

# prepare data
nt <- ocd_data %>% filter(group == "NT") %>% select(2:3) %>% t()
bt <- ocd_data %>% filter(group == "BT") %>% select(2:3) %>% t()
cbt <- ocd_data %>% filter(group == "CBT") %>% select(2:3) %>% t()

# run the tests
mvnrmtest::mshapiro.test(nt)
mvnrmtest::mshapiro.test(bt)
mvnrmtest::mshapiro.test(cbt)

```

Solution

Subtask 1

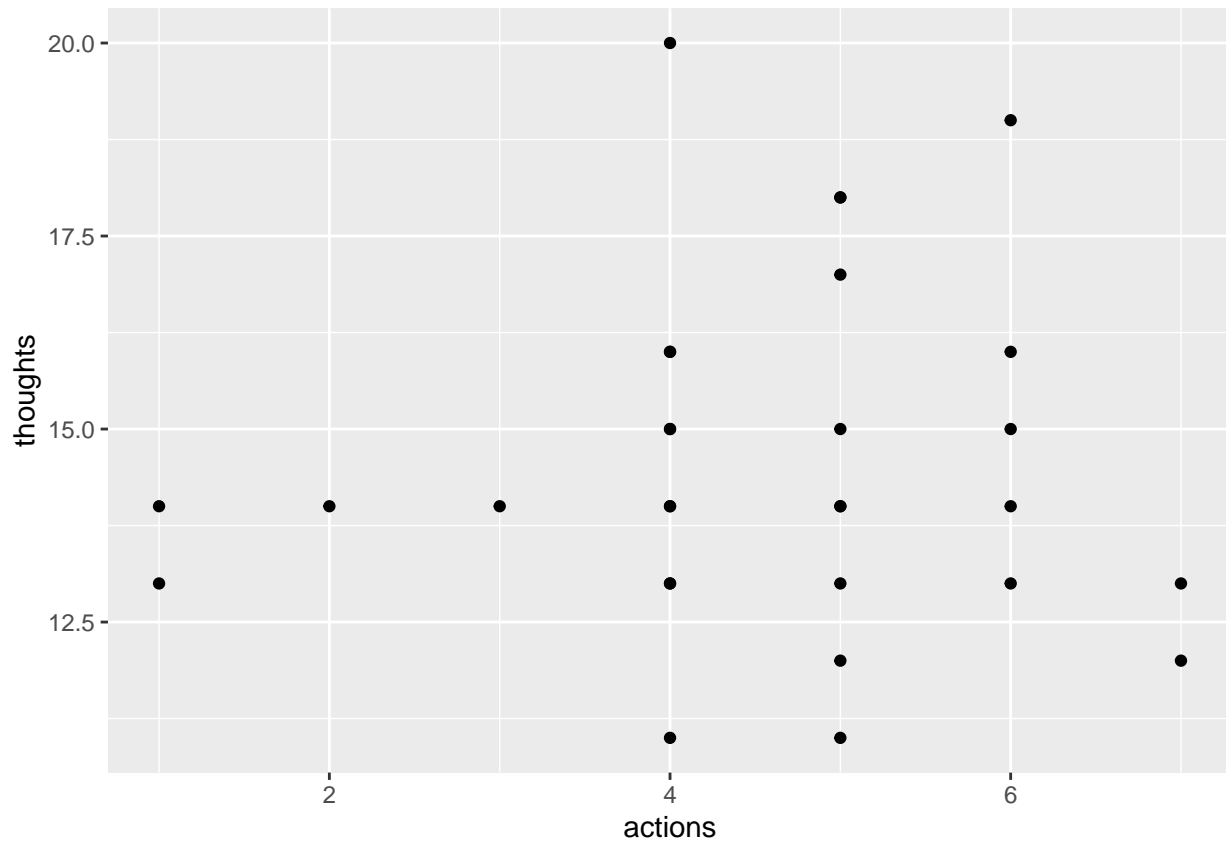
```

# Create plot object
baseplot <- ggplot(ocd_data, aes(x = actions, y = thoughts))

# Add point geom
scatterplot <- baseplot + geom_point()

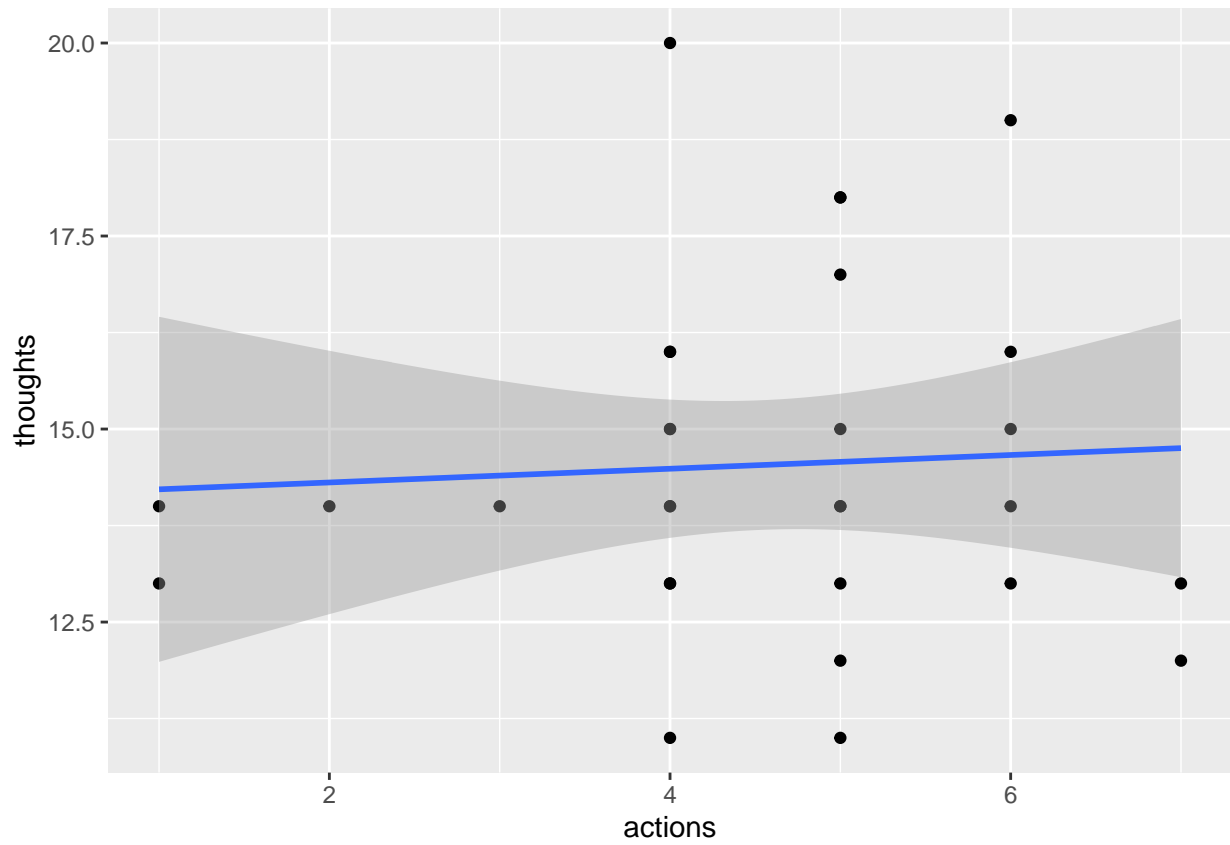
# Display plot
scatterplot

```



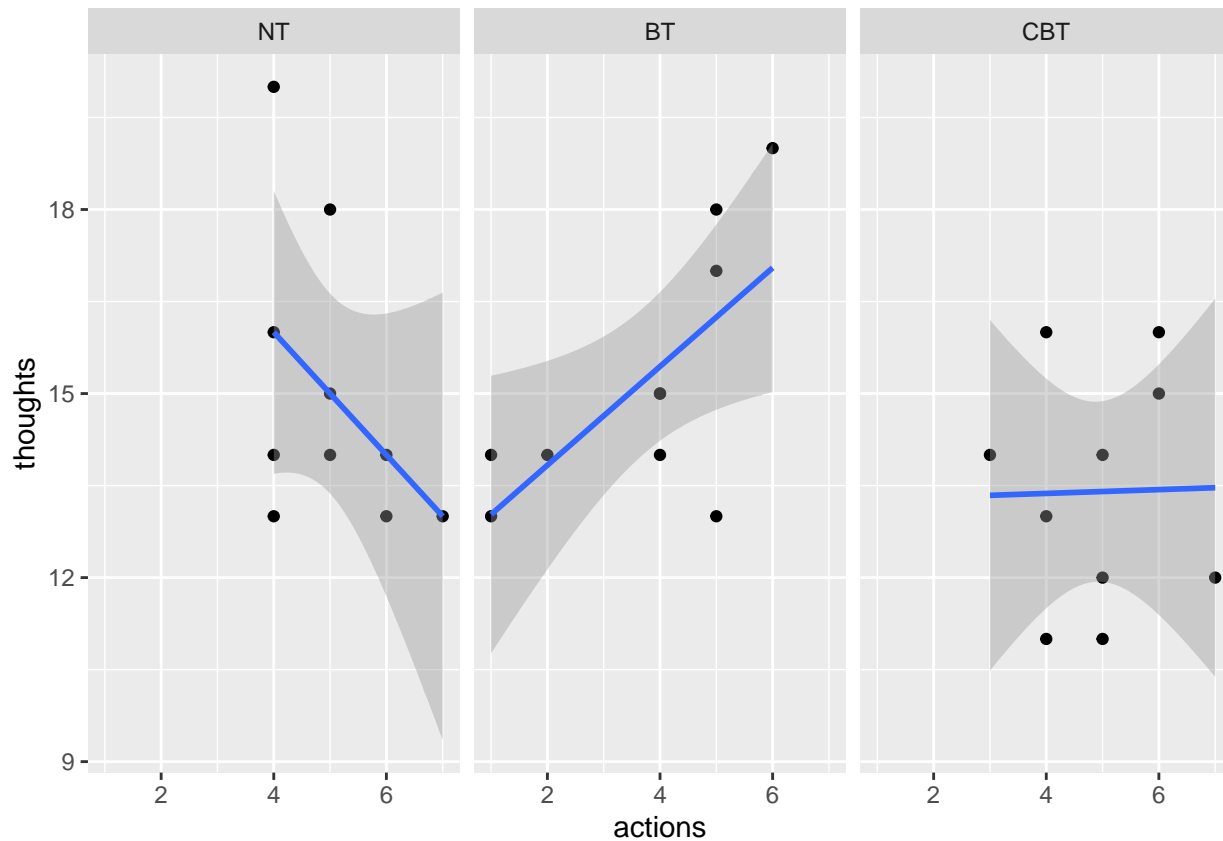
Subtask 2

```
# Add regression line  
lineplot <- scatterplot + geom_smooth(method = "lm")  
  
# Display plot  
lineplot  
  
## `geom_smooth()` using formula 'y ~ x'
```

Subtask 3

```
# Split plot  
groupplot <- lineplot + facet_wrap(~ group)  
  
# Display plot  
groupplot  
  
## `geom_smooth()` using formula 'y ~ x'
```



Subtask 4

```
oed_data %>% select(actions, thoughts) %>% by(oed_data$group, cov)
```

```
## oed_data$group: NT
##      actions thoughts
## actions  1.111111 -1.111111
## thoughts -1.111111  5.555556
## -----
## oed_data$group: BT
##      actions thoughts
## actions  3.122222  2.511111
## thoughts 2.511111  4.400000
## -----
## oed_data$group: CBT
##      actions thoughts
## actions  1.43333333 0.04444444
## thoughts 0.04444444 3.60000000
```

Subtask 5

```
oed_data %>% by(oed_data$group, psych::describe)
```

```
## oed_data$group: NT
##      vars  n mean  sd median trimmed mad min max range skew kurtosis  se
## group*   1 10   1 0.00   1   1.00 0.00  1  1    0 NaN      NaN 0.00
## actions  2 10   5 1.05   5   4.88 1.48  4  7    3 0.51  -1.22 0.33
## thoughts 3 10  15 2.36  14  14.62 1.48 13 20    7 0.96  -0.54 0.75
```

```
## -----
## ocd_data$group: BT
##      vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
## group*    1 10  2.0 0.00   2.0   2.00 0.00   2  2    0  NaN      NaN 0.00
## actions   2 10  3.7 1.77   4.0   3.75 1.48   1  6    5 -0.46  -1.45 0.56
## thoughts  3 10 15.2 2.10  14.5  15.00 1.48  13 19    6  0.61  -1.28 0.66
## -----
## ocd_data$group: CBT
##      vars  n mean  sd median trimmed  mad min max range skew kurtosis  se
## group*    1 10  3.0 0.0   3.0   3.00 0.00   3  3    0  NaN      NaN 0.00
## actions   2 10  4.9 1.2   5.0   4.88 1.48   3  7    4 0.17  -1.18 0.38
## thoughts  3 10 13.4 1.9  13.5  13.38 2.22  11 16    5 0.09  -1.67 0.60
```

Subtask 6

The `t()` command transposes the matrices containing the group data. Before, the information was coded from top to bottom (i.e., in columns). Now, the information's saved left to right (i.e., in rows). This is pretty unusual, but necessary for this test.

```
# Prepare data
nt <- ocd_data %>% filter(group == "NT") %>% select(2:3) %>% t()
bt <- ocd_data %>% filter(group == "BT") %>% select(2:3) %>% t()
cbt <- ocd_data %>% filter(group == "CBT") %>% select(2:3) %>% t()

# Run tests
mvnormttest::mshapiro.test(nt)
```

```
##
## Shapiro-Wilk normality test
##
## data: Z
## W = 0.82605, p-value = 0.02998
mvnormttest::mshapiro.test(bt)
```

```
##
## Shapiro-Wilk normality test
##
## data: Z
## W = 0.89122, p-value = 0.175
mvnormttest::mshapiro.test(cbt)
```

```
##
## Shapiro-Wilk normality test
##
## data: Z
## W = 0.9592, p-value = 0.7767
```

We get a significant result for the first group, indicating that the data in that group is not multivariately normally distributed.

Inspite of this, we continue the analysis for the sake of this sheet.

Subtask 7

```
res.boxm <- heplots::boxM(ocd_data[,c('actions', 'thoughts')], group=ocd_data$group)
res.boxm
```

```
##
```

```
## Box's M-test for Homogeneity of Covariance Matrices
##
## data:  ocd_data[, c("actions", "thoughts")]
## Chi-Sq (approx.) = 8.8932, df = 6, p-value = 0.1797
# summary(res.boxm) # for details
```

The p-value of our BoxM-test is not below 0.05, therefore we stay with H0 and state, that the variance-covariance-matrices don't differ significantly.

3) Run the MANOVA

First, some explanation

You can run a MANOVA in R using the command `manova()`. This works exactly like the `lm()` and `aov()` commands you already know, the pattern being `manova(outcome ~ predictor, data = data)`. The difference, however, is that you have to bind together all relevant outcome variables beforehand, using the `cbind()` command.

1. Set the contrasts for this analysis. This works the same way as for ANOVAs and regressions. If you're having trouble, consult your notes on earlier sheets covering these topics.
 - a) First contrast: comparison between BT and NT
 - b) Second contrast: comparison between CBT and NT *Note: These contrasts are non-orthogonal. Here, that's ok because we only have one predictor variable (cf. Field, ch. 16.6.6: Setting Contrasts) Note: We don't have to specify contrasts manually. Default contrasts would take BT as reference group and the difference to group CBT would be one of our effects.*
2. Create the necessary outcome object by using `cbind()` to join `ocd_data$thoughts` and `ocd_data$actions` together.
3. Use the `manova()` command to run the analysis. Save the result as an object.
4. Use the `summary()` command on the MANOVA object, including the additional argument `intercept = TRUE`.
5. What conclusions can you draw from the output?

Solution

Subtask 1

```
.bt.vs.nt <- c(-1,1,0)
.cbt.vs.nt <- c(-1,0,1)

# Attach contrasts to factor
contrasts(ocd_data$group) <- cbind(.bt.vs.nt, .cbt.vs.nt)

mean(ocd_data$actions)
```

```
## [1] 4.533333
```

```
ocd_data %>% group_by(group) %>% summarize(mean(actions))
```

```
## # A tibble: 3 x 2
##   group `mean(actions)`
##   <fct>      <dbl>
## 1 NT          5
## 2 BT         3.7
## 3 CBT        4.9
```

Subtask 2

```
outcome <- cbind(ocd_data$actions, ocd_data$thoughts)
```

Subtask 3

```
modell1 <- manova(outcome ~ group, data = ocd_data)
```

Subtask 4

```
summary(modell1, intercept = TRUE)
```

```
##              Df  Pillai approx F num Df den Df  Pr(>F)
## (Intercept)  1 0.98285   745.23     2    26 < 2e-16 ***
## group        2 0.31845     2.56     4    54 0.04904 *
## Residuals   27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Subtask 5

The F test for the groups is significant ($F(4,54) = 2.56$, $p = .049$). This means that the type of therapy received had an influence on the frequency of obsessive symptoms, measured through both thoughts and actions. From these results alone, this is the most detailed conclusions we're able to draw right now.

4) Interpreting the MANOVA

1. Use the command `summary.aov()` on the MANOVA model.
2. Briefly interpret the results.
 - a) Do these results justify analysing the planned contrasts?
3. Create a separate ANOVA model for each of the outcome variables and investigate the output in relation to your contrasts.
4. What do you conclude from these results?

Solution

Subtask 1

```
summary.aov(modell1)
```

```
## Response 1 :
##              Df Sum Sq Mean Sq F value Pr(>F)
## group        2  10.467   5.2333  2.7706 0.08046 .
## Residuals   27  51.000   1.8889
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Response 2 :
##              Df Sum Sq Mean Sq F value Pr(>F)
## group        2  19.467   9.7333  2.1541 0.1355
## Residuals   27 122.000   4.5185
```

Subtask 2

Neither of the separate ANOVAs yield significant results. This is an indication for neither obsessive thoughts nor obsessive actions being especially influenced by the kind of therapy received. That's highly interesting because we do have found an effect on the combination of these two components.

Based on this, we don't really have a good reason for investigating our planned contrasts. For the sake of this sheet, we'll do it anyway.

Subtask 3

```
out1 <- lm(actions ~ group, data = ocd_data)
out2 <- lm(thoughts ~ group, data = ocd_data)
```

```
summary(out1)
```

```
##
## Call:
## lm(formula = actions ~ group, data = ocd_data)
##
## Residuals:
##   Min     1Q   Median     3Q    Max
## -2.700 -0.975  0.100  1.075  2.300
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      4.5333     0.2509  18.067  <2e-16 ***
## group.bt.vs.nt  -0.8333     0.3549  -2.348   0.0264 *
## group.cbt.vs.nt  0.3667     0.3549   1.033   0.3106
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.374 on 27 degrees of freedom
## Multiple R-squared:  0.1703, Adjusted R-squared:  0.1088
## F-statistic: 2.771 on 2 and 27 DF,  p-value: 0.08046
```

```
summary(out2)
```

```
##
## Call:
## lm(formula = thoughts ~ group, data = ocd_data)
##
## Residuals:
##   Min     1Q   Median     3Q    Max
##  -2.40  -1.40  -0.70   1.45   5.00
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     14.5333     0.3881  37.448  <2e-16 ***
## group.bt.vs.nt   0.6667     0.5488   1.215   0.2350
## group.cbt.vs.nt -1.1333     0.5488  -2.065   0.0487 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.126 on 27 degrees of freedom
## Multiple R-squared:  0.1376, Adjusted R-squared:  0.07372
## F-statistic: 2.154 on 2 and 27 DF,  p-value: 0.1355
```

Subtask 4

For obsessive actions, the behavioral therapy works significantly better than the no treatment control. However, for obsessive thoughts, it's the cognitive behavioral therapy beating the control group.

5) Rendering (knit)

Render this file using `Ctrl + Shift + K` (Windows) or `Cmd + Shift + K` (Mac). In the viewer you should see a pretty versionn of your file. If this works: Congratulations! Your code can be rendered completely and without error codes! If it doesn't: No worries, you'll get there! Go hunting for errors in your code! Otherwise, we'll get it to render in the next seminar session!

Literature

Note: These sheets are based partially on exercises from the book *Discovering Statistics Using R* (Field, Miles & Field, 2012). They've been modified for the porpuses of this seminar, and the R code was updated.

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