



# Programmierung und Deskriptive Statistik

BSc Psychologie WiSe 2022/23

Belinda Fleischmann

Inhalte basieren auf Programmierung und Deskriptive Statistik von Dirk Ostwald, lizenziert unter CC BY-NC-SA 4.0

## (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



# Beispieldatensatz

## Evidenzbasierte Evaluation von Psychotherapieformen bei Depression

### Becks Depressions-Inventar (BDI) zur Depressionsdiagnostik

**BDI-II** Fragebogen

Name	Alter	Seitwertsch.	Prüfung	Datum

**Anleitung:** Dieser Fragebogen enthält 21 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 herausgerufen haben (0, 1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichwohl zutreffend für Sie zutreffen, kreuzen Sie die Aussage mit der höchsten Zahl an. Antworten Sie bitte darauf, dass Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, das gilt auch für Gruppe 16 (Veränderungen der Schlafgewohnheiten) oder Gruppe 16 (Veränderungen des Appetits).

**1.) Traurigkeit**

- Ich bin nicht traurig.
- Ich bin oft traurig.
- Ich bin ständig traurig.
- Ich bin so traurig oder unglücklich, dass ich es nicht aushalte.

**2.) Pessimismus**

- Ich sehe nicht mutlos in die Zukunft.
- Ich sehe mutlos in die Zukunft als sonst.
- Ich bin mutlos und erwarte nicht, dass meine Situation besser wird.
- Ich glaube, dass meine Zukunft hoffnungslos ist und nur noch schlechter wird.

**3.) Versagensgefühle**

- Ich fühle mich nicht als Versager.
- Ich habe häufiger Versagensgefühle.
- Wenn ich zurückblicke, sehe ich eine Menge Fehlentscheidungen.
- Ich habe das Gefühl, ich werde ein völliger Versager zu sein.

**4.) Verlust von Freude**

- Ich kann die Dinge genauso gut genießen wie früher.
- Ich kann die Dinge nicht mehr so genießen wie früher.
- Dinge, die mir früher Freude gemacht haben, kann ich kaum mehr genießen.
- Dinge, die mir früher Freude gemacht haben, kann ich überhaupt nicht mehr genießen.

**5.) Schuldgefühle**

- Ich habe keine besonderen Schuldgefühle.
- Ich habe oft Schuldgefühle wegen Dingen, die ich getan habe oder hätte tun sollen.
- Ich habe die meiste Zeit Schuldgefühle.
- Ich habe ständig Schuldgefühle.

**6.) Bestrafungsgefühle**

- Ich habe nicht das Gefühl, für etwas bestraft zu sein.
- Ich habe das Gefühl, vielleicht bestraft zu sein.
- Ich erwarte, bestraft zu werden.
- Ich habe das Gefühl, bestraft zu sein.

**7.) Selbstablehnung**

- Ich halte von mir genauso viel wie immer.
- Ich habe Vertrauen in mich verloren.
- Ich bin von mir enttäuscht.
- Ich lehne mich völlig ab.

**8.) Selbstvorwürfe**

- Ich kritisiere oder tadle mich nicht mehr als sonst.
- Ich bin mir gegenüber kritischer als sonst.
- Ich kritisieren mich für all meine Mängel.
- Ich gebe mir die Schuld für alles Schlechte, was passiert.

**9.) Selbstmordgedanken**

- Ich denke nicht daran, mir etwas anzutun.
- Ich denke manchmal an Selbstmord.
- Ich möchte mich umbringen, aber ich würde mich umbringen, wenn ich die Gelegenheit dazu hätte.
- Ich möchte mich umbringen, aber ich kann nicht.

**10.) Weinen**

- Ich weine nicht öfter als früher.
- Ich weine jetzt mehr als früher.
- Ich weine beim geringsten Anlass.
- Ich möchte gern weinen, aber ich kann nicht.

**11.) Unruhe**

- Ich bin nicht unruhiger als sonst.
- Ich bin unruhiger als sonst.
- Ich bin so unruhig, dass es mir schwerfällt, still zu sitzen.
- Ich bin so unruhig, dass ich mich ständig bewege oder etwas tun muss.

**12.) Interessenverlust**

- Ich habe das Interesse an anderen Menschen oder an Tätigkeiten nicht verloren.
- Ich habe weniger Interesse an anderen Menschen oder an Dingen als sonst.
- Ich habe das Interesse an anderen Menschen oder Dingen zum größten Teil verloren.
- Es fällt mir schwer, mich überhaupt für irgend etwas zu interessieren.

**13.) Entscheidungsfähigkeit**

- Ich bin so entscheidungsfröhlich wie immer.
- Es fällt mir schwerer als sonst, Entscheidungen zu treffen.
- Es fällt mir sehr viel schwerer als sonst, Entscheidungen zu treffen.
- Ich habe Mühe, überhaupt Entscheidungen zu treffen.

**14.) Wertlosigkeit**

- Ich fühle mich nicht wertlos.
- Ich habe mich für weniger wertvoll und nützlich als sonst.
- Versehen mit anderen Menschen fühle ich mich viel weniger wert.
- Ich fühle mich völlig wertlos.

**15.) Energieverlust**

- Ich habe so viel Energie wie immer.
- Ich habe weniger Energie als sonst.
- Ich habe so wenig Energie, dass ich kaum noch etwas schaffe.
- Ich habe keine Energie mehr, um überhaupt noch etwas zu tun.

**16.) Veränderungen der Schlafgewohnheiten**

- Meine Schlafgewohnheiten haben sich nicht verändert.
- Ich schlafe etwas mehr als sonst.
- Ich schlafe etwas weniger als sonst.
- Ich schlafe viel mehr als sonst.
- Ich schlafe viel weniger als sonst.
- Ich schlafe fast den ganzen Tag.
- Ich wache 1-2 Stunden früher auf als gewöhnlich und kann dann nicht mehr einschlafen.

**17.) Reizbarkeit**

- Ich bin nicht reizbarer als sonst.
- Ich bin reizbarer als sonst.
- Ich bin viel reizbarer als sonst.
- Ich fühle mich dauernd gereizt.

**18.) Veränderungen des Appetits**

- Mein Appetit hat sich nicht verändert.
- Mein Appetit ist etwas schlechter als sonst.
- Mein Appetit ist etwas größer als sonst.
- Mein Appetit ist viel schlechter als sonst.
- Mein Appetit ist viel größer als sonst.
- Ich habe überhaupt keinen Appetit.
- Ich habe ständig Heißhunger.

**19.) Konzentrations-schwierigkeiten**

- Ich kann mich so gut konzentrieren wie immer.
- Ich kann mich nicht mehr so gut konzentrieren wie sonst.
- Es fällt mir schwer, mich längere Zeit auf irgend etwas zu konzentrieren.
- Ich kann überhaupt nicht mehr konzentrieren.

**20.) Ermüdung oder Erschöpfung**

- Ich fühle mich nicht müde oder erschöpft als sonst.
- Ich werde schneller müde oder erschöpft als sonst.
- Für viele Dinge, die ich üblicherweise tue, bin ich so müde oder erschöpft, dass ich fast nichts mehr tun kann.
- Ich fühle mich müde oder erschöpft, dass ich fast nichts mehr tun kann.

**21.) Verlust an sexuellem Interesse**

- Mein Interesse an Sexualität hat sich in letzter Zeit nicht verändert.
- Ich interessiere mich weniger für Sexualität als früher.
- Ich interessiere mich jetzt viel weniger für Sexualität.
- Ich habe das Interesse an Sexualität völlig verloren.

0 - 8 keine Depression

9 - 13 minimale Depression

14 - 19 leichte Depression

20 - 28 mittelschwere Depression

29 - 63 schwere Depression

Summe Seite 1:

Übertrag Seite 1:

Summe Seite 2:

Summe Seite 3:

Übertrag Seite 3:

Summe Seite 4:

Übertrag Seite 4:

Summe Seite 5:

Übertrag Seite 5:

Summe Seite 6:

Übertrag Seite 6:

Summe Seite 7:

Übertrag Seite 7:

Summe Seite 8:

Übertrag Seite 8:

Summe Seite 9:

Übertrag Seite 9:

Summe Seite 10:

Übertrag Seite 10:

Summe Seite 11:

Übertrag Seite 11:

Summe Seite 12:

Übertrag Seite 12:

Summe Seite 13:

Übertrag Seite 13:

Summe Seite 14:

Übertrag Seite 14:

Summe Seite 15:

Übertrag Seite 15:

Summe Seite 16:

Übertrag Seite 16:

Summe Seite 17:

Übertrag Seite 17:

Summe Seite 18:

Übertrag Seite 18:

Summe Seite 19:

Übertrag Seite 19:

Summe Seite 20:

Übertrag Seite 20:

Summe Seite 21:

Übertrag Seite 21:

Summe Seite 22:

Übertrag Seite 22:

Summe Seite 23:

Übertrag Seite 23:

Summe Seite 24:

Übertrag Seite 24:

Summe Seite 25:

Übertrag Seite 25:

Summe Seite 26:

Übertrag Seite 26:

Summe Seite 27:

Übertrag Seite 27:

Summe Seite 28:

Übertrag Seite 28:

Summe Seite 29:

Übertrag Seite 29:

Summe Seite 30:

Übertrag Seite 30:

Summe Seite 31:

Übertrag Seite 31:

Summe Seite 32:

Übertrag Seite 32:

Summe Seite 33:

Übertrag Seite 33:

Summe Seite 34:

Übertrag Seite 34:

Summe Seite 35:

Übertrag Seite 35:

Summe Seite 36:

Übertrag Seite 36:

Summe Seite 37:

Übertrag Seite 37:

Summe Seite 38:

Übertrag Seite 38:

Summe Seite 39:

Übertrag Seite 39:

Summe Seite 40:

Übertrag Seite 40:

Summe Seite 41:

Übertrag Seite 41:

Summe Seite 42:

Übertrag Seite 42:

Summe Seite 43:

Übertrag Seite 43:

Summe Seite 44:

Übertrag Seite 44:

Summe Seite 45:

Übertrag Seite 45:

Summe Seite 46:

Übertrag Seite 46:

Summe Seite 47:

Übertrag Seite 47:

Summe Seite 48:

Übertrag Seite 48:

Summe Seite 49:

Übertrag Seite 49:

Summe Seite 50:

Übertrag Seite 50:

Summe Seite 51:

Übertrag Seite 51:

Summe Seite 52:

Übertrag Seite 52:

Summe Seite 53:

Übertrag Seite 53:

Summe Seite 54:

Übertrag Seite 54:

Summe Seite 55:

Übertrag Seite 55:

Summe Seite 56:

Übertrag Seite 56:

Summe Seite 57:

Übertrag Seite 57:

Summe Seite 58:

Übertrag Seite 58:

Summe Seite 59:

Übertrag Seite 59:

Summe Seite 60:

Übertrag Seite 60:

Summe Seite 61:

Übertrag Seite 61:

Summe Seite 62:

Übertrag Seite 62:

Summe Seite 63:

Übertrag Seite 63:

Summe Seite 64:

Übertrag Seite 64:

Summe Seite 65:

Übertrag Seite 65:

Summe Seite 66:

Übertrag Seite 66:

Summe Seite 67:

Übertrag Seite 67:

Summe Seite 68:

Übertrag Seite 68:

Summe Seite 69:

Übertrag Seite 69:

Summe Seite 70:

Übertrag Seite 70:

Summe Seite 71:

Übertrag Seite 71:

Summe Seite 72:

Übertrag Seite 72:

Summe Seite 73:

Übertrag Seite 73:

Summe Seite 74:

Übertrag Seite 74:

Summe Seite 75:

Übertrag Seite 75:

Summe Seite 76:

Übertrag Seite 76:

Summe Seite 77:

Übertrag Seite 77:

Summe Seite 78:

Übertrag Seite 78:

Summe Seite 79:

Übertrag Seite 79:

Summe Seite 80:

Übertrag Seite 80:

Summe Seite 81:

Übertrag Seite 81:

Summe Seite 82:

Übertrag Seite 82:

Summe Seite 83:

Übertrag Seite 83:

Summe Seite 84:

Übertrag Seite 84:

Summe Seite 85:

Übertrag Seite 85:

Summe Seite 86:

Übertrag Seite 86:

Summe Seite 87:

Übertrag Seite 87:

Summe Seite 88:

Übertrag Seite 88:

Summe Seite 89:

Übertrag Seite 89:

Summe Seite 90:

Übertrag Seite 90:

Summe Seite 91:

Übertrag Seite 91:

Summe Seite 92:

Übertrag Seite 92:

Summe Seite 93:

Übertrag Seite 93:

Summe Seite 94:

Übertrag Seite 94:

Summe Seite 95:

Übertrag Seite 95:

Summe Seite 96:

Übertrag Seite 96:

Summe Seite 97:

Übertrag Seite 97:

Summe Seite 98:

Übertrag Seite 98:

Summe Seite 99:

Übertrag Seite 99:

Summe Seite 100:

Übertrag Seite 100:

Summe Seite 101:

Übertrag Seite 101:

Summe Seite 102:

Übertrag Seite 102:

Summe Seite 103:

Übertrag Seite 103:

Summe Seite 104:

Übertrag Seite 104:

Summe Seite 105:

Übertrag Seite 105:

Summe Seite 106:

Übertrag Seite 106:

Summe Seite 107:

Übertrag Seite 107:

Summe Seite 108:

Übertrag Seite 108:

Summe Seite 109:

Übertrag Seite 109:

Summe Seite 110:

Übertrag Seite 110:

Summe Seite 111:

Übertrag Seite 111:

Summe Seite 112:

Übertrag Seite 112:

Summe Seite 113:

Übertrag Seite 113:

Summe Seite 114:

Übertrag Seite 114:

Summe Seite 115:

Übertrag Seite 115:

Summe Seite 116:

Übertrag Seite 116:

Summe Seite 117:

Übertrag Seite 117:

Summe Seite 118:

Übertrag Seite 118:

Summe Seite 119:

Übertrag Seite 119:

Summe Seite 120:

Übertrag Seite 120:

Summe Seite 121:

Übertrag Seite 121:

Summe Seite 122:

Übertrag Seite 122:

Summe Seite 123:

Übertrag Seite 123:

Summe Seite 124:

Übertrag Seite 124:

Summe Seite 125:

Übertrag Seite 125:

Summe Seite 126:

Übertrag Seite 126:

Summe Seite 127:

Übertrag Seite 127:

## Beispiel: Evaluation von Psychotherapieformen bei Depression

Experimentelle Bedingung  
(Gruppen von  $n = 50$ )

Psychotherapie

Klassisch

Pre-BDI



Post-BDI



Online

Pre-BDI



Post-BDI



## Einlesen des Datensatzes mit read.table()

```
fname      = file.path(data_dir, "psychotherapie_datensatz.csv")  
D          = read.table(fname, 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



## Datensatzübersicht mit View()



	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
9	Klassisch	18	11
10	Klassisch	18	14
11	Klassisch	20	10
12	Klassisch	17	15
13	Klassisch	16	17
14	Klassisch	18	12
15	Klassisch	16	10
16	Klassisch	18	13

---

Beispieldatensatz

**Datenvorverarbeitung**

Deskriptive Statistiken

Visualisierung

Parameterschätzung

Konfidenzintervalle

Hypothesentests

## Datenvorverarbeitung

- Studienfokus ist die **Veränderung** der Depressionsymptomatik durch Therapieformen.
- Für jede Proband:in der ergibt sich diese Veränderung als **Differenz** von Post.BDI - Pre.BDI.
- Eine Reduktion der Depressionssymptomatik ergibt dabei eine **negative Zahl**.
- Es ist sinnvoller, Verbesserungen mit **positiven Zahlen** zu repräsentieren.
- Als Maß des Therapieeffekts bei Proband:in  $i$  bietet sich also 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 folgender Interpretation

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

# Beispieldatensatz

## Datenvorverarbeitung

Hinzufügen einer  $\Delta$ BDI Spalte zum Dataframe

```
fname      = file.path(data_dir, "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

## Bedingungsabhängige Auswertung deskriptiver Statistiken

```
# Initialisierung eines Dataframes
ntp      = c("Klassisch", "Online")
ntp      = length(tp)
S        = data.frame(
  n      = rep(NaN,ntp),
  Max    = rep(NaN,ntp),
  Min    = rep(NaN,ntp),
  Median = rep(NaN,ntp),
  Mean   = rep(NaN,ntp),
  Var    = rep(NaN,ntp),
  Std    = rep(NaN,ntp),
  row.names = tp)

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

# Iterationen über Therapiebedingungen
for(i in 1:ntp){
  data = D$Delta.BDI[D$Bedingung == tp[i]]
  S$n[i] = length(data)
  S$Max[i] = max(data)
  S$Min[i] = min(data)
  S$Median[i] = median(data)
  S$Mean[i] = mean(data)
  S$Var[i] = var(data)
  S$Std[i] = sd(data)
}
```

## Bedingungsabhängige Auswertung deskriptiver Statistiken

*# Ausgabe*

```
print.AsIs(S)
```

```
>           n Max Min Median Mean  Var  Std
> Klassisch 50 12 -1      6 6.16 7.08 2.66
> Online    50  9  1      5 4.92 3.91 1.98
```

- Die Anzahl der Proband:innen in beiden Therapiegruppen ist gleich.
- Die Spannweite 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.

## Bedingungsabhängige Visualisierung 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 Boz
# Horizontale Achsenbeschriftung
# x-Achse bei y = 0
# y-Achse bei x = 0
# Non-Bold Titel
# Textvergrößerungsfaktor
# Titelttextvergrößerungsfaktor

# Linkes Panel: Balkendiagramm mit Fehlerbalken
mw      = S$Mean
sd      = S$Std
names(mw) = tp
x = barplot(
  mw,
  col      = "gray90",
  ylim     = c(0,12),
  xlim     = c(0,3),
  xlab     = "Bedingung",
  main     = TeX("$\\Delta$ BDI$"))

# Gruppenmittelwert
# Gruppenstandardabweichung
# barplot braucht x-Werte als names
# Ausgabe der x-Ordinaten (?barplot für Details)
# Mittelwerte = Balkenhöhe
# Balkenfarbe
# y-Achsenbegrenzung
# x-Achsenbegrenzung
# x-Achsenbeschriftung
# Titel

arrows(
  x0      = x,
  y0      = mw - sd,
  x1      = x,
  y1      = mw + sd,
  code    = 3,
  angle   = 90,
  length  = 0.05)

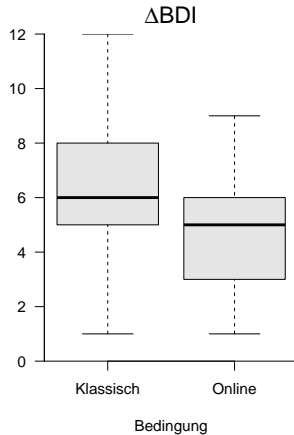
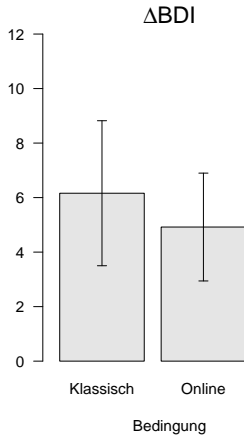
# arrows() für Fehlerbalken (siehe ?arrows)
# arrow start x-ordinate
# arrow start y-ordinate
# arrow end x-ordinate
# arrow end y-ordinate
# Pfeilspitzen beiderseits
# Pfeilspitzenwinkel -> Linie
# Linielänge

# Rechtes Panel: Bozplot
boxplot(
  D$Delta.BDI ~ D$Bedingung,
  ylim     = c(0,12),
  col      = "gray90",
  ylab     = "",
  xlab     = "Bedingung",
  main     = TeX("$\\Delta$ BDI$"))

# Gruppierung der Delta.BDI Daten nach D$Bedingung
# y-Achsenbegrenzung
# Bozfarbe
# y-Achsenbeschriftung
# x-Achsenbeschriftung
# Titel
```



## Bedingungsabhängige Visualisierung deskriptiver Statistiken



---

Beispieldatensatz

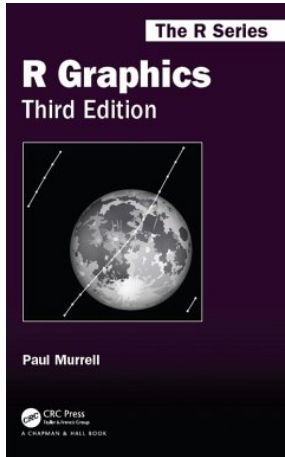
Deskriptive Statistiken

**Visualisierung**

Parameterschätzung

Konfidenzintervalle

Hypothesentests



## 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

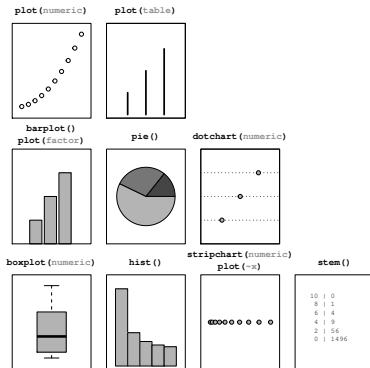
- **Erstellung und bedarfsgerechte Anpassung von Abbildungen**
- **Eher low-level, fine tuning orientiert**

### 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.

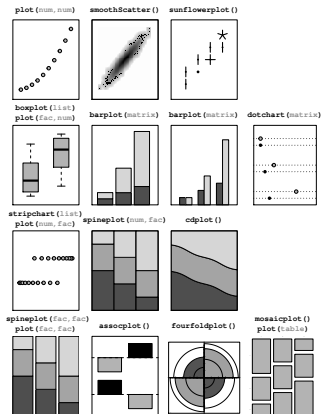
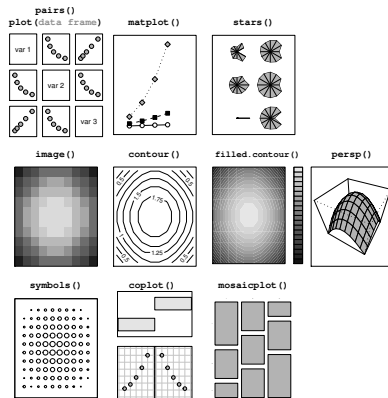


Figure 2.6

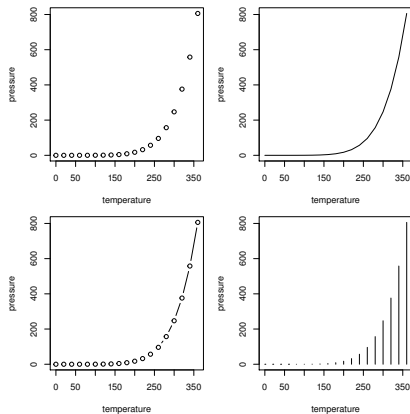
High-level base graphics plotting functions for producing plots of two variables. Where the function can be used to produce more than one type of plot, the relevant data type is shown (in gray). For example `plot(num, fac)` represents calling the `plot()` function with a numeric vector as the first argument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



**Figure 2.7**

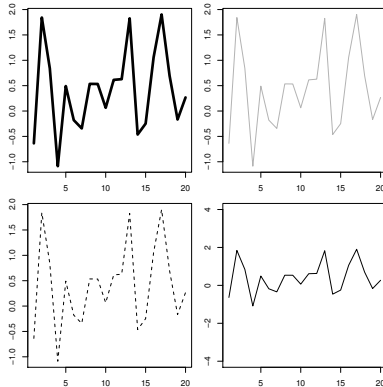
High-level base graphics plotting functions for producing plots of many variables. Where the function can be used to produce more than one type of plot, the relevant data type is shown (in gray).





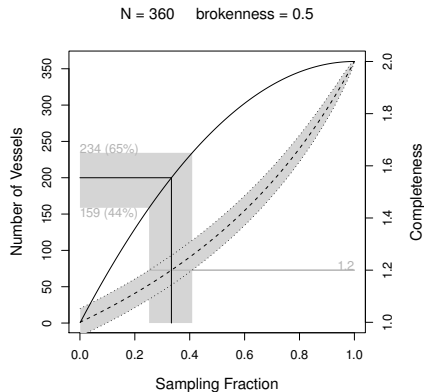
**Figure 2.2**

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()
```

## Histogramme

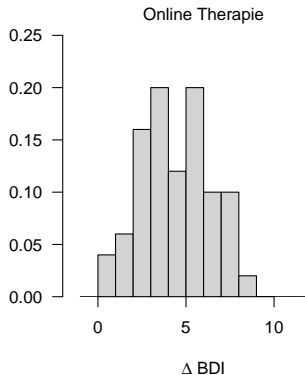
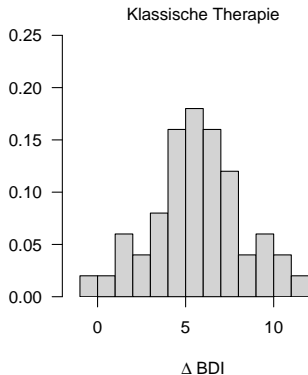
```
# Histogrammparameter
h          = 1                                # gewünschte Klassenbreite
b_0        = min(D$Delta.BDI)                 # b_0
b_k        = max(D$Delta.BDI)                 # b_0
k          = ceiling((b_k - b_0)/h)           # Anzahl der Klassen
b          = seq(b_0, b_k, by = h)             # Klassen [b_{j-1}, b_j[
ylimits    = c(0, 25)                         # y-Achsenlimits
xlimits    = c(-2, 14)                        # x-Achsenlimits
therapie   = c("Klassisch", "Online")         # Therapiebedingungen
labs       = c("Klassische Therapie",        # Abbildungslabel
               "Online Therapie")

# Abbildungsparameter
par(                                              # für Details siehe ?par
mfcol      = c(1, 2),                          # 1 x 2 Panelstruktur
family     = "sans",                          # Serif-freier Fonttyp
pty        = "m",                             # Maximale Abbildungsregion
bty        = "l",                             # L förmige Box
las        = 1,                               # Horizontale Achsenbeschriftung
xaxs       = "i",                             # x-Achse bei y = 0
 yaxs      = "i",                             # y-Achse bei x = 0
font.main  = 1,                               # Non-Bold Titel
cex        = 1,                               # Textvergrößerungsfaktor
cex.main   = 1)                              # Titeltextvergrößerungsfaktor

# Iteration über Therapiebedingungen
for(i in 1:2){
  hist(
    D$Delta.BDI[D$Bedingung == therapie[i]],    # Delta.BDI Werte von Therapiebedingung i
    breaks = b,                                # Histogrammklassen
    freq   = F,                               # normierte relative Häufigkeit
    xlim   = xlimits,                         # x-Achsenlimits
    ylim   = ylimits,                         # y-Achsenlimits
    xlab    = TeX("$\\Delta$ BDI"),           # x-Achsenbeschriftung
    ylab    = "",                             # y-Achsenbeschriftung
    main    = labs[i])                       # Titelbeschriftung
}

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

## Histogramme



---

Beispieldatensatz

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 nun das Normalverteilungsmodell zugrunde.

Wir nehmen also an, dass die  $\Delta$  BDI Werte Realisierungen von unabhängig verteilten Zufallsvariablen

$$v_{ij} \sim N(\mu_i, \sigma_i^2), i = 1, 2, j = 1, \dots, 50 \quad (2)$$

sind, wobei  $i$  die Therapiebedingung (1 = Klassisch, 2 = Online) und  $j$  den Proband:innen Index in der  $i$ ten experimentellen Bedingung bezeichnen. Innerhalb einer Bedingung sind diese Zufallsvariablen also unabhängig und identisch verteilt.

Dies entspricht der Annahme, dass sich der  $\Delta$ BDI Wert einer Proband:in durch Addition einer normalverteilten Fehlervariable mit Erwartungswertparameter 0 und Varianzparameter  $\sigma_i^2$  zu den innerhalb einer Therapiebedingung identischen Wert  $\mu_i$  ergibt.



# Parameterschätzung

Zur Parameterschätzung im vorliegenden Modell nutzen wir

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

```
# Initialisierung eines Dataframes
tp      = c("Klassisch", "Online")
ntp     = length(tp)
theta_hats = data.frame(
  mu_ML      = rep(NA, ntp),
  sigsqv_VAR = rep(NA, ntp))

# Therapiebedingungen
# Anzahl Therapiebedingungen
# Dataframeerzeugung
# ML Schätzer für \mu_i
# Varianzschätzer für \sigma^2_i

# Iterationen über Therapiebedingungen
for(i in 1:ntp){
  data      = D$Delta.BDI[D$Bedingung == tp[i]]
  theta_hats$mu_ML[i] = mean(data)
  theta_hats$sigsqr_VAR[i] = var(data)
}

# Ausgabe
print(theta_hats)

> mu_ML sigsqv_VAR
> 1  6.16      7.08
> 2  4.92      3.91
```

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

$$\hat{\mu}_1 = 6.16, \quad \hat{\mu}_2 = 4.92, \quad \hat{\sigma}_1^2 = 7.08, \quad \hat{\sigma}_2^2 = 3.91. \quad (3)$$

Die mit diesen Tipps assoziierte Unsicherheit ist hier nicht angegeben.

---

Beispieldatensatz

Visualisierung

Deskriptive Statistiken

Parameterschätzung

**Konfidenzintervalle**

Hypothesentest

# Konfidenzintervalle

## Konfidenzintervalle für die Erwartungswertparameterschätzer

```
# Analyseparameter
t      = c("Klassisch", "Online")
ntp    = length(tp)
n      = 50
C      = data.frame(
  G_u    = rep(NA,n,tp),
  mu_hat = rep(NA,n,tp),
  G_o    = rep(NA,n,tp),
  row.names = tp)

# Therapiebedingungen
# Anzahl an Therapiebedingungen
# Anzahl von Beobachtungen pro Therapiebedingung
# Dataframeerzeugung
# untere KI Grenze
# Erwartungswertparameterschätzer
# obere KI Grenze
# Therapiebedingungen

# Konfidenzintervallparameter
delta  = 0.95
psi_inv = qt((1+delta)/2,n-1)

# Konfidenzintervallevaluation
for(i in 1:ntp){
  data      = D$Delta.BDI[D$Bedingung == t[i]]
  X_bar     = mean(data)
  S         = sd(data)
  C$G_u[i]  = X_bar - (S/sqrt(n))*psi_inv
  C$mu_hat[i] = X_bar
  C$G_o[i]  = X_bar + (S/sqrt(n))*psi_inv
}

# Ausgabe
print.AsIs(C)

>           G_u mu_hat G_o
> Klassisch 5.40  6.16 6.92
> Online    4.36  4.92 5.48
```

# Konfidenzintervalle

## Konfidenzintervalle für die Varianzparameterschätzer

```
# Analyseparameter
t      = c("Klassisch", "Online")
ntp    = length(tp)
n      = 50
C      = data.frame(
  G_u   = rep(NaN,ntp),
  sigsq_r_hat = rep(NaN,ntp),
  G_o   = rep(NaN,ntp),
  row.names = tp)

# Therapiebedingungen
# Anzahl an Therapiebedingungen
# Anzahl von Beobachtungen pro Therapiebedingung
# Dataframeerzeugung
# untere KI Grenze
# Varianzparameterschätzer
# obere KI Grenze
# Therapiebedingungen

# Konfidenzintervallparameter
delta  = 0.95
xi_1   = qchisq((1-delta)/2, n - 1)
xi_2   = qchisq((1+delta)/2, n - 1)

# Konfidenzintervallevaluation
for(i in 1:ntp){
  data      = D$Delta.BDI[D$Bedingung == t[i]] # Stichprobenrealisierung
  S2        = var(data)                        # Stichprobenvarianz
  C$G_u[i]  = (n-1)*S2/xi_2                    # untere KI Grenze
  C$sigsqr_hat[i] = S2                        # Varianzparameterschätzer
  C$G_o[i]  = (n-1)*S2/xi_1                    # obere KI Grenze
}

# Ausgabe
print.AsIs(C)

>           G_u sigsq_r_hat  G_o
> Klassisch 4.94         7.08 10.99
> Online    2.73         3.91  6.07
```

---

Beispieldatensatz

Deskriptive Statistiken

Visualisierung

Parameterschätzung

Konfidenzintervalle

**Hypothesentests**

## Anwendungsszenario und statistisches Modell

Wir nehmen an, dass die  $\Delta$  BDI Werte, also der uns vorliegende Datensatz Realisierungen von unabhängig verteilten Zufallsvariablen

$$v_{ij} \sim N(\mu_i, \sigma_i^2), i = 1, 2, j = 1, \dots, 50 \quad (4)$$

sind, wobei  $i$  die Therapiebedingung ( $1 = \text{Klassisch}$ ,  $2 = \text{Online}$ ) und  $j$  den Proband:innen Index in der  $i$ ten experimentellen Bedingung bezeichnen. Innerhalb einer Bedingung sind diese Zufallsvariablen also unabhängig und identisch verteilt.

Die Parameter  $\mu_1, \mu_2, \sigma^2$  sind unbekannt sind. Wir beabsichtigen das Quantifizieren der Unsicherheit beim inferentiellen Vergleich von  $\mu_1$  mit  $\mu_2$ .

Dafür können wir einen Zweistichproben-T-Test bei unabhängigen Stichproben unter Annahme identischer Varianz durchführen.

Wir wollen die Hypothesen  $H_0 : \mu_1 = \mu_2$  und  $H_1 : \mu_1 \neq \mu_2$  mit einem Signifikanzniveau von  $\alpha_0 = 0.05$  testen und verwenden dafür einen **zweiseitigen Zweistichproben-Test**.

## (1) Statistisches Modell und Testhypothesen

Wir betrachten die einfache Nullhypothese und die zusammengesetzte Alternativhypothese

$$H_0 : \mu_1 - \mu_2 = 0 \Leftrightarrow \mu_1 = \mu_2 \text{ und} \quad (5)$$

$$H_1 : \mu_1 - \mu_2 \neq 0 \Leftrightarrow \mu_1 \neq \mu_2 \quad (6)$$

## (2) Definition und Analyse der Teststatistik

Die T-Teststatistik für den Zweistichproben-T-Test ist gegeben durch

$$T = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\bar{y}_1 - \bar{y}_2 - \mu_0}{s_{12}} \right) \quad (7)$$

und dessen Verteilung durch

$$T \sim t(\delta, n_1 + n_2 - 2) \text{ mit } \delta = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \left( \frac{\mu_1 - \mu_2 - \mu_0}{\sigma} \right). \quad (8)$$



## (3) Definition des Tests

$$\phi(v) := 1_{\{|T| \geq k\}} = \begin{cases} 1 & |T| \geq k \\ 0 & |T| < k \end{cases}. \quad (9)$$

## (4) Analyse der Testgütefunktion

### Theorem (Testgütefunktion)

Es sei  $\phi$  der im obigen Modell formulierte Zweistichproben-T-Test. Dann ist die Testgütefunktion von  $\phi$  gegeben durch

$$\begin{aligned} q_\phi : \mathbb{R}^2 \rightarrow [0, 1], (\mu_1, \mu_2) &\mapsto q_\phi(\mu_1, \mu_2) \\ &:= 1 - \psi(k; \delta, n_1 + n_2 - 2) + \psi(-k; \delta, n_1 + n_2 - 2) \end{aligned} \quad (10)$$

wobei  $\psi(\cdot; \delta, n_1 + n_2 - 2)$  die KVF der nichtzentralen  $t$ -Verteilung mit Nichtzentralitätsparameter

$$\delta := \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\mu_1 - \mu_2}{\sigma} \quad (11)$$

und Freiheitsgradparameter  $n_1 + n_2 - 2$  bezeichnet.

## (5) Testumfangkontrolle

### Theorem (Testumfangkontrolle)

$\phi$  sei der im obigen Testszenario definierte Test. Dann ist  $\phi$  ein Level- $\alpha_0$ -Test mit Testumfang  $\alpha_0$ , wenn der kritische Wert definiert ist durch

$$k_{\alpha_0} := \psi^{-1} \left( 1 - \frac{\alpha_0}{2}; n_1 + n_2 - 2 \right), \quad (12)$$

wobei  $\psi^{-1}(\cdot; n_1 + n_2 - 2)$  die inverse KVF der  $t$ -Verteilung mit  $n_1 + n_2 - 2$  Freiheitsgraden ist.

## Manueller Zweistichproben-T-Test

```
# Datenauswahl
x_1      = D$Delta.BDI[D$Bedingung == "Klassisch"]
x_2      = D$Delta.BDI[D$Bedingung == "Online"]
n_1      = length(x_1)
n_2      = length(x_2)
alpha_0  = 0.05
k_alpha_0 = qt(1 - (alpha_0/2), n_1+n_2-2)
x_bar_1  = mean(x_1)
x_bar_2  = mean(x_2)
s_12     = sqrt((sum((x_1-x_bar_1)^2)+sum((x_2-x_bar_2)^2))/
               (n_1+n_2-2))
t        = sqrt((n_1*n_2)/(n_1+n_2))*((x_bar_1-x_bar_2)/s_12)
if(abs(t) >= k_alpha_0){
  phi = 1
} else {
  phi = 0
}
pval     = 2*(1 - pt(abs(t), n_1+n_2-2))

# \Delta.BDI Daten Klassische Therapie
# \Delta.BDI Daten Klassische Therapie
# Stichprobengröße n_1
# Stichprobengröße n_2
# Signifikanzniveau
# kritischer Wert
# x_bar_1
# x_bar_2
# gepoolte Standardabweichung s_12
# Zweistichproben-T-Teststatistik
# Test 1_{|T(X)| >= k_alpha_0}
# Ablehnen von H_0
# Nicht Ablehnen von H_0
# p-Wert
```

## Manueller Zweistichproben-T-Test

```
# Ausgabe
cat("\nx_bar_1      = ", x_bar_1,
    "\nx_bar_2      = ", x_bar_2,
    "\nfg          = ", n_1 + n_2 - 2,
    "\nSignifikanzlevel = ", alpha_0,
    "\nKritischer Wert  = ", k_alpha_0,
    "\nTeststatistik   = ", t,
    "\nTestwert       = ", phi,
    "\np-Wert        = ", pval)
```

```
>
> x_bar_1      = 6.16
> x_bar_2      = 4.92
> fg          = 98
> Signifikanzlevel = 0.05
> Kritischer Wert  = 1.98
> Teststatistik   = 2.65
> Testwert       = 1
> p-Wert        = 0.00951
```

⇒ Wir lehnen die Nullhypothese  $H_0 : \mu_1 = \mu_2$  ab.

## R Implementation des Zweistichproben-T-Tests

```
# Automatischer Zweistichproben-T-Test
varphi = t.test(                                # ?t.test für Details
  x_1,                                           # Datensatz x_1
  x_2,                                           # Datensatz x_2
  var.equal = TRUE,                             # \sigma_1^2 = \sigma_2^2
  alternative = c("two.sided"),                 # H_1: \mu_1 \neq \mu_2
  conf.level = 1-alpha_0)                       # \delta = 1 - \alpha_0 (sic!)
```

```
# Ausgabe
print(varphi)
```

```
>
> Two Sample t-test
>
> data: x_1 and x_2
> t = 3, df = 98, p-value = 0.01
> alternative hypothesis: true difference in means is not equal to 0
> 95 percent confidence interval:
>  0.31 2.17
> sample estimates:
> mean of x mean of y
>    6.16    4.92
```

```
# Genauere Ausgabe t
paste(varphi[1])
```

```
> [1] "c(t = 2.64516155336263)"
```

```
# Genauere Ausgabe p
paste(varphi[3])
```

```
> [1] "0.00951137026459394"
```

## References

---

Murrell, Paul. 2019. *R Graphics*. Third edition. The R Series. Boca Raton: CRC Press, Taylor & Francis Group.