



# Programmierung und Deskriptive Statistik

BSc Psychologie WiSe 2024/25

Belinda Fleischmann

Datum	Einheit	Thema	Form
15.10.24	R Grundlagen	(1) Einführung	Seminar
22.10.24	R Grundlagen	(2) R und Visual Studio Code	Seminar
29.10.24	R Grundlagen	(2) R und Visual Studio Code	Übung
05.11.24	R Grundlagen	(3) Vektoren, (4) Matrizen	Seminar
12.11.24	R Grundlagen	(5) Listen und Dataframes	Seminar
	<i>Leistungsnachweis 1</i>		
19.11.24	R Grundlagen	(6) Datenmanagement	Seminar
26.11.24	R Grundlagen	(2)-(6) R Grundlagen	Übung
03.12.24	Deskriptive Statistik	(7) Häufigkeitsverteilungen	Seminar
10.12.24	Deskriptive Statistik	(8) Verteilungsfunktionen und Quantile	Seminar
	<i>Leistungsnachweis 2</i>		
17.12.24	Deskriptive Statistik	(9) Maße der zentralen Tendenz und Datenvariabilität	Seminar
	Weihnachtspause		
07.01.25	R Grundlagen	(10) Strukturiertes Programmieren: Kontrollfluss, Debugging	Seminar
14.01.25	Deskriptive Statistik	(11) Anwendungsbeispiel	Übung
	<i>Leistungsnachweis 3</i>		
<b>21.01.25</b>	<b>Deskriptive Statistik</b>	<b>(11) Anwendungsbeispiel</b>	<b>Seminar</b>
28.01.25	Deskriptive Statistik	(11) Anwendungsbeispiel, Q&A	Seminar

## (11) Anwendungsbeispiel

Beispieldatensatz

Datenvorverarbeitung

Deskriptive Statistiken

Visualisierung

Parameterschätzung

Konfidenzintervalle

Hypothesentests

## **Beispieldatensatz**

Datenvorverarbeitung

Deskriptive Statistiken

Visualisierung

Parameterschätzung

Konfidenzintervalle

Hypothesentests

## Evidenzbasierte Evaluation von Psychotherapieformen bei Depression

Welche Therapieform ist bei Depression wirksamer?

### Online Psychotherapie



### Klassische Psychotherapie



## Beispiel: Evaluation von Psychotherapieformen bei Depression

Experimentelle Bedingung  
(Gruppen von  $n = 50$ )

Psychotherapie

Klassisch

Pre-BDI



Post-BDI

Online

Pre-BDI



Post-BDI

# Mittel der Datenerhebung

## Becks Depressions-Inventar (BDI) zur Depressionsdiagnostik

BDI-II Fragebogen			
Name	Alter	Bemerkungen	Status
<p><b>Anleitung:</b> Dieser Fragebogen enthält 13 Gruppen von Aussagen. Bitte lesen Sie jede dieser Gruppen von Aussagen sorgfältig durch und wählen Sie sich dann in jeder Gruppe eine Aussage heraus, die am besten beschreibt, wie Sie sich in den letzten zwei Wochen, einschließlich heute, gefühlt haben. Kreuzen Sie die Zahl neben der Aussage an, die Sie sich am besten beschreiben (0, 1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichwohl auf Sie zutreffen, kreuzen Sie die Aussage mit der höchsten Zahl an. Achten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, dies gilt auch für Gruppe 1b (Veränderungen der Schlafgewohnheiten) oder Gruppe 1d (Veränderungen des Appetits).</p>			
<p><b>1.) Traurigkeit</b></p> <p>0 Ich bin nicht traurig. 1 Ich bin oft traurig. 2 Ich bin ständig traurig. 3 Ich bin so traurig oder unglücklich, dass ich es nicht aushalte.</p>			
<p><b>2.) Pessimismus</b></p> <p>0 Ich sehe nicht mutlos in die Zukunft. 1 Ich sehe mutlos in die Zukunft als sonst. 2 Ich bin müde und erwarte nicht, dass meine Situation besser wird. 3 Ich glaube, dass meine Zukunft hoffnungslos ist und nur noch schlechter wird.</p>			
<p><b>3.) Versagensgefühle</b></p> <p>0 Ich fühle mich nicht als Versager. 1 Ich habe häufiger Versagensgefühle. 2 Wenn ich zurückblicke, sehe ich eine Menge Fehlschläge. 3 Ich habe das Gefühl, als Mensch ein richtiges Versagen zu sein.</p>			
<p><b>4.) Verlust von Freude</b></p> <p>0 Ich kann die Dinge genauso gut genießen wie früher. 1 Ich kann die Dinge nicht mehr so genießen wie früher. 2 Dinge, die mir früher Freude gemacht haben, kann ich kaum mehr genießen. 3 Dinge, die mir früher Freude gemacht haben, kann ich überhaupt nicht mehr genießen.</p>			
<p><b>5.) Schuldgefühle</b></p> <p>0 Ich habe keine besonderen Schuldgefühle. 1 Ich habe ein Schuldgefühl wegen Dingen, die ich getan habe oder hätte tun sollen. 2 Ich habe die meiste Zeit Schuldgefühle. 3 Ich habe ständig Schuldgefühle.</p>			
<p><b>6.) Bestrafungsgefühle</b></p> <p>0 Ich habe nicht das Gefühl, für etwas bestraft zu sein. 1 Ich habe das Gefühl, vielleicht bestraft zu werden. 2 Ich erwarte, bestraft zu werden. 3 Ich habe das Gefühl, bestraft zu sein.</p>			
<p><b>7.) Selbstablehnung</b></p> <p>0 Ich habe von mir genauso viel wie immer. 1 Ich habe Vertrauen in mich verloren. 2 Ich bin von mir enttäuscht. 3 Ich lehne mich völlig ab.</p>			
<p><b>8.) Selbstvorwürfe</b></p> <p>0 Ich kritisiere oder tadle mich nicht mehr als sonst. 1 Ich bin mir gegenüber kritischer als sonst. 2 Ich kritisiere mich für all meine Mängel. 3 Ich gebe mir die Schuld für alles Schlechte, was passiert.</p>			
<p><b>9.) Selbstmordgedanken</b></p> <p>0 Ich denke nicht daran, mir etwas anzutun. 1 Ich denke manchmal an Selbstmord, aber ich würde es nicht tun. 2 Ich würde mich vielleicht unterbringen. 3 Ich würde mich umbringen, wenn ich die Gelegenheit dazu hätte.</p>			
<p><b>10.) Werten</b></p> <p>0 Ich werte nicht oft ab, ob ich lebe. 1 Ich werte jetzt mehr ab, ob ich lebe. 2 Ich werte dem geringsten Anreiz. 3 Ich möchte gern weinens, aber ich kann nicht.</p>			

PEARSON  
© 2008 Pearson Assessment & Information GmbH, Frankfurt

<p><b>11.) Unruhe</b></p> <p>0 Ich bin nicht unruhiger als sonst. 1 Ich bin unruhiger als sonst. 2 Ich bin so unruhig, dass es mir schwerfällt, still zu sitzen. 3 Ich bin so unruhig, dass ich mich ständig bewege oder etwas tun muss.</p>	<p><b>12.) Interessenverlust</b></p> <p>0 Ich habe das Interesse an anderen Menschen oder an Tätigkeiten nicht verloren. 1 Ich habe weniger Interesse an anderen Menschen oder an Dingen als sonst. 2 Ich habe das Interesse an anderen Menschen oder Dingen zum größten Teil verloren. 3 Es fällt mir schwer, mich überhaupt für irgend etwas zu interessieren.</p>	<p><b>13.) Entscheidungsfähigkeit</b></p> <p>0 Ich bin so entscheidungsfähig wie immer. 1 Es fällt mir schwerer als sonst, Entscheidungen zu treffen. 2 Es fällt mir sehr viel schwerer als sonst, Entscheidungen zu treffen. 3 Ich habe Mühe, überhaupt Entscheidungen zu treffen.</p>	<p><b>17.) Reizbarkeit</b></p> <p>0 Ich bin nicht reizbarer als sonst. 1 Ich bin reizbarer als sonst. 2 Ich bin viel reizbarer als sonst. 3 Ich fühle mich dauernd gereizt.</p>
<p><b>14.) Wertlosigkeit</b></p> <p>0 Ich fühle mich nicht wertlos. 1 Ich habe mich für weniger wertvoll und nützlich als sonst. 2 Vergleichen mit anderen Menschen fühle ich mich viel weniger wert. 3 Ich fühle mich völlig wertlos.</p>	<p><b>18.) Veränderung des Appetits</b></p> <p>0 Mein Appetit hat sich nicht verändert. 1a Mein Appetit ist etwas schlechter als sonst. 1b Mein Appetit ist etwas größer als sonst. 2a Mein Appetit ist viel schlechter als sonst. 2b Mein Appetit ist viel größer als sonst. 3a Ich habe überhaupt keinen Appetit. 3b Ich habe ständig Heißhunger.</p>	<p><b>19.) Konzentrationschwierigkeiten</b></p> <p>0 Ich kann mich so gut konzentrieren wie immer. 1 Ich kann mich nicht mehr so gut konzentrieren wie sonst. 2 Es fällt mir schwer, mich längere Zeit auf irgend etwas zu konzentrieren. 3 Ich kann mich überhaupt nicht mehr konzentrieren.</p>	<p><b>20.) Ermüdung oder Erschöpfung</b></p> <p>0 Ich fühle mich nicht müder oder erschöpfter als sonst. 1 Ich werde schneller müde oder erschöpft als sonst. 2 Für viele Dinge, die ich üblicherweise tue, bin ich zu müde oder erschöpft. 3 Ich bin zu müde oder erschöpft, dass ich fast nichts mehr tun kann.</p>
<p><b>15.) Energielverlust</b></p> <p>0 Ich habe so viel Energie wie immer. 1 Ich habe weniger Energie als sonst. 2 Ich habe so wenig Energie, dass ich kaum noch etwas schaffe. 3 Ich habe keine Energie mehr, um überhaupt noch etwas zu tun.</p>	<p><b>21.) Verlust an sexuellem Interesse</b></p> <p>0 Mein Interesse an Sexualität hat sich in letzter Zeit nicht verändert. 1 Ich interessiere mich weniger für Sexualität als früher. 2 Ich interessiere mich jetzt viel weniger für Sexualität. 3 Ich habe das Interesse an Sexualität völlig verloren.</p>	<p><b>22.) Veränderungen der Schlafgewohnheiten</b></p> <p>0 Meine Schlafgewohnheiten haben sich nicht verändert. 1a Ich schlafe etwas weniger als sonst. 1b Ich schlafe viel mehr als sonst. 2a Ich schlafe viel weniger als sonst. 2b Ich schlafe fast den ganzen Tag. 3a Ich wache 1-2 Stunden früher auf als gewöhnlich und kann dann nicht mehr einschlafen.</p>	<p><b>Summe Skala 1:</b></p> <p><b>Übersicht Skala 1:</b></p> <p><b>Summe Skala 1 (24):</b></p>

0 - 8 keine Depression

9 - 13 minimale Depression

14 - 19 leichte Depression

20 - 28 mittelschwere Depression

29 - 63 schwere Depression



# Beispieldatensatz

## Einlesen des Datensatzes mit read.table()

```
file_path <- file.path(data_dir_path, "psychotherapie_datensatz.csv")

# file_path könnte beispielsweise so aussehen:
# "/home/username/uni/progr-und-deskr-stat-24/Daten/psychotherapie_datensatz.csv"

D <- read.table(file_path, sep = ",", header = TRUE)
```

## Daten der ersten acht Proband:innen jeder Gruppe

	Bedingung	Pre.BDI	Post.BDI
1	Klassisch	17	9
2	Klassisch	20	14
3	Klassisch	16	13
4	Klassisch	18	12
5	Klassisch	21	12
6	Klassisch	17	14
7	Klassisch	17	12
8	Klassisch	17	9
51	Online	22	16
52	Online	19	15
53	Online	21	13
54	Online	18	15
55	Online	19	13
56	Online	17	16
57	Online	20	13
58	Online	19	16

## Bonus: Datensimulation

```
# Seed setzen
set.seed(5)                                     # Startwert für den Zufallsgenerator setzen

# Simulationsparameter
n      <- 50                                     # Proband:innen pro Gruppe
mu     <- c(                                     # Erwartungswertparameter
  18, 12,                                       # Pre und Post der Gruppe Klassisch
  19, 14)                                     # Pre und Post der Gruppe Online
sigsqr <- 3                                     # Varianzparameter (gleich für alle Gruppen)

# Datensimulation
D <- data.frame(
  "Bedingung" = c(
    rep("Klassisch", n), rep("Online", n)), # n-mal "Klassisch", n-mal "Online"
  "Pre BDI" = c(
    round(rnorm(n, mu[1], sqrt(sigsqr))), # n Zufallswerte aus Normalverteilung mit mu[1]
    round(rnorm(n, mu[3], sqrt(sigsqr))), # n Zufallswerte aus Normalverteilung mit mu[3]
  "Post BDI" = c(
    round(rnorm(n, mu[2], sqrt(sigsqr))), # n Zufallswerte aus Normalverteilung mit mu[2]
    round(rnorm(n, mu[4], sqrt(sigsqr)))  # n Zufallswerte aus Normalverteilung mit mu[4]
  )
)

# Datenspeicherung
fname <- file.path(data_path, "psychotherapie_datensatz.csv")
write.csv(D, file = fname)
```

Beispieldatensatz

## **Datenvorverarbeitung**

Deskriptive Statistiken

Visualisierung

Parameterschätzung

Konfidenzintervalle

Hypothesentests

## Überlegungen für die Datenvorverarbeitung

- Studienfokus ist die **Veränderung** der Depressionsymptomatik durch Therapieformen.
- Für jede Proband:in ergibt sich diese Veränderung als **Differenz** zwischen Post.BDI und Pre.BDI.
- Eine Reduktion der Depressionssymptomatik ergibt dabei einen **negativen Wert**.
- Es ist intuitiver, Verbesserungen mit **positiven Zahlen** zu repräsentieren.
- Als Quantifizierung des Therapieeffekts bei Proband:in  $i$  bietet sich also folgendes Maß an

$$\Delta\text{BDI}[i] := -(\text{Post.BDI}[i] - \text{Pre.BDI}[i]) \quad (1)$$

- Wir betrachten in der Folge also das  $\Delta\text{BDI}$  Maß mit folgenden Interpretationen

$\Delta\text{BDI} > 0$	Verminderung der Depressionsymptomatik	Wirksame Therapie
$\Delta\text{BDI} = 0$	Keine Veränderung der Depressionsymptomatik	Wirkungslose Therapie
$\Delta\text{BDI} < 0$	Verstärkung der Depressionsymptomatik	Schädigende Therapie

## Hinzufügen einer $\Delta$ BDI Spalte zum Dataframe

```
fname      <- file.path(data_path, "psychotherapie_datensatz.csv")    # Einlesen
D          <- read.table(fname, sep = ",", header = TRUE)            # Rohdaten
D$Delta.BDI <- -(D$Post.BDI - D$Pre.BDI)                             # \Delta BDI Maß
```

## Daten der ersten acht Proband:innen jeder Gruppe

	Bedingung	Pre.BDI	Post.BDI	Delta.BDI
1	Klassisch	17	9	8
2	Klassisch	20	14	6
3	Klassisch	16	13	3
4	Klassisch	18	12	6
5	Klassisch	21	12	9
6	Klassisch	17	14	3
7	Klassisch	17	12	5
8	Klassisch	17	9	8
51	Online	22	16	6
52	Online	19	15	4
53	Online	21	13	8
54	Online	18	15	3
55	Online	19	13	6
56	Online	17	16	1
57	Online	20	13	7
58	Online	19	16	3

Beispieldatensatz

Datenvorverarbeitung

**Deskriptive Statistiken**

Visualisierung

Parameterschätzung

Konfidenzintervalle

Hypothesentests

## Bedingungsunabhängige Auswertung

```
# Initialisierung eines Dataframes
data      <- D$Delta.BDI
descr_stat <- data.frame(
  n      = length(data),
  Max     = max(data),
  Min     = min(data),
  Median  = median(data),
  Mean    = mean(data),
  Var     = var(data),
  Std     = sd(data)
)

print(descr_stat)

# Datenvektor aus Dataframe kopieren
# Dataframeerzeugung
# Stichprobengröße
# Maximum
# Minimum
# Mediane
# Mittelwert
# Varianz
# Standardabweichung

# Ausgabe
```

	n	Max	Min	Median	Mean	Var	Std
1	100	12	-1	6	5.54	5.826667	2.413849

## Bedingungsabhängige Auswertung

```
# Initialisierung eines Dataframes
th_bed      <- c("Klassisch", "Online")
n_th_bed    <- length(th_bed)
deskr_stat  <- data.frame(
  n      = rep(NaN, n_th_bed),
  Max    = rep(NaN, n_th_bed),
  Min    = rep(NaN, n_th_bed),
  Median = rep(NaN, n_th_bed),
  Mean   = rep(NaN, n_th_bed),
  Var    = rep(NaN, n_th_bed),
  Std    = rep(NaN, n_th_bed),
  row.names = th_bed
)

# Therapiebedingungen
# Anzahl Therapiebedingungen
# Dataframeerzeugung
# Stichprobengrößen
# Maxima
# Minima
# Mediane
# Mittelwerte
# Varianzen
# Standardabweichungen
# Zeilenbenennung

# Iterationen über Therapiebedingungen
for (i in seq_along(th_bed)){
  data      <- D$Delta.BDI[D$Bedingung == th_bed[i]] # Daten filtern
  deskr_stat$n[i]      <- length(data)                # Stichprobengröße
  deskr_stat$Max[i]    <- max(data)                   # Maxima
  deskr_stat$Min[i]    <- min(data)                   # Minima
  deskr_stat$Median[i] <- median(data)                 # Mediane
  deskr_stat$Mean[i]   <- mean(data)                  # Mittelwerte
  deskr_stat$Var[i]    <- var(data)                   # Varianzen
  deskr_stat$Std[i]    <- sd(data)                   # Standardabweichungen
}
```



## Bedingungsabhängige Auswertung

```
# Ausgabe  
print(deskr_stat)
```

	n	Max	Min	Median	Mean	Var	Std
Klassisch	50	12	-1	6	6.16	7.075918	2.660060
Online	50	9	1	5	4.92	3.911837	1.977836

- Die Anzahl der Proband:innen in beiden Therapiegruppen ist gleich.
- Die Spannbreite der  $\Delta$ BDI Daten ist in der klassischen Therapieform leicht erhöht.
- Median und Mittelwert nehmen für die klassische Therapieform leicht höhere Werte an.
- Ein  $\Delta$ BDI Mittelwertsunterschied von 1 ist klinisch wohl eher vernachlässigbar.
- Median und Mittelwert sind in beiden Therapieformen ähnlich (unimodale Verteilung).
- Die Variabilitätsmaße zeigen eine etwas erhöhte Variabilität in der klassischen Therapieform.

Beispieldatensatz

Datenvorverarbeitung

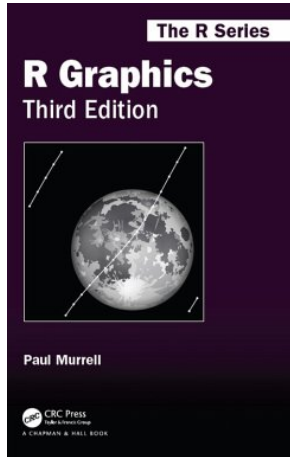
Deskriptive Statistiken

**Visualisierung**

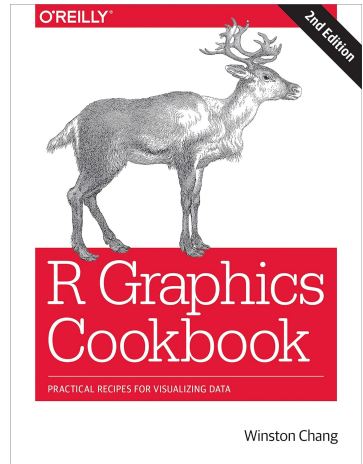
Parameterschätzung

Konfidenzintervalle

Hypothesentests



Murrell (2019)



Online-Buch, Chang (2013)

## R Funktionalitäten für Abbildungen

### Base Graphics

- Erstellung und bedarfsgerechte Anpassung von Abbildungen
- Eher low-level, fine tuning orientiert

### Lattice und ggplot2

- Erstellung und bedarfsgerechte Anpassung von Abbildungen
- Eher high level, an der eigenen Philosophie orientiert

Base Graphics, lattice und ggplot2 können ähnliche Abbildungen generieren

LaTeX Typesetting ist in allen Paketen unterentwickelt

## R Funktionalitäten für Abbildungen

### Base Graphics

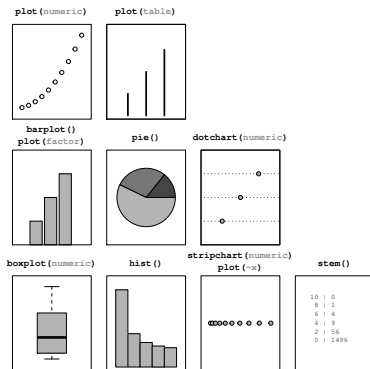
- **Erstellung und bedarfsgerechte Anpassung von Abbildungen**
- **Eher low-level, fine tuning orientiert**

### Lattice und ggplot2

- Erstellung und bedarfsgerechte Anpassung von Abbildungen
- Eher high level, an der eigenen Philosophie orientiert

Base Graphics, lattice und ggplot2 können ähnliche Abbildungen generieren

LaTeX Typesetting ist in allen Paketen unterentwickelt



**Figure 2.5**

High-level base graphics plotting functions for producing plots of a single variable. Where the function can be used to produce more than one type of plot, the relevant data type is shown (in gray). For example, `plot(numeric)` means that this is what the `plot()` produces when it is given a single numeric argument.

Murrell (2019)

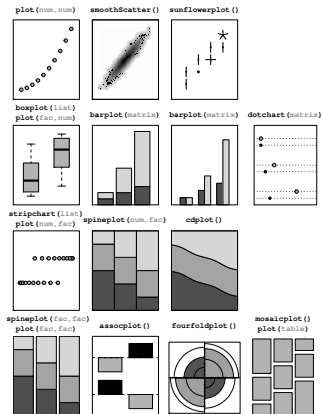
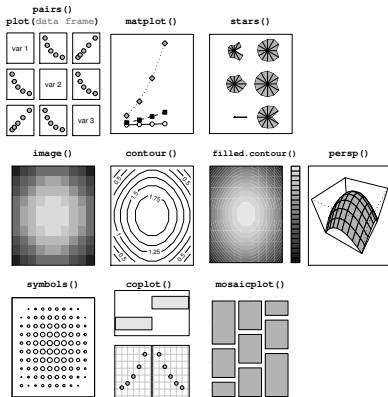


Figure 2.6

High-level base graphics plotting functions for producing plots of two variables. Where the function can be used to produce more than one type of plot, the relevant data type is shown (in gray). For example `plot(num, fac)` represents calling the `plot()` function with a numeric vector as the first argument and a factor as the second argument.

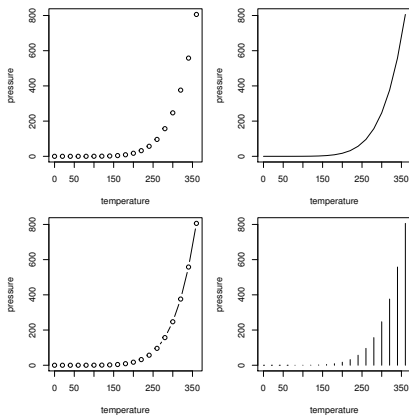


**Figure 2.7**

High-level base graphics plotting functions for producing plots of many variables. Where the function can be used to produce more than one type of plot, the relevant data type is shown (in gray).

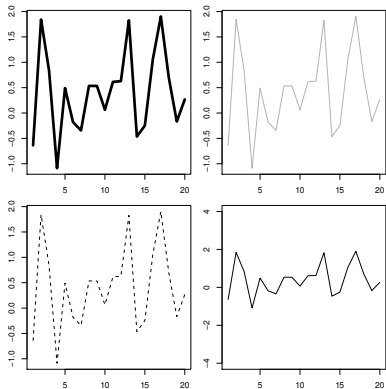
Murrell (2019)





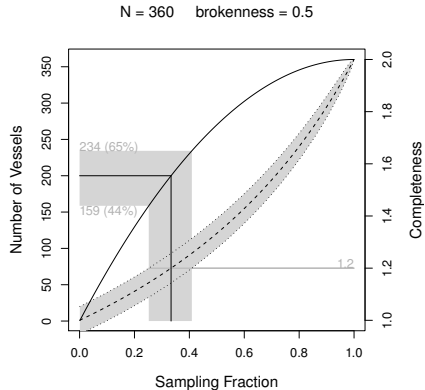
**Figure 2.2**

Four variations on a scatterplot. In each case, the plot is produced by a call to the `plot()` function with the same data; all that changes is the value of the `type` argument. At top-left, `type="p"` to give points (data symbols), at top-right, `type="l"` to give lines, at bottom-left, `type="b"` to give both, and at bottom-right, `type="h"` to give histogram-like vertical lines.



**Figure 2.9**

Standard arguments for high-level functions. All four plots are produced by calls to the `plot()` function with the same data, but with different standard plot function arguments specified: the top-left plot makes use of the `lwd` argument to control line thickness; the top-right plot uses the `col` argument to control line color; the bottom-left plot makes use of the `lty` argument to control line type; and the bottom-right plot uses the `ylim` argument to control the scale on the y-axis.



**Figure 1.3**

A customized scatterplot produced using R. This is created by starting with a simple scatterplot and augmenting it by adding an additional y-axis and several additional sets of lines, polygons, and text labels.

Murrell (2019)

## Code Outline

```
# Initialisierung einer neuen Abbildung
dev.new()

# Abbildungsparameter
par(
  z.B. Arrangement von Panels, Begrenzungsstile, Schriftfonts, etc
)

# Higher-level Abbildungsfunktion wie plot(), hist(), barplot(), ...
plot(
  z.B. x- und y-Daten, Achsenlimits, Achsenbeschriftungen, Titel, Farben, etc.
  Jeder Aufruf einer higher-level Graphikfunktion belegt ein neues Subpanel!
)

# Hinzufügen weiterer Daten mit lower-level Abbildungsfunktionen zum aktuellen Panel
z.B. points(), lines(), abline()

# Weitere Graphikannotation zu aktuellem Panel
z.B. legend(), text()

# Speichern der Abbildung (Größenverhältnisse erst hier final festgelegt)
z.B. dev.copy2pdf()
```

# Visualisierung

## Visualisierung der bedingungsabhängigen Histogramme

```
# Histogrammparameter
h      <- 1
b_0    <- min(D$Delta.BDI)
b_k    <- max(D$Delta.BDI)
k      <- ceiling((b_k - b_0) / h)
b      <- seq(b_0, b_k, by = h)
ylimits <- c(0, .25)
xlimits <- c(-2, 14)
th_bed <- c("Klassisch", "Online")
labs   <- c("Klassische Therapie", "Online Therapie")

# Abbildungsparameter
par(
  mfcol      = c(1, 2),
  family     = "sans",
  pty        = "m",
  bty        = "n",
  las        = 1,
  xaxs       = "i",
  yaxs       = "i",
  font.main  = 1,
  cex        = 1,
  cex.main   = 1
)

# Iteration über Therapiebedingungen
for(i in seq_along(th_bed)){
  hist(
    D$Delta.BDI[D$Bedingung == th_bed[i]],
    breaks = b,
    freq   = F,
    xlim   = xlimits,
    ylim   = ylimits,
    xlab    = TeX("$\\Delta$ BDI"),
    ylab    = "",
    main    = labs[i]
  )
}

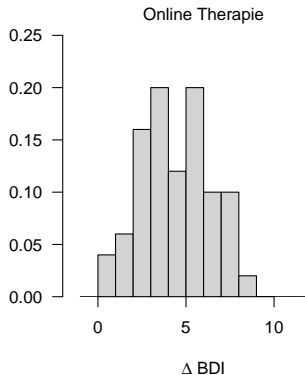
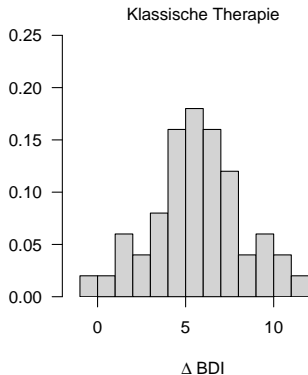
# PDF Speicherung
dev.copy2pdf(
  file      = file.path(abb_dir, "pds_11_histogramm.pdf"),
  width     = 8,
  height    = 4
)
```

```
# gewünschte Klassenbreite
# b_0
# b_k
# Anzahl der Klassen
# Klassen [b_{j-1}, b_j[
# y-Achsenlimits
# x-Achsenlimits
# Therapiebedingungen
# Abbildungslabel

# für Details siehe ?par
# 1 x 2 Panelstruktur
# Serif-freier Fonttyp
# Maximale Abbildungsregion
# L förmige Box
# Horizontale Achsenbeschriftung
# x-Achse bei y = 0
# y-Achse bei x = 0
# Non-Bold Titel
# Textvergrößerungsfaktor
# Titeltextrvergrößerungsfaktor

# Delta.BDI Werte von Therapiebedingung i
# Histogrammklassen
# normierte relative Häufigkeit
# x-Achsenlimits
# y-Achsenlimits
# x-Achsenbeschriftung
# y-Achsenbeschriftung
# Titelbeschriftung
```

## Visualisierung der bedingungsabhängigen Histogramme



## Visualisierung bedingungsabhängiger deskriptiver Statistiken

```
# Abbildungsparameter
par(
  mfcol      = c(1,2),
  family     = "sans",
  pty       = "m",
  bty       = "l",
  las       = 1,
  xaxs      = "i",
  yaxs      = "i",
  font.main  = 1,
  cex       = 1,
  cex.main   = 1.5
)

# für Details siehe ?par
# 1 x 2 Panelstruktur
# Serif-freier Fonttyp
# Maximale Abbildungsregion
# L-förmige Box
# Horizontale Achsenbeschriftung
# x-Achse bei y = 0
# y-Achse bei x = 0
# Non-Bold Titel
# Textvergrößerungsfaktor
# Titeltextrvergrößerungsfaktor
```

## Visualisierung bedingungsabhängiger deskriptiver Statistiken

```
# Linkes Panel: Balkendiagramm mit Fehlerbalken
# -----

# Stichprobenmittelwert und Standardabweichung extrahieren
mw      <- desk_r_stat$Mean      # Gruppenmittelwert
sd      <- desk_r_stat$Std       # Gruppenstandardabweichung
names(mw) <- th_bed              # barplot braucht x-Werte als names

# Mit der Funktion barplot() ein Balkendiagramm plotten
x <- barplot(
  height = mw,                  # Speichern der der x-Ordinaten (?barplot für Details)
  col    = "gray90",           # Mittelwerte als Balkenhöhe
  ylim   = c(0,12),            # Balkenfarbe
  xlim   = c(0,3),             # y-Achsenbegrenzung
  xlab    = "Bedingung",        # x-Achsenbegrenzung
  main    = TeX("$\\Delta$ BDI$") # x-Achsenbeschriftung
)                                # Titel

# Mit der Funktion arrows() Fehlerbalken zeichnen
arrows(
  x0      = x,                  # arrow start x-ordinate
  y0      = mw - sd,            # arrow start y-ordinate
  x1      = x,                  # arrow end x-ordinate
  y1      = mw + sd,            # arrow end y-ordinate
  code    = 3,                  # Pfeilspitzen beiderseits
  angle   = 90,                 # Pfeilspitzenwinkel -> Linie
  length  = 0.05                # Linielänge
)
```

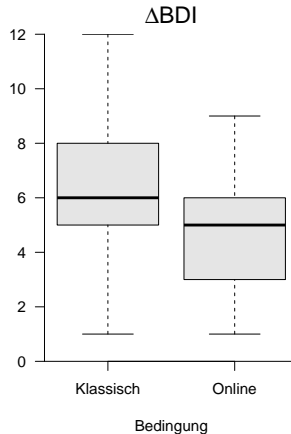
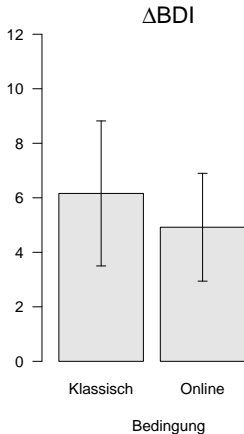


## Visualisierung bedingungsabhängiger deskriptiver Statistiken

```
# Rechtes Panel: Boxplot
# -----

# Mit der Funktion boxplot() boxplots zeichnen
boxplot(
  D$Delta.BDI ~ D$Bedingung,          # Gruppierung der Delta.BDI Daten nach D$Bedingung mit "-"
  ylim = c(0, 12),                  # y-Achsenbegrenzung
  col = "gray90",                   # Boxfarbe
  ylab = "",                         # y-Achsenbeschriftung
  xlab = "Bedingung",               # x-Achsenbeschriftung
  main = TeX("$\\Delta$ BDI$")      # Titel
)
```

## Visualisierung bedingungsabhängiger deskriptiver Statistiken



Beispieldatensatz

Datenvorverarbeitung

Deskriptive Statistiken

Visualisierung

**Parameterschätzung**

Konfidenzintervalle

Hypothesentests

## Modellannahmen für Parameterschätzung und Konfidenzintervalle

Motiviert durch die therapieabhängige Visualisierung der  $\Delta$ BDI-Daten und unseren wissenschaftssoziologischen Kontext legen wir das folgende Normalverteilungsmodell zugrunde:

Für die  $\Delta$ BDI-Reduktion  $y_{ij}$  der  $j$ ten Proband:in in der  $i$ ten Therapiebedingung nehmen wir an:

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma_i^2), \quad i = 1, 2, j = 1, \dots, 50. \quad (2)$$

Die Reduktion  $y_{ij}$  wird also durch eine innerhalb der  $i$ ten Therapiebedingung identische mittlere Reduktion  $\mu_i \in \mathbb{R}$  und eine Proband:innen-spezifische normalverteilte Abweichung  $\epsilon_{ij}$  erklärt.

Dieses Modell ist äquivalent zur Aussage:

$$y_{i1}, \dots, y_{i50} \sim N(\mu_i, \sigma_i^2). \quad (3)$$

Innerhalb jeder Therapiebedingung werden die  $\Delta$ BDI-Werte somit als unabhängig und identisch verteilte Zufallsvariablen modelliert.

Die Standardproblemstellungen der Frequentistischen Inferenz führen zu folgenden Fragen:

1. Was sind sinnvolle Schätzwerte für die wahren, aber unbekannten Parameter  $\mu_1, \mu_2$  und  $\sigma_1^2, \sigma_2^2$ ?
2. Wie gelingt im Sinne einer Intervallschätzung eine möglichst sichere Schätzung dieser Parameter?
3. Entscheiden wir uns sinnvollerweise für die Hypothese, dass gilt  $\mu_1 = \mu_2$ ?

# Parameterschätzung

Zur Parameterschätzung im vorliegenden Modell nutzen wir

- den Maximum Likelihood Schätzer für  $\mu_i$
- den Varianzschätzer für  $\sigma_i^2$

```
# Initialisierung eines Dataframes
th_bed      <- c("Klassisch", "Online")
n_th_bed    <- length(th_bed)
theta_hats  <- data.frame(
  mu_ML      = rep(NA, n_th_bed),
  sigsqr_VAR = rep(NA, n_th_bed)
)

# Iterationen über Therapiebedingungen
for(i in 1:n_th_bed){
  data      <- D$Delta.BDI[D$Bedingung == th_bed[i]] # Daten
  theta_hats$mu_ML[i] <- mean(data)                  # ML Schätzer für \mu_i
  theta_hats$sigsqr_VAR[i] <- var(data)               # Varianzschätzer für \sigma^2_i
}

# Ausgabe
print(theta_hats)
```

```
mu_ML sigsqr_VAR
1  6.16  7.075918
2  4.92  3.911837
```

Tipps für  $\mu_i$  und  $\sigma_i^2$  auf Grundlage dieser unverzerrten Schätzer sind also

$$\hat{\mu}_1 = 6.16, \quad \hat{\mu}_2 = 4.92, \quad \hat{\sigma}_1^2 = 7.08, \quad \hat{\sigma}_2^2 = 3.91. \quad (4)$$

Die mit diesen Tipps assoziierte Unsicherheit ist hier nicht angegeben.

Beispieldatensatz

Datenvorverarbeitung

Deskriptive Statistiken

Visualisierung

Parameterschätzung

**Konfidenzintervalle**

Hypothesentests

# Konfidenzintervalle

## Konfidenzintervalle für die Erwartungswertparameterschätzer

```
# Analyseparameter
th_bed  <- c("Klassisch", "Online")
n_th_bed <- length(th_bed)
n        <- 50

# Therapiebedingungen
# Anzahl an Therapiebedingungen
# Anzahl von Beobachtungen pro Therapiebedingung

kappa <- data.frame(
  G_u      = rep(NA, n_th_bed),
  mu_hat   = rep(NA, n_th_bed),
  G_o      = rep(NA, n_th_bed),
  row.names = th_bed
)

# Dataframeerzeugung
# untere KI Grenze
# Erwartungswertparameterschätzer
# obere KI Grenze
# Therapiebedingungen

# Konfidenzintervallparameter
delta  <- 0.95
t_delta <- qt((1 + delta) / 2, n - 1)

# Konfidenzintervallevaluation
for(i in 1:n_th_bed){
  data <- D$Delta.BDI[D$Bedingung == th_bed[i]]
  y_bar <- mean(data)
  S <- sd(data)
  kappa$G_u[i] <- y_bar - (S / sqrt(n)) * t_delta
  kappa$mu_hat[i] <- y_bar
  kappa$G_o[i] <- y_bar + (S / sqrt(n)) * t_delta
}

# Ausgabe
print(kappa)
```

	G_u	mu_hat	G_o
Klassisch	5.404019	6.16	6.915981
Online	4.357905	4.92	5.482095



# Konfidenzintervalle

## Konfidenzintervalle für die Varianzparameterschätzer

```
# Analyseparameter
th_bed  <- c("Klassisch", "Online")
n_th_bed <- length(th_bed)
n        <- 50

# Therapiebedingungen
# Anzahl an Therapiebedingungen
# Anzahl von Beobachtungen pro Therapiebedingung

kappa <- data.frame(
  G_u      = rep(NA, n_th_bed),
  sigsq_r_hat = rep(NA, n_th_bed),
  G_o      = rep(NA, n_th_bed),
  row.names = th_bed
)

# Dataframeerzeugung
# untere KI Grenze
# Varianzparameterschätzer
# obere KI Grenze
# Therapiebedingungen

# Konfidenzintervallparameter
delta <- 0.95
u_delta <- qchisq((1 - delta) / 2, n - 1)
u_delta_p <- qchisq((1 + delta) / 2, n - 1)

# Konfidenzintervallevaluation
for(i in 1:n_th_bed){
  data <- D$Delta.BDI[D$Bedingung == th_bed[i]]
  S2 = var(data)
  kappa$G_u[i] = (n - 1) * S2 / u_delta_p
  kappa$sigsqr_hat[i] = S2
  kappa$G_o[i] = (n - 1) * S2 / u_delta
}

# Ausgabe
print(kappa)
```

	G_u	sigsqr_hat	G_o
Klassisch	4.937455	7.075918	10.987828
Online	2.729613	3.911837	6.074489

# Programmierungsübungen und Selbstkontrollfragen

---

1. Führe alle in dieser Einheit gezeigten Datenanalysen selbstständig durch und halte sie in einem strukturierten R-Skript fest.
2. Simuliere einen Beispieldatensatz mit Daten einer Evaluation von 3 verschiedenen Psychotherapieformen bei Depression mit 100 Versuchspersonen pro Gruppe und zwei Messzeitpunkten (vor Intervention und nach Intervention).
3. Variiere die Parameter der Simulationen für zwei Szenarien, in denen jeweils in *nur einer* Gruppe im Mittel ein Unterschied zwischen Pre- und Post-BDI-Werten besteht.
4. Berechne die bedingungsabhängigen deskriptiven Statistiken und visualisiere diese.
5. Berechne die Parameterschätzer für Erwartungswert und Varianz und bestimme deren Konfidenzintervalle.
6. Erläutere die zugrundeliegenden Modellannahmen im Rahmen der Frequentistischen Inferenz.
7. Wählen Sie zwei Gruppen, und führen Sie einen Zweistichproben-T-Test durch, um die Hypothese testen, dass ein Unterschied in der BDI-Veränderung zwischen diesen Gruppen besteht. Wählen Sie dafür sinnvolle Null- und Alternativhypothesen. Welche Annahmen müssen Sie vor Durchführung des Tests treffen?
8. Erläutere die zugrundeliegenden Modellannahmen im Rahmen der Frequentistischen Inferenz.

Chang, Winston. 2013. *R Graphics Cookbook*. Beijing Cambridge Farnham Köln Sebastopol Tokyo: O'Reilly.  
Murrell, Paul. 2019. *R Graphics*. Third edition. The R Series. Boca Raton: CRC Press, Taylor & Francis Group.