



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>		
21.01.25	Deskriptive Statistik	(11) Anwendungsbeispiel	Seminar
28.01.25	Deskriptive Statistik	(11) Anwendungsbeispiel, Q&A	Seminar

(7) Häufigkeitsverteilungen

Motivation

Beispieldatensatz

Häufigkeitsverteilungen

Histogramme

Programmierübungen und Selbstkontrollfragen

Motivation

Beispieldatensatz

Häufigkeitsverteilungen

Histogramme

Programmierübungen und Selbstkontrollfragen

Definition und Ziele der Deskriptive Statistik

- Die Deskriptive Statistik ist die *beschreibende* Statistik.
- Ziel der Deskriptiven Statistik ist es, Daten übersichtlich darzustellen.
- Deskriptive Statistik ist insbesondere bei großen Datensätzen sinnvoll.
- Die Deskriptive Statistik berechnet zusammenfassende Maße aus Daten.

Typische Methoden der Deskriptiven Statistik

- Häufigkeitsverteilungen und Histogramme
- Verteilungsfunktionen und Quantile
- Maße der zentralen Tendenz und der Datenvariabilität
- Zusammenhangsmaße

Die Deskriptive Statistik benutzt keine probabilistischen Modelle, aber die Methoden der Deskriptiven Statistik ergeben nur vor dem Hintergrund probabilistischer Modelle Sinn.

Motivation

Beispieldatensatz

Häufigkeitsverteilungen

Histogramme

Programmierübungen und Selbstkontrollfragen

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		Wahr	Schwach	Stark
		0-1	2-3	4-5
<p>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 der letzten zwei Wochen, einschließlich heute, gefühlt haben. Kennen Sie die Zahl neben der Aussage an, die Sie sich herausgehört haben (0, 1, 2 oder 3). Falls in einer Gruppe mehrere Aussagen gleichwahrscheinlich für zutreffend, können Sie für Aussage mit der höchsten Zahl an Antworten Ihre Antwort, aber Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, die gilt auch für Gruppe 18 (Veränderungen der Schlafgewohnheiten) oder Gruppe 19 (Veränderungen des Appetits).</p>				
<p>1.) Traurigkeit</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>6.) Bestürzungsempfinden</p> <p>0 Ich habe keine 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>11.) Unruhe</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 bewegen oder etwas tun muss.</p>	<p>17.) Reizbarkeit</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>2.) pessimismus</p> <p>0 Ich sehe nicht mutlos in die Zukunft. 1 Ich sehe mutlos in die Zukunft als Ganzes. 2 Ich bin mutlos und erwarte nicht, dass meine Situation besser wird. 3 Ich glaube, dass meine Zukunft hoffnungslos ist und nur noch schlechter wird.</p>	<p>7.) Selbstabwertung</p> <p>0 Ich halte von mir genauso viel wie immer. 1 Ich habe Vertrauen in mich verloren. 2 Ich vermute von mir enttäuscht. 3 Ich lehne mich völlig ab.</p>	<p>12.) Interessenverlust</p> <p>0 Ich habe das Interesse an anderen Menschen oder an Dingen zum größten Teil 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>18.) Veränderungen des Appetits</p> <p>0 Mein Appetit hat sich nicht verändert. 1a Mein Appetit ist etwas schlechter als sonst. 2a Mein Appetit ist etwas größer als sonst. 3a Mein Appetit ist viel schlechter als sonst. 3b Mein Appetit ist viel größer als sonst. 3c Ich habe überhaupt keinen Appetit. 3d Ich habe ständig Heißhunger.</p>	
<p>3.) Versagensgefühle</p> <p>0 Ich fühle mich nicht als Versager. 1 Ich habe häufiger Versagensgefühle. 2 Wenn ich zurückblicke, sehe ich eine Menge Fehlentscheidungen. 3 Ich habe das Gefühl, ich Mensch ein völliger Versager zu sein.</p>	<p>8.) Selbstvorwürfe</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>13.) Entscheidungsfähigkeit</p> <p>0 Ich bin so entscheidungsfreudig 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>19.) Konzentrationschwierigkeiten</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>4.) Verlust von Freude</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>9.) Selbstmordgedanken</p> <p>0 Ich denke nicht daran, mir etwas anzutun. 1 Ich denke manchmal an Selbstmord, aber ich würde es nicht tun. 2 Ich möchte mich am liebsten umbringen. 3 Ich würde mich umbringen, wenn ich die Gelegenheit dazu hätte.</p>	<p>14.) Wertlosigkeit</p> <p>0 Ich fühle mich nicht wertlos. 1 Ich fühle 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>20.) Ermüdung oder Erschöpfung</p> <p>0 Ich fühle mich nicht müde 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 so müde oder erschöpft, dass ich fast nichts mehr tun kann.</p>	
<p>5.) Schuldgefühle</p> <p>0 Ich habe keine besonderen Schuldgefühle. 1 Ich habe oft Schuldgefühle wegen Dingen, die ich getan habe oder hätte tun sollen. 2 Ich habe die meisten Zeit Schuldgefühle. 3 Ich habe ständig Schuldgefühle.</p>	<p>10.) Weinen</p> <p>0 Ich weine nicht öfter als früher. 1 Ich weine jetzt mehr als früher. 2 Ich weine beim geringsten Anlass. 3 Ich möchte gar weinen, aber ich kann nicht.</p>	<p>15.) Energieverlust</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 schaffen. 3 Ich habe keine Energie mehr, um überhaupt noch etwas zu tun.</p>	<p>21.) Verlust an sexuellem Interesse</p> <p>0 Mein Interesse an Sexualität hat sich in letzter Zeit 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>	

0 - 8 keine Depression

9 - 13 minimale Depression

14 - 19 leichte Depression

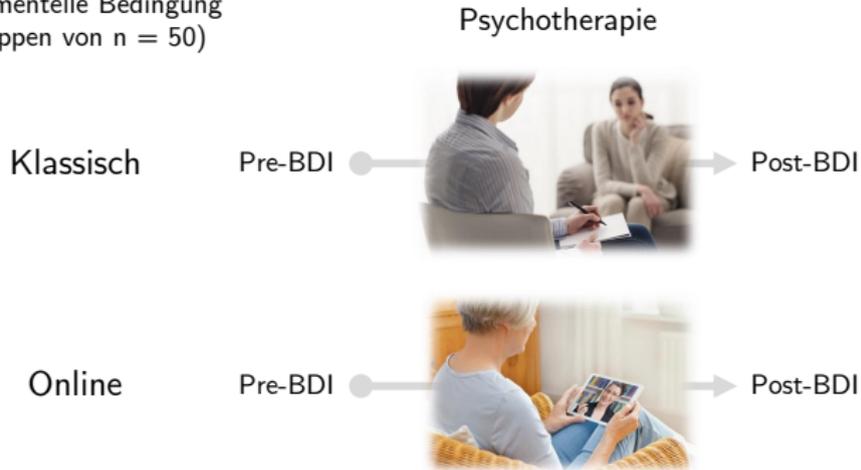
20 - 28 mittelschwere Depression

29 - 63 schwere Depression

<p>16.) Veränderungen der Schlafgewohnheiten</p> <p>0 Meine Schlafgewohnheiten haben sich nicht verändert. 1a Ich schlafe etwas mehr als sonst. 2a Ich schlafe viel mehr als sonst. 3a Ich schlafe weniger als sonst. 3b Ich schlafe fast den ganzen Tag. 3c Ich wache 1-2 Stunden früher auf als gewöhnlich und kann dann nicht mehr einschlafen.</p>	<p>22.) Somatischer Status 1:</p> <p>0</p>
<p>23.) Übertrag Status 1:</p> <p>0</p>	<p>24.) Somatischer Status 1-2:</p> <p>0</p>

Beispiel: Evaluation von Psychotherapieformen bei Depression

Experimentelle Bedingung
(Gruppen von $n = 50$)



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

VSCode Interactive Viewers

Table Viewer

```
# Skript zur Demonstration des VSCode-Table-Viewers
# -----
# Datum: 18.09.2024
# Belinda Fleischmann

# Dateipfad definieren
data_dir_path <- file.path(
  getwd(), "~/0VGU/2024_wise_PDS/pragr-und-deskr-stat-25/Daten"
)
f <- file.path(data_dir_path, "psychotherapie_datensatz.csv")
file_path <- file.path(data_dir_path, "psychotherapie_datensatz.csv")

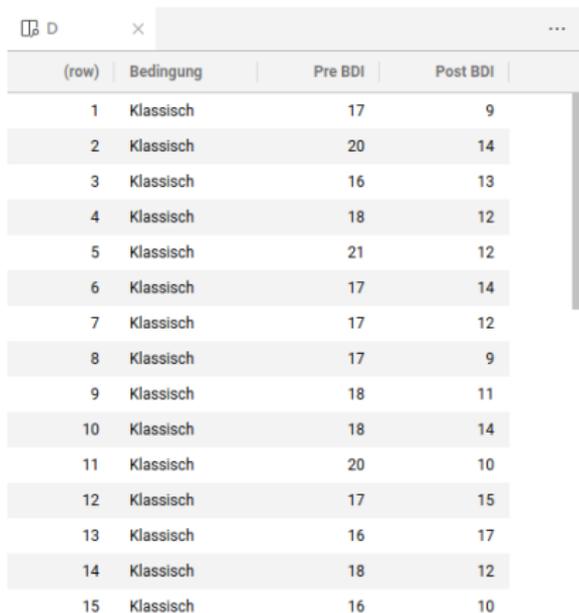
# Daten einlesen
D <- read.table(file_path, sep = ",", header = T)
D
```

(row)	Endpreis	Pro.BEI	Post.BEI
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
17	Klassisch	17	9
18	Klassisch	14	13
19	Klassisch	18	12

Mit dem Befehl `View()` oder im R **WORKSPACE** → **Global Environment** über das View Symbol  neben entsprechendem Objekt

[VS Code Wiki - Interactive viewers](#)

Datensatzübersicht mit View()



(row)	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

Motivation

Beispieldatensatz

Häufigkeitsverteilungen

Histogramme

Programmierübungen und Selbstkontrollfragen

Definition (Absolute und relative Häufigkeitsverteilungen)

$x := (x_1, \dots, x_n)$ mit $x_i \in \mathbb{R}$ sei ein *Datensatz* (manchmal auch “Urliste” genannt) und $A := \{a_1, \dots, a_k\}$ mit $k \leq n$ seien die im Datensatz vorkommenden verschiedenen Zahlenwerte (manchmal auch “Merkmalsausprägungen” genannt). Dann heißt die Funktion

$$h : A \rightarrow \mathbb{N}, a \mapsto h(a) := \text{Anzahl der } x_i \text{ aus } x \text{ mit } x_i = a \quad (1)$$

die *absolute Häufigkeitsverteilung* der Zahlenwerte von x und die Funktion

$$r : A \rightarrow [0, 1], a \mapsto r(a) := \frac{h(a)}{n} \quad (2)$$

die *relative Häufigkeitsverteilung* der Zahlenwerte von x .

Bemerkungen

- Absolute und relative Häufigkeitsverteilungen fassen Datensätze zusammen
- Absolute und relative Häufigkeitsverteilungen können einen ersten Datenüberblick geben

Berechnung der Häufigkeitsverteilungen

Die **absoluten Häufigkeiten** einer kategorialen Variable können in R mit der Funktion `table()` bestimmt werden.

```
x      <- D$Pre.BDI           # double vector der Pre BDI Werte
n      <- length(x)           # Anzahl der Datenwerte (hier: 100)
H      <- as.data.frame(table(x)) # absolute Häufigkeitsverteilung als Dataframe
names(H) <- c("a", "h")      # Benennen der Spalten im Dataframe
```

Die **relativen Häufigkeiten** erhält man, indem man die absoluten Häufigkeiten durch die Stichprobengröße n teilt.

```
H$r    <- H$h / n             # Neue Spalte mit relativen Häufigkeiten
print(H)                       # Ausgabe des Dataframes
```

```
      a h    r
1  14  1 0.01
2  15  3 0.03
3  16  6 0.06
4  17 17 0.17
5  18 21 0.21
6  19 20 0.20
7  20 17 0.17
8  21 12 0.12
9  22  2 0.02
10 23  1 0.01
```

Visualisierung der absoluten Häufigkeitsverteilung mit `barplot()`

```
h          <- H$h          # Kopie der absoluten Häufigkeiten in einen Vektor
names(h)   <- H$a          # Den Häufigkeitswerten die entsprechenden Kategorien zuweisen
barplot(   # Balkendiagramm
  h,       # absolute Häufigkeiten
  col = "gray90", # Balkenfarbe
  xlab = "a",    # x Achsenbeschriftung
  ylab = "h(a)", # y Achsenbeschriftung
  ylim = c(0, 25), # y Achsengrenzen
  las = 2,      # x Tick Orientierung (2: perpendicular to the axis)
  main = "Pre BDI" # Titel
)
```

Anmerkungen

- Allgemeine **R Graphics** Plot Parameter können mit `?par` und Plotspezifische Parameter auf der jeweiligen Plot Help Page (z.B. `?barplot`) nachgeschlagen werden

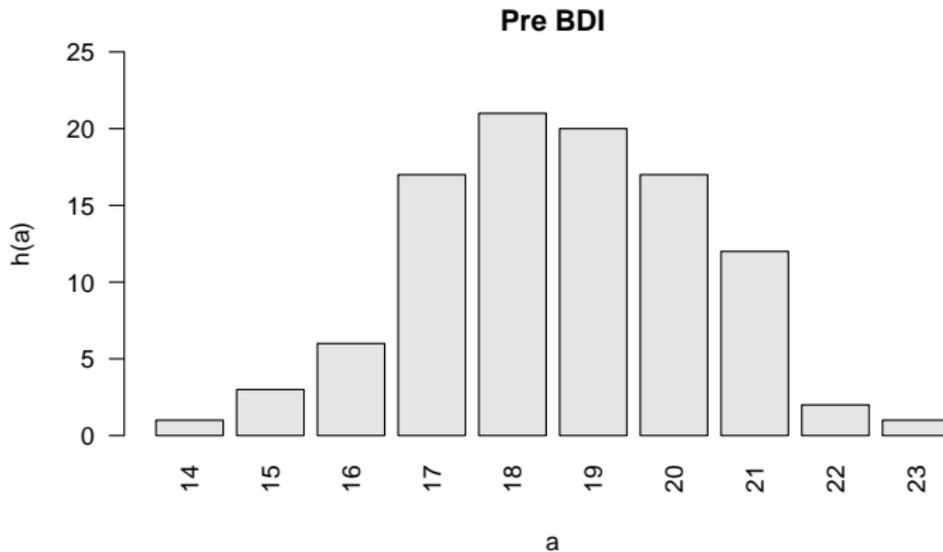
Speichern von Abbildungen mit `dev.copy2pdf()`

```
# Variante 1: Dateinamen im selben Verzeichnis wie das Skript erstellen
skript_dir_path <- sys.frame(1)$ofile # Verzeichnis des aktuellen Skripts ermitteln
abb_fname      <- file.path(         # Absoluter Pfad zur Zieldatei im Skriptverzeichnis
  skript_verzeichnis,
  "pds_7_ha_prebdi.pdf"
)

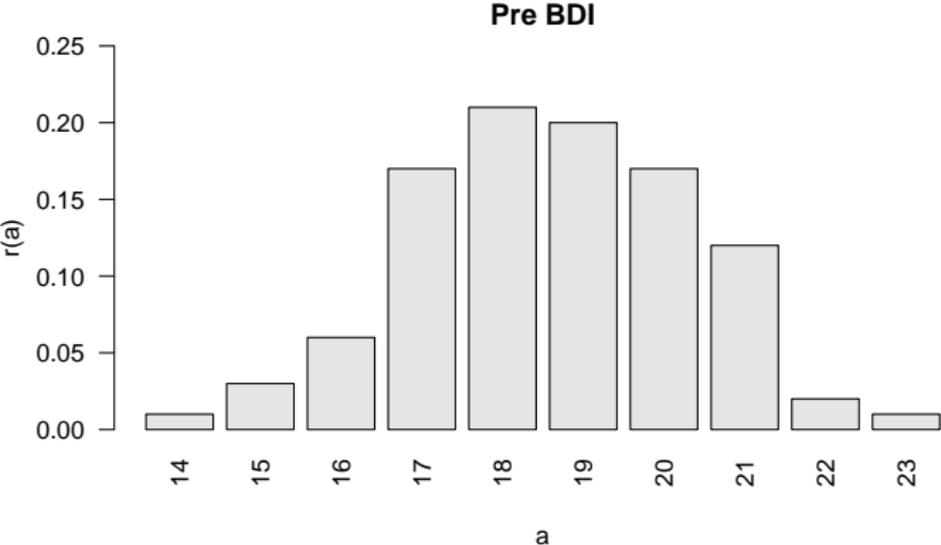
# Variante 2: Dateinamen in einem Unterverzeichnis "Abbildungen" erstellen
abb_dir_path   <- file.path(         # Absoluter Pfad zum Unterverzeichnis "Abbildungen"
  getwd(),     # Im Beispiel hier: im aktuellen working directory
  "VL_Abbildungen"
)
abb_fname     <- file.path(         # Absoluter Pfad zur Zieldatei im Abbildungsverzeichnis
  abb_dir_path,
  "pds_7_ha_prebdi.pdf"
)

# Speichern der Abbildung als PDF
dev.copy2pdf( # PDF Kopierfunktion
  file = abb_fname, # Dateiname
  width = 7, # Breite (inch)
  height = 4 # Höhe (inch)
)
```

Absolute Häufigkeitsverteilung aller Pre-BDI Werte



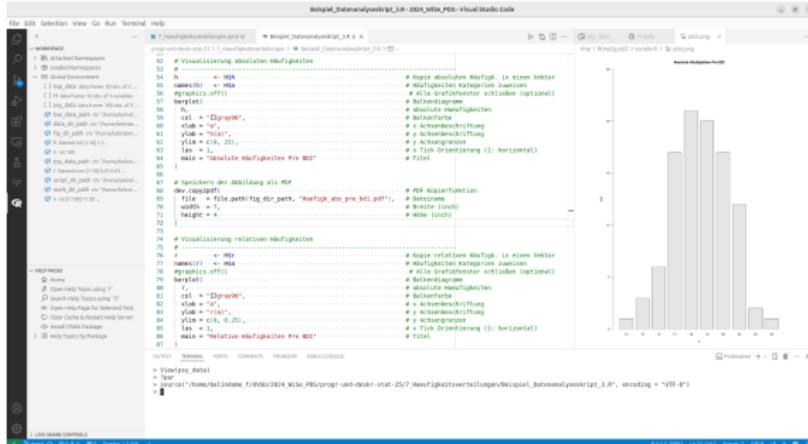
Relative Häufigkeitsverteilung aller Pre-BDI Werte



Grafikfenster und Plot Viewer in VS Code

Der base VSCode-R plot viewer (png device)

Wenn Visualisierungen in R erstellt werden, öffnet sich standardmäßig ein Grafikfenster (graphical device). Dieses Fenster wird verwendet, um die Plots anzuzeigen.



```
## Visualisierung absoluter Häufigkeiten
n <- nrow(Haerfig)
mapply(Haerfig, MARGIN="category", FUN=function(x) sum(Haerfig[,x]))
graphics.off()
par(mfrow=c(2,2))
col = "Dyspare"
ylim = "y"
ylim = "x"
ylim = c(0, 25)
lty = 1
mtext("absolute Häufigkeiten für MCC")
# Speichern der Abbildung als PDF
dev.capdf()
file = file.path(fig_dir_path, "Haerfig_abs_pre_MCC.pdf")
width = 7
height = 6

## Visualisierung relativer Häufigkeiten
n <- nrow(Haerfig)
mapply(Haerfig, MARGIN="category", FUN=function(x) sum(Haerfig[,x])/n)
graphics.off()
par(mfrow=c(2,2))
col = "Dyspare"
ylim = "y"
ylim = "x"
ylim = c(0, 25)
lty = 1
mtext("relative Häufigkeiten für MCC")
```

The screenshot shows the VS Code editor with an R script. The script contains two main plotting functions: one for absolute frequencies and one for relative frequencies. Both functions use `mapply()` to iterate over categories, `graphics.off()` to clear the plot area, `par(mfrow=c(2,2))` to set a 2x2 grid, and `col = "Dyspare"` to specify the variable. The absolute frequency plot uses `ylim = "y"` and `ylim = "x"`, while the relative frequency plot uses `ylim = "y"` and `ylim = "x"`. Both plots use `lty = 1` for the legend and `mtext()` to add a title. The script also includes a `dev.capdf()` call to save the plots as a PDF file. On the right side of the editor, a plot viewer window titled "Absolute Häufigkeiten" displays a histogram with a y-axis ranging from 0 to 25 and an x-axis with categories. The plot shows the distribution of absolute frequencies for the variable "Dyspare".

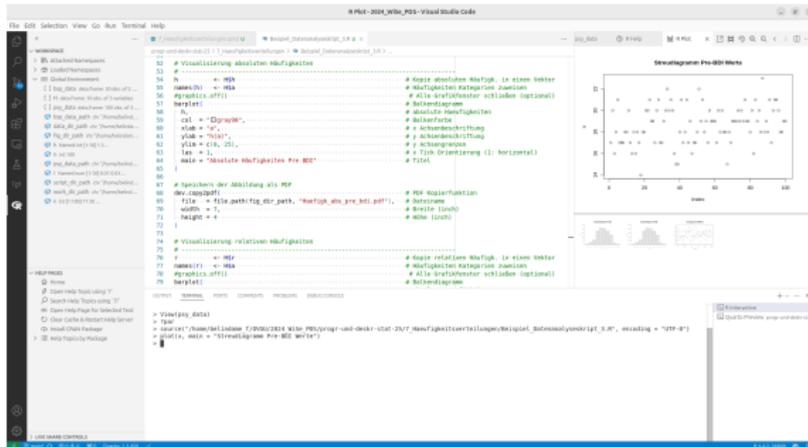
In Theorie können Grafikfenster mit `dev.new()` neu geöffnet und mit `dev.off()` geschlossen werden. Allerdings ist dieses Verhalten in VS Code nicht immer konsistent. Mit `graphics.off()` können jedoch alle noch offenen Grafikfenster geschlossen werden, bevor eine neue erstellt werden soll.

Für eine optimierte Anzeige von Grafiken innerhalb der VS Code-Umgebung ist es empfehlenswert, den [httpgd-based Plot Viewer](#) zu verwenden.

Grafikfenster und Plot Viewer in VS Code

httpgd-based Plot Viewer

Das R-Paket `httpgd` vereinfacht die Arbeit mit Grafiken über den VS Code `Plot viewer`.



The screenshot shows the VS Code interface with an R script editor on the left and a Plot Viewer window on the right. The R code defines a function `plot_absolute` that takes a data frame `df` and a variable `var` as input. It uses `ggplot2` to create a bar chart showing the absolute frequencies of the variable. The plot is styled with a white background, black axes, and blue bars. The x-axis is labeled with the variable name, and the y-axis is labeled 'Absolute Häufigkeiten'. The plot is displayed in the Plot Viewer window, which also shows a list of other plots in the background.

- Nach der Installation des Pakets kann dieser Plot Viewer in den VS Code-Einstellungen über die Option `r.plot.useHttpgd` aktiviert werden.
- Nach dem Ausführen eines Plot-Befehls öffnen sich die Grafiken automatisch im Plot Viewer.
- Nacheinander erstellte Grafiken werden nicht überschrieben, sondern als separate Ansichten hintereinander angezeigt und können durchgescrollt werden.

VS Code Wiki - Interactive viewers

Motivation

Beispieldatensatz

Häufigkeitsverteilungen

Histogramme

Programmierübungen und Selbstkontrollfragen

Definition (Histogramm)

Ein *Histogramm* ist ein Diagramm, in dem zu einem Datensatz $x = (x_1, \dots, x_n)$ mit verschiedenen Zahlenwerten $A := \{a_1, \dots, a_m\}$, $m \leq n$ über benachbarten Intervallen $[b_{j-1}, b_j[$, welche *Klassen* oder *Bins* genannt werden, für $j = 1, \dots, k$ Rechtecke mit

$$\text{Breite} \quad d_j = b_j - b_{j-1}$$

$$\text{Höhe} \quad h(a) \text{ oder } r(a) \text{ mit } a \in [b_{j-1}, b_j[$$

abgebildet sind, wobei $b_0 := \min A$ und $b_k := \max A$ angenommen werden soll.

Bemerkungen

- Das Aussehen eines Histogramms ist stark von der Anzahl k der Klassen abhängig.
- Im Folgenden sind einige konventionelle Methoden zur Bestimmung von k aufgeführt ($\lceil \cdot \rceil$ ist die Aufrundungsfunktion).

$$k := \lceil \frac{b_k - b_0}{h} \rceil$$

Bestimmung der Anzahl der Klassen aus gewünschter Klassenbreite h

$$k := \lceil \sqrt{n} \rceil$$

Klassenanzahl in Excel

$$k := \lceil \log_2 n + 1 \rceil$$

Implizite Normalverteilungsannahme (Sturges, 1926)

$$h := 3.49 \cdot \frac{S_n}{\sqrt[3]{n}}$$

Berechnung h nach Scott (1979): Min. MSE Dichteschätzung bei Normalverteilung

Berechnung und Visualisierung von Histogrammen

Berechnung und Visualisierung von Histogrammen mit `hist()`

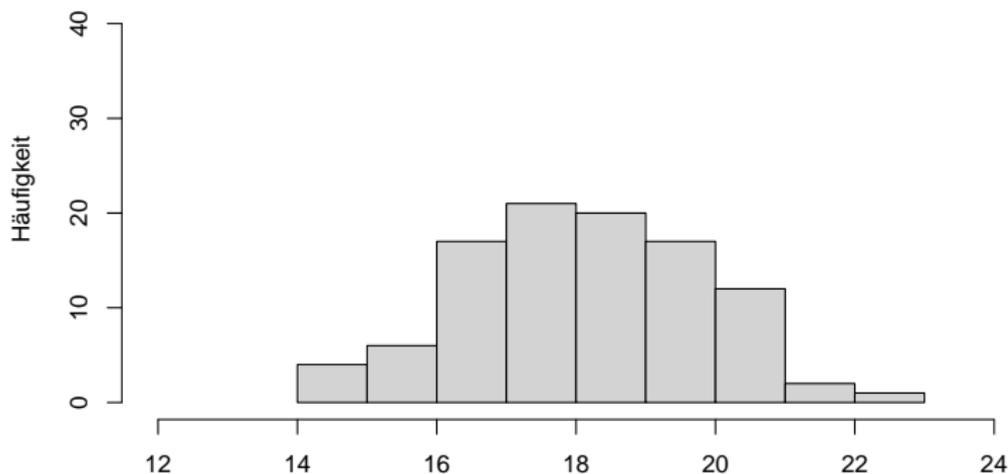
- Die Klassen $[b_{j-1}, b_j], j = 1, \dots, k$ werden als Argument `breaks` festgelegt
- `breaks` ist der atomic vector $c(b_0, b_1, \dots, b_k)$ mit Länge $k + 1$
- Per default benutzt `hist()` eine Modifikation der Sturges Empfehlung $k = \lceil \log_2 n + 1 \rceil$
- `hist()` bietet eine Vielzahl weiterer Spezifikationsmöglichkeiten

```
# Default Histogramm
x      <- D$Pre.BDI           # Datensatz
x_min  <- 12                  # x Achsengrenze (links)
x_max  <- 25                  # x Achsengrenze (rechts)
y_min  <- 0                   # y Achsengrenze (unten)
y_max  <- 45                  # y Achsengrenze (oben)

hist(                          # Histogramm
  x,                            # Datensatz
  xlim = c(x_min, x_max),      # x Achsengrenzen
  ylim = c(y_min, y_max),      # y Achsengrenzen
  ylab = "Häufigkeit",        # y-Achsenbezeichnung
  xlab = "",                   # x-Achsenbezeichnung
  main = "Pre-BDI, R Default" # Titel
)
```

Visualisierung

Pre-BDI, R Default



Histogramm mit hist() und default breaks - Beispiel

Ausgabe der Ergebnisse

```
# Histogrammdaten berechnen, ohne Plot auszugeben
histogram_data <- hist(      # Speichert die Ergebnisse von hist()
  x,                        # Datensatz
  plot = FALSE              # Verhindert die direkte Darstellung des Plots
)

print(histogram_data)      # Ausgabe der Ergebnisse
```

\$breaks

```
[1] 14 15 16 17 18 19 20 21 22 23
```

\$counts

```
[1] 4 6 17 21 20 17 12 2 1
```

\$density

```
[1] 0.04 0.06 0.17 0.21 0.20 0.17 0.12 0.02 0.01
```

\$mids

```
[1] 14.5 15.5 16.5 17.5 18.5 19.5 20.5 21.5 22.5
```

\$xname

```
[1] "x"
```

\$equidist

```
[1] TRUE
```

attr(,"class")

```
[1] "histogram"
```

Histogramme mit alternativen Klassenrößen

Berechnung von Klassenanzahlen und breaks Argument

```
# Histogramm mit gewünschter Klassenbreite
h <- 1 # gewünschte Klassenbreite
b_0 <- min(x) # b_0
b_k <- max(x) # b_k
k <- ceiling((b_k - b_0) / h) # Anzahl der Klassen
b <- seq(b_0, b_k, len = k + 1) # Klassen [b_{j-1}, b_j[

# Excelstandard
n <- length(x) # Anzahl Datenwerte
k <- ceiling(sqrt(n)) # Anzahl der Klassen
b <- seq(b_0, b_k, len = k + 1) # Klassen [b_{j-1}, b_j[
h <- b[2] - b[1] # Klassenbreite

# Sturges
n <- length(x) # Anzahl Datenwerte
k <- ceiling(log2(n)+1) # Anzahl der Klassen
b <- seq(b_0, b_k, len = k + 1) # Klassen [b_{j-1}, b_j[
h <- b[2] - b[1] # Klassenbreite

# Scott
n <- length(x) # Anzahl Datenwerte
S <- sd(x) # Stichprobenstandardabweichung
h <- ceiling(3.49*S/(n^(1/3))) # Klassenbreite
k <- ceiling((b_k - b_0)/h) # Anzahl der Klassen
b <- seq(b_0, b_k, len = k + 1) # Klassen [b_{j-1}, b_j[
```

Berechnung und Visualisierung von Histogrammen mit `hist()`

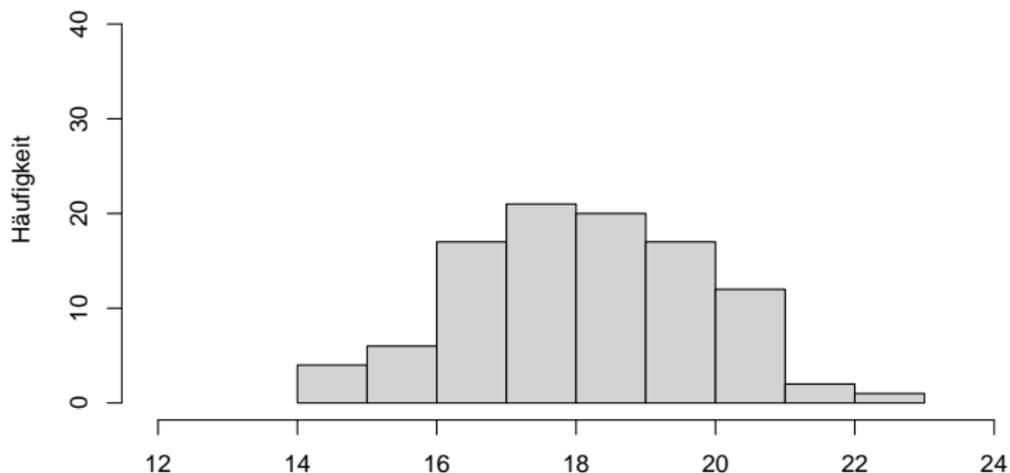
- Die Klassen $[b_{j-1}, b_j], j = 1, \dots, k$, die in der Variable `b` gespeichert sind, werden als Argument mit `breaks` festgelegt
- `breaks` ist der atomic vector $c(b_0, b_1, \dots, b_k)$ mit Länge $k + 1$

```
# Default Histogramm
x      <- D$Pre.BDI                # Datensatz
x_min  <- 12                       # x Achsengrenze (unten)
x_max  <- 25                       # x Achsengrenze (oben)
y_min  <- 0                         # y Achsengrenze (oben)
y_max  <- 30                       # y Achsengrenze (unten)
hist(                                # Histogramm
  x,                                  # Datensatz
  breaks = b,                        # breaks
  xlim   = c(x_min, x_max),          # x Achsengrenzen
  ylim   = c(y_min, y_max),          # y Achsengrenzen
  ylab   = "Häufigkeit",            # y-Achsenbezeichnung
  xlab   = "",                      # x-Achsenbezeichnung
  main   = sprintf("Pre-BDI, k = %.0f, h = %.2f", k, h) # Titel
)
```

Histogramme - Beispiel

Gewünschte Klassenbreite $h := 1$

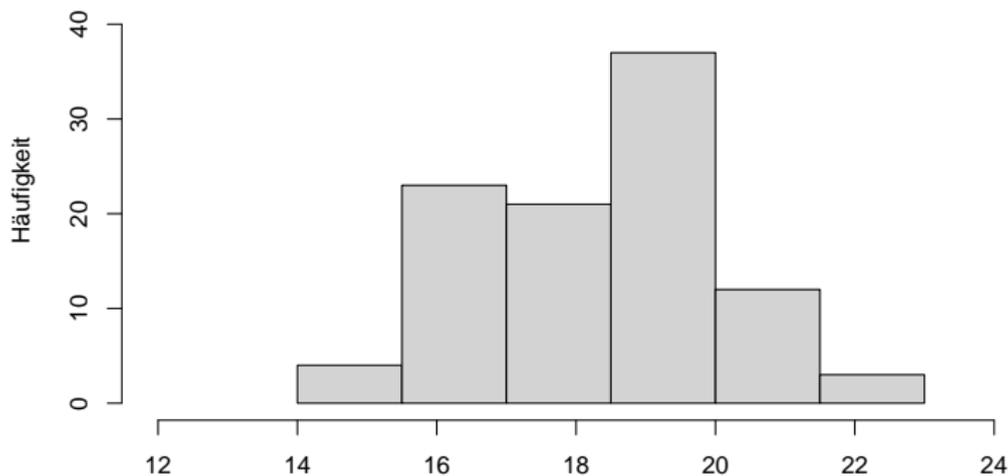
Pre-BDI, $k = 9$, $h = 1.00$



Histogramme - Beispiel

Gewünschte Klassenbreite $h := 1.5$

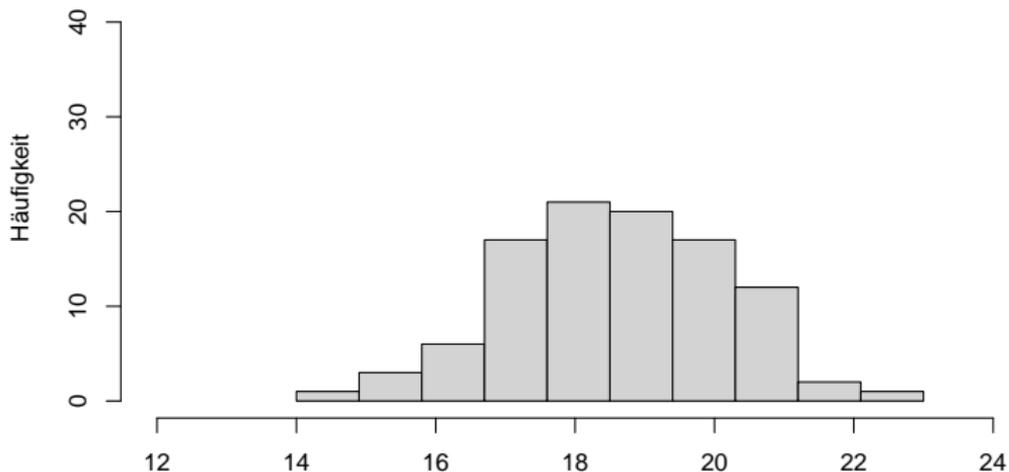
Pre-BDI, $k = 6$, $h = 1.50$



Histogramme - Beispiel

Excelstandard $k := \lceil \sqrt{n} \rceil$

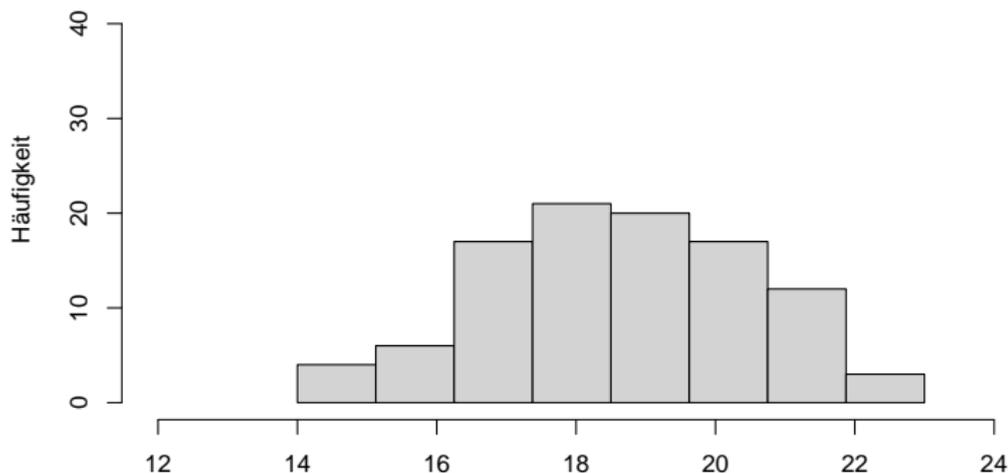
Pre-BDI, $k = 10$, $h = 0.90$



Histogramme - Beispiel

nach Sturges (1926) , $k := \lceil \log_2 n + 1 \rceil$

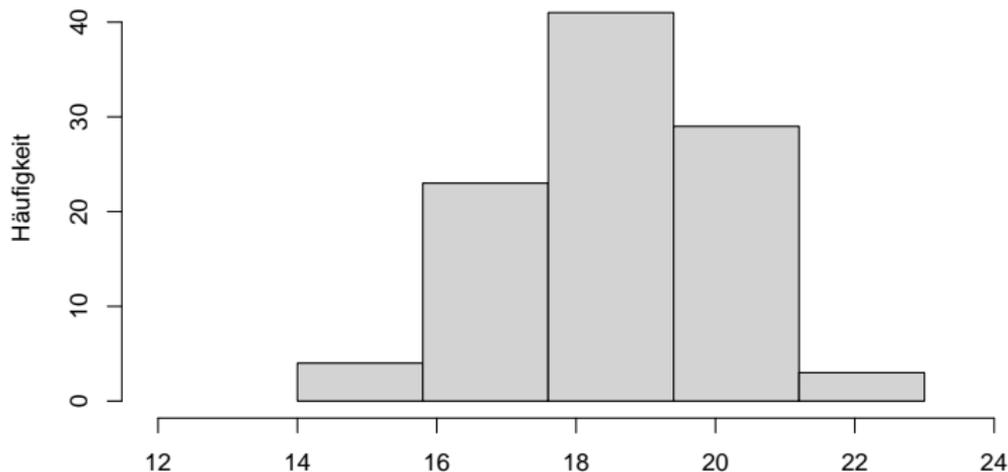
Pre-BDI, k = 8, h = 1.12



Histogramme - Beispiel

nach Scott (1979) , $h := 3.49S_n / \sqrt[3]{n}$

Pre-BDI, k = 5, h = 1.80



Motivation

Beispieldatensatz

Häufigkeitsverteilungen

Histogramme

Programmierübungen und Selbstkontrollfragen

1. Visualisiere die absoluten und relativen Häufigkeitsverteilungen der Post-BDI-Daten des Beispieldatensatzes *psychotherapie_datensatz.csv*. Speichere die erstellten Diagramme als PDF-Dateien im gleichen Verzeichnis, in dem das Skript für diese Aufgabe liegt. Stelle sicher, dass sowohl die Achsentitel als auch die Legenden klar und beschreibend sind.
2. Visualisiere die Häufigkeitsverteilungen der Differenzen zwischen Post- und Pre-BDI-Daten. Speichere die Diagramme in einem Abbildungsverzeichnis und achte darauf, dass die Darstellung für spätere Vergleiche konsistent ist.
3. Visualisiere die Häufigkeitsverteilungen der Differenzen zwischen Post- und Pre-BDI-Daten getrennt nach den experimentellen Bedingungen („Klassisch“ und „Online“). Nutze dein Wissen zu Indizierung in R, um die Daten korrekt zu filtern, und beschrifte die Diagramme und PDF-Dateien entsprechend.
4. Beschreibe die in den vorherigen Aufgaben erstellten Häufigkeitsverteilungen. Gehe dabei auf Unterschiede in Form, Lage und Streuung ein.

5. Erstelle Histogramme der Differenzen zwischen Post- und Pre-BDI-Daten mithilfe verschiedener Methoden zur Bestimmung der Klassengröße. Beschrifte alle Diagramme klar und lege sie im Abbildungsverzeichnis ab.
6. Vergleiche und beschreibe die Unterschiede zwischen den in Aufgabe 5 erstellten Histogrammen. Diskutiere insbesondere, wie die Wahl der Klassengrößen und -anzahlen die Form und die Interpretation der Verteilungen beeinflusst.

1. Definiere die Begriffe der absoluten und relativen Häufigkeitsverteilungen.
2. Definiere den Begriff des Histogramms.
3. Erläutere die Bedeutung der Klassenanzahl für das Erscheinungsbild eines Histogramms.

References

Scott, David W. 1979. "On Optimal and Data-Based Histograms," 6.

Sturges, Herbert A. 1926. "The Choice of a Class Interval." *Journal of the American Statistical Association* 21 (153): 65–66. <https://doi.org/10.1080/01621459.1926.10502161>.