



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

(7) Häufigkeitsverteilungen

Datum	Einheit	Thema
12.10.22	Einführung	(1) Einführung
19.10.22	R Grundlagen	(2) R und RStudio I
26.10.22	R Grundlagen	(2) R und RStudio II
02.11.22	R Grundlagen	(3) Vektoren
09.11.22	R Grundlagen	(4) Matrizen
16.11.22	R Grundlagen	(5) Listen und Dataframes
23.11.22	R Grundlagen	(6) Datenmanagement
30.11.22	Deskriptive Statistik	(7) Häufigkeitsverteilungen I
07.12.22	Deskriptive Statistik	(7) Häufigkeitsverteilungen II
14.12.22	Deskriptive Statistik	(8) Verteilungsfunktionen und Quantile
19.12.22	<i>Leistungsnachweis Teil 1</i>	
21.12.22	Deskriptive Statistik	(9) Maße der zentralen Tendenz
	Weihnachtspause	
04.01.23	Deskriptive Statistik	(10) Maße der Datenvariabilität
11.01.23	Deskriptive Statistik	(12) Anwendungsbeispiel (Deskriptive Statistik)
18.01.23	Inferenzstatistik	(13) Anwendungsbeispiel (Parameterschätzung, Konfidenzintervalle)
25.01.23	Inferenzstatistik	(14) Anwendungsbeispiel (Hypothesentest)
???	<i>Leistungsnachweis Teil 2</i>	

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.

Beispieldatensatz

Häufigkeitsverteilungen

Histogramme

Übungen und Selbstkontrollfragen

Beispieldatensatz

Häufigkeitsverteilungen

Histogramme

Übungen 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
Name		Nr. 2		
<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 den 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 gleichwertig für Sie zutreffend sind, können Sie für Aussage mit der höchsten Zahl ein „Aber Sie bin ich nicht“, da Sie in jeder Gruppe nicht mehr als eine Aussage ankreuzen, die gut für Sie passt. (Die Veränderungen der Schilddrüsenfunktion oder Gruppe 16 (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.) Bestürzungsfühle</p> <p>0 Ich habe viele 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 mal in die Zukunft. 1 Ich sehe mittelmäßig in die Zukunft als blass. 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>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 als sonst. 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 mache 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 umbringen, um meine Probleme zu lösen. 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 früher bewerkstelligen konnte, 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 vermindert. 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>FEARSON PsychCorp</p>	<p>Skizze Seite 1</p>	<p>Skizze Seite 2</p>	<p>Skizze Seite 3</p>	

0 - 8 keine Depression

9 - 13 minimale Depression

14 - 19 leichte Depression

20 - 28 mittelschwere Depression

29 - 63 schwere Depression

Beispiel: Evaluation von Psychotherapieformen bei Depression

Experimentelle Bedingung
(Gruppen von $n = 50$)

Psychotherapie

Klassisch

Pre-BDI



Post-BDI

Online

Pre-BDI



Post-BDI

Beispieldatensatz


Einlesen des Datensatzes mit read.table()

```
pfad_zu_datei = file.path(pfad_zu_Datenordner, "psychotherapie_datensatz.csv")  
# z.B. könnte fpath so aussehen:  
# "/home/belindame_f/ovgu/progr-und-deskr-stat-23/Data/psychotherapie_datensatz.csv"  
  
D = read.table(pfad_zu_datei, sep = ",", header = T)
```

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()



The image shows a screenshot of a data table interface. At the top, there is a toolbar with a grid icon, a close button 'D x', and a 'Filter' button. Below the toolbar is a table with 16 rows and 4 columns. The columns are labeled 'Bedingung', 'Pre BDI', and 'Post BDI'. The first column contains row numbers from 1 to 16. The 'Bedingung' column contains the word 'Klassisch' for all rows. The 'Pre BDI' and 'Post BDI' columns contain numerical values for each 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
16	Klassisch	18	13

Beispieldatensatz

Häufigkeitsverteilungen

Histogramme

Übungen 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

Häufigkeitsverteilungen

Erzeugen der absoluten Häufigkeitsverteilung mit `table()`

Erzeugen der relativen Häufigkeitsverteilung durch Division mit n

```
x = D$Pre.BDI           # Double vector der Pre BDI Werte
n = length(x)           # Anzahl der Datenwerte (100)
H = as.data.frame(table(x)) # absolute Haeufigkeitsverteilung (dataframe)
names(H) = c("a", "h")   # Spaltenbenennung
H$r = H$h/n             # relative Haeufigkeitsverteilung
```

a	h	r
14	1	0.01
15	3	0.03
16	6	0.06
17	17	0.17
18	21	0.21
19	20	0.20
20	17	0.17
21	12	0.12
22	2	0.02
23	1	0.01

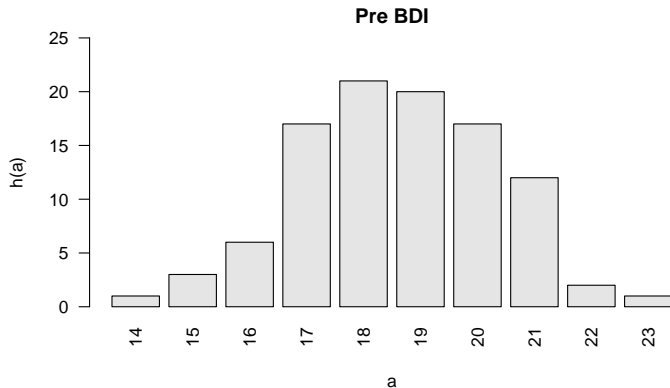
Visualisierung der absoluten Häufigkeitsverteilung mit `barplot()`

```
h           = H$h           # h(a) Werte
names(h)    = H$a           # barplot braucht a Werte als names
dev.new()   # Abbildungsinitialisierung
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
  main     = "Pre BDI"    # Titel
)
```

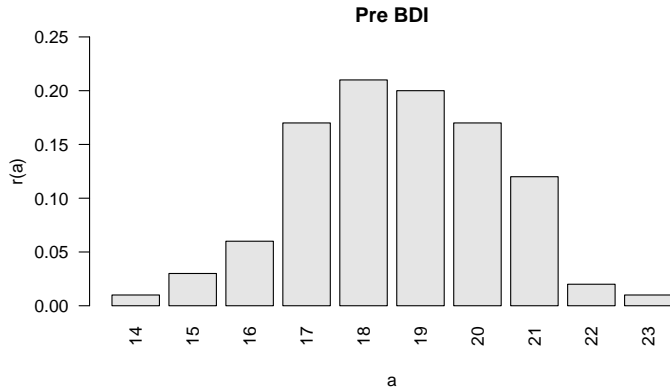
Speichern von Abbildungen mit `dev.copy2pdf()`

```
dev.copy2pdf(                                     # PDF Kopierfunktion
  file      = file.path(pfad_zu_abbildungen_ordner, "pds_7_ha_prebdi.pdf"), # Dateiname
  width     = 7,                                   # Breite (inch)
  height    = 4,                                   # Höhe (inch)
)
```

Absolute Häufigkeitsverteilung aller Pre-BDI Werte



Relative Häufigkeitsverteilung aller Pre-BDI Werte



Beispieldatensatz

Häufigkeitsverteilungen

Histogramme

Übungen 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

Breite $d_j = b_j - b_{j-1}$

Höhe $h(a)$ oder $r(a)$ 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.
- Mit der Aufrundungsfunktion $\lceil \cdot \rceil$ sind konventionelle Werte für k

$k := \lceil (b_k - b_0)h \rceil$ h ist die gewünschte Klassenbreite

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

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

$k := 3.49 S_n / \sqrt[3]{n}$ Min. MSE Dichteschätzung bei Normalverteilung (Scott, 1979)

Berechnung und Visualisierung von Histogrammen mit hist()

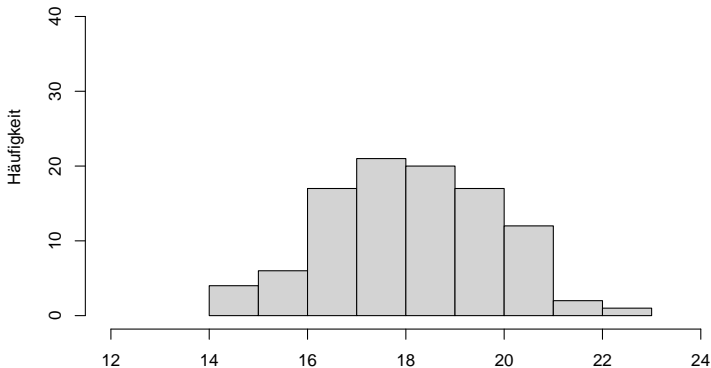
- Die Klassen $[b_{j-1}, b_j[, j = 1, \dots, k$ werden als Argument breaks festgelegt
- breaks ist der atomic vector $c(b_0, b_1, \dots, b_k)$ mit Länge $k + 1$
- Per default benutzt hist() eine Modifikation der Sturges Empfehlung $k = \lceil \log_2 n + 1 \rceil$
- hist() bietet eine Vielzahl weiterer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

```
# Default Histogramm
x      = D$Pre.BDI           # Datensatz
x_min  = 12                  # x Achsengrenze (unten)
x_max  = 25                  # x Achsengrenze (oben)
y_min  = 0                   # y Achsengrenze (oben)
y_max  = 30                  # y Achsengrenze (unten)
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
)
```

Histogramme

Berechnung und Visualisierung von Histogrammen mit `hist()`

Pre-BDI, R Default



Berechnung von Klassenanzahlen und breaks Argument

```
# Histogramm mit gewuenschter Klassenbreite
h = 1                                # gewuenschte 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, by = h)             # 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)           # Klassen [b_{j-1}, b_j[
h = b[2] - b[1]                       # Klassenbreite

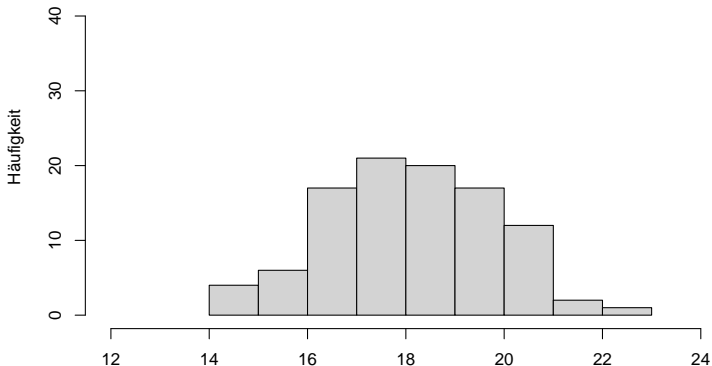
# Sturges
n = length(x)                         # Anzahl Datenwerte
k = ceiling(log2(n)+1)                # Anzahl der Klassen
b = seq(b_0, b_k, len = k)           # Klassen [b_{j-1}, b_j[
h = b[2] - b[1]                       # Klassenbreite

# 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)           # Klassen [b_{j-1}, b_j[
```

Histogramme

Gewünschte Klassenbreite $h := 1$

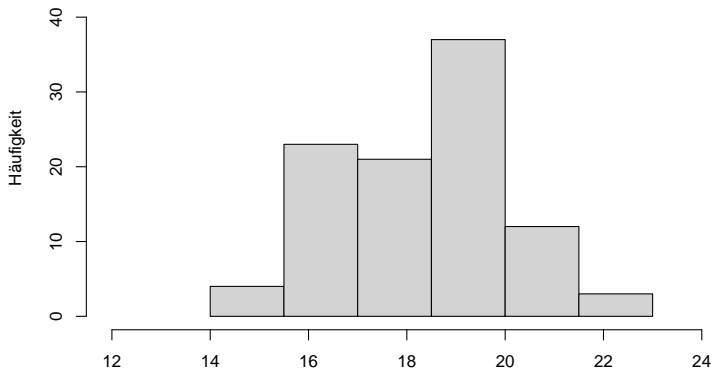
Pre-BDI, $k = 9$, $h = 1.00$



Histogramme

Gewünschte Klassenbreite $h := 1.5$

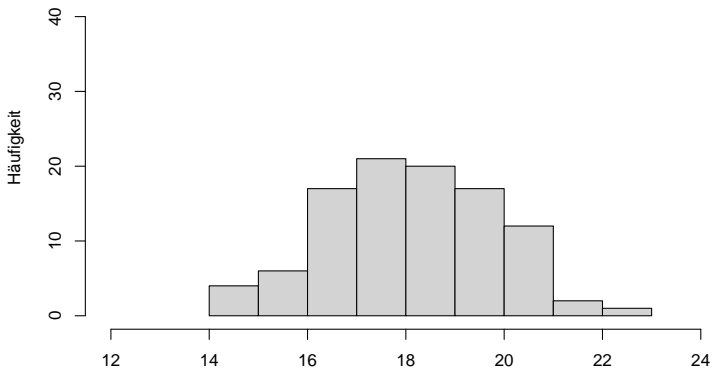
Pre-BDI, $k = 6$, $h = 1.50$



Histogramme

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

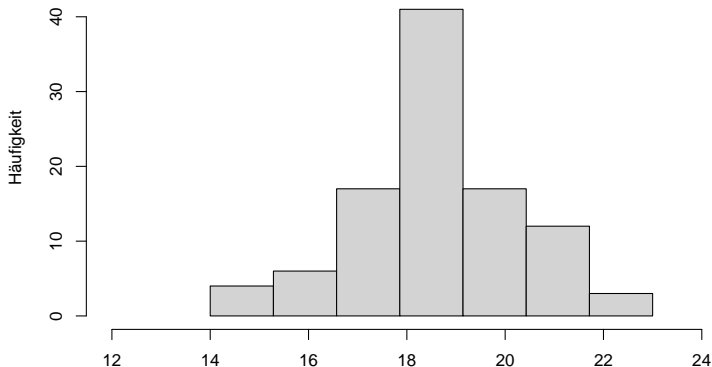
Pre-BDI, $k = 10$, $h = 1.00$



Histogramme

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

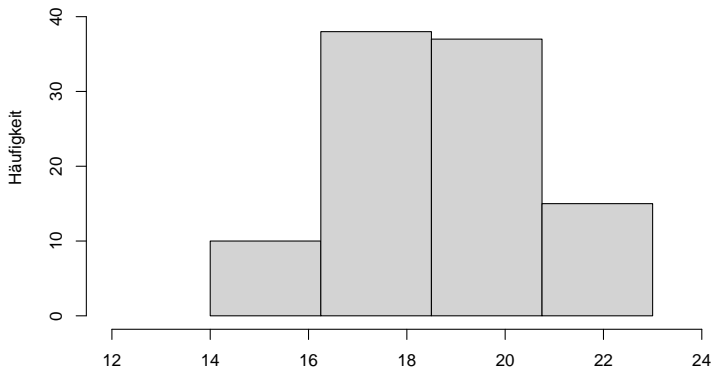
Pre-BDI, k = 8, h = 1.29



Histogramme

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

Pre-BDI, k = 5, h = 2.25



Beispieldatensatz

Häufigkeitsverteilungen

Histogramme

Übungen und Selbstkontrollfragen

Übungen und Selbstkontrollfragen

1. Definieren Sie die Begriffe der absoluten und relativen Häufigkeitsverteilungen.
2. Visualisieren Sie die Häufigkeitsverteilungen der Post-BDI Daten.
3. Visualisieren Sie die Häufigkeitsverteilungen der Differenzen von Post- und Pre-BDI Daten.
4. Visualisieren Sie die Häufigkeitsverteilungen der Differenzen von Post- und Pre-BDI Daten getrennt nach den experimentellen Bedingungen "Klassisch" und "Online". Nutzen Sie dazu Ihr Wissen zu den Prinzipien der Indizierung in R.
5. Beschreiben Sie die in der vorherigen Aufgabe erstellten Häufigkeitsverteilungen.
6. Definieren Sie den Begriff des Histogramms.
7. Erläutern Sie die Bedeutung der Klassenanzahl für das Erscheinungsbild eines Histogramms.
8. Visualisieren Sie Histogramme der Daten wie in Aufgabe 4. mit einer Klassenbreite von 3, dem Excelstandard, der Sturges Klassenanzahl und der Scott Klassenanzahl.
9. Beschreiben Sie die in der vorherigen Aufgabe erstellten Histogramme.

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>.