

Kapitel 1: Deskriptive Statistik

Grafiken	1
Statistische Kennwerte	5
z-Standardisierung	6
Literatur	7

Grafiken

Mit Hilfe von R lassen sich eine Vielzahl unterschiedlicher Grafiken für unterschiedliche Zwecke erstellen. Wir besprechen hier die zwei in Kapitel 1.1 thematisierten Varianten Histogramm und Kreisdiagramm. Lesen Sie zunächst den Datensatz ein. Hierfür muss das bereits per Standardeinstellung installierte Paket `foreign` mit der Funktion `library()` geladen werden

```
library(foreign)
beispiel <- read.spss("Beispieldatensatz.sav",
                    to.data.frame = TRUE)
```

Für das Erstellen von Grafiken wird das Paket `ggplot2` verwendet. Installieren und laden Sie es:

```
install.packages("ggplot2")
library(ggplot2)
```

Histogramm

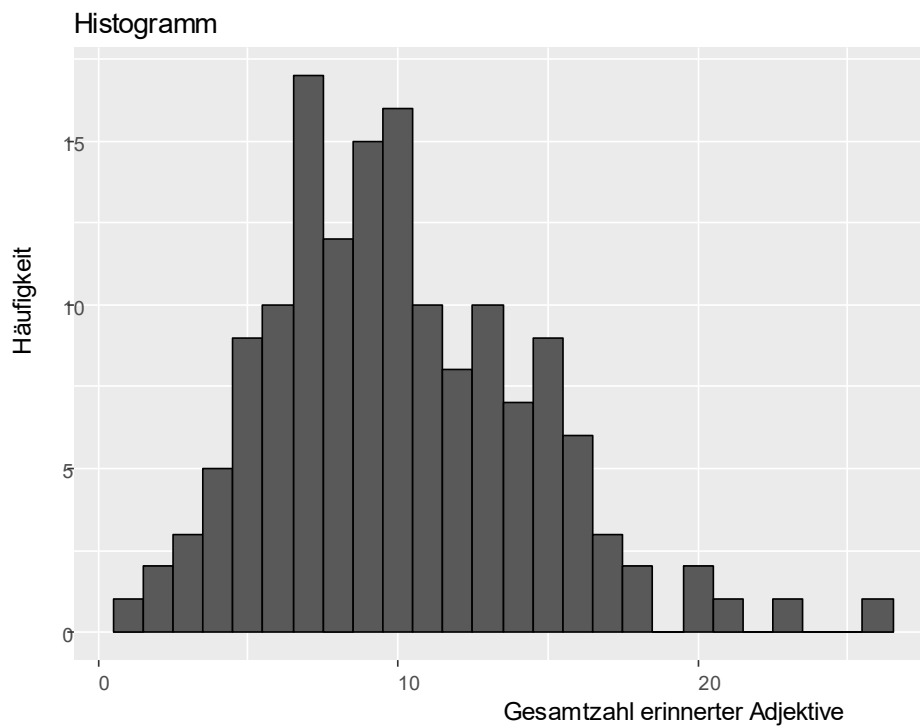
Um eine Häufigkeitsverteilung von der Gesamtzahl erinnerter Adjektive zu erstellen, verwenden Sie die Funktion `ggplot()` des Pakets `ggplot2`:

```
hist <- ggplot(beispiel, aes(x=ges)) +
  geom_histogram(binwidth = 1, color = "black") +
  labs(title = "Histogramm",
       x = "Gesamtzahl erinnerter Adjektive",
       y = "Häufigkeit")
```

Stellen Sie sich jede Zeile des Codes wie eine Schicht vor, aus der sich das Diagramm am Ende zusammensetzt. Diese unterschiedlichen Schichten werden jeweils mit einem „+“ verbunden. In der ersten Zeile spezifizieren Sie zunächst den Datensatz und danach welche Variable auf welcher Achse abgebildet werden soll. In der zweiten Zeile wurde die Art des Diagramms angegeben, die Intervallbreite und die Rahmenfarbe der Balken. Als nächstes wurde das Diagramm entsprechend beschriftet. Geben Sie `hist` in die Konsole ein, um sich nun die Grafik anzeigen zu lassen.

<https://lehrbuch-psychologie.springer.com/content/zusatztexte-mit-anleitungen-zu-spss-r-und-gpower-sowie-datensätze>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 1*, 5. Auflage. Heidelberg: Springer.

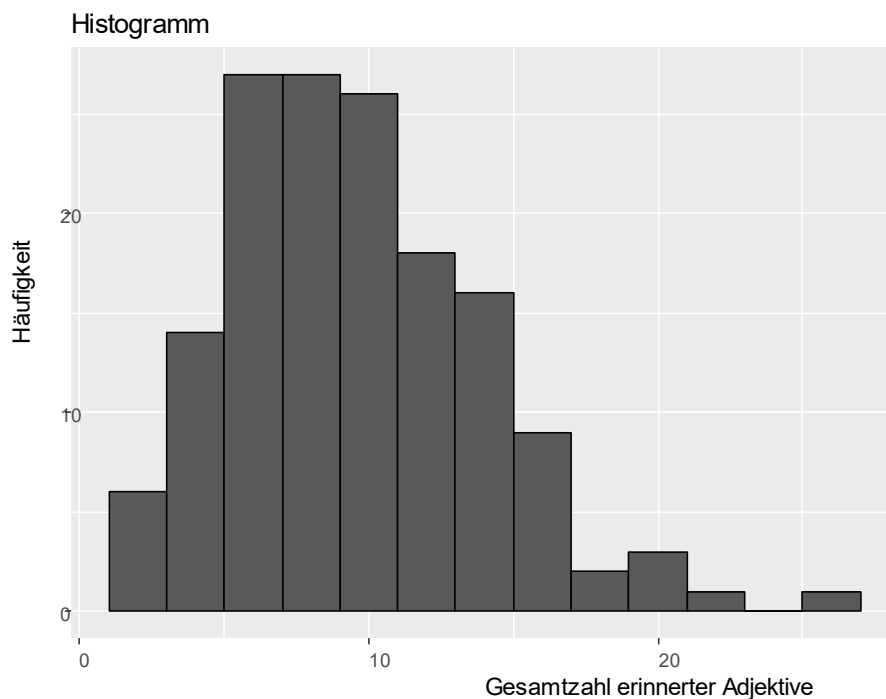


Dieser Darstellung können Sie entnehmen, dass nur eine Person lediglich ein Adjektiv erinnert hat. Ebenfalls hat nur eine Person mehr als 25 Worte erinnert.

Das Paket `ggplot2` bietet Ihnen vielfältige Optionen, die optische Darstellung zu verändern. Beispielweise können Sie die Intervallbreite verändern, also die Anzahl der Einheiten, die in einem Balken dargestellt werden. Ändern Sie dafür lediglich die Zahl mit dem Argument `binwidth`. Wir verändern im Beispiel die Intervallbreite auf 2. Entsprechend weniger Balken werden im resultierenden Diagramm dargestellt.

<https://lehrbuch-psychologie.springer.com/content/zusatztexte-mit-anleitungen-zu-spss-r-und-gpower-sowie-datensätze>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 1*, 5. Auflage. Heidelberg: Springer.



Sie können mit Hilfe des Pakets `ggplot2` das Aussehen des Diagramms stark verändern und die Interpretierbarkeit für Ihre speziellen Zwecke anpassen. Beispielsweise lassen sich die Achsen mit den verwendeten Skalenformaten ändern, die Beschriftungen des Diagramms oder die Farben der Balken. Probieren Sie es selber aus!

Kreisdiagramm

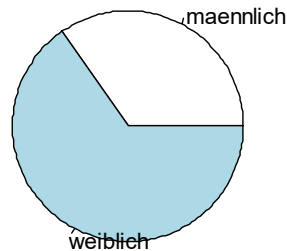
Die Ausnahme bilden Kreisdiagramme, weil sich das Paket `ggplot2` hierfür nicht eignet. Verwenden Sie stattdessen die Funktion `pie()`. Geben Sie die Variable „Geschlecht“ mit der Funktion `summary()` ein. Mit dem Argument `main` können Sie einen Diagrammtitel angeben.

```
pie(summary(beispiel$sex),  
     main = "Kreisdiagramm Anzahl von Geschlecht")
```

<https://lehrbuch-psychologie.springer.com/content/zusatztexte-mit-anleitungen-zu-spss-r-und-gpower-sowie-datensätze>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 1*, 5. Auflage. Heidelberg: Springer.

Kreisdiagramm Anzahl vo



Das entstandene Diagramm können Sie wieder vielfältig selber gestalten, z.B. in dem Sie sich anzeigen lassen, wie viel Prozent der Stichprobe durch jedes Stück des Kreises repräsentiert wird. Dafür gehen Sie folgendermaßen vor. Weisen Sie zunächst `summary(beispiel$sex)` einem Objekt zu.

```
sex <- summary(beispiel$sex)
```

Weisen Sie einem weiteren Objekt die beiden prozentualen Anteile der Geschlechter zu. In den eckigen Klammern geben Sie an, auf welche Zelle Sie zugreifen möchten. Dann teilen Sie diese Zelle durch die Gesamtanzahl, die Sie mit der Funktion `sum()` berechnen können und multiplizieren Sie diese Ergebnis mit 100. Verwenden Sie die Funktion `round()`, um ganze Zahlen zu erhalten.

```
percent <- c(round(sex[1]/sum(sex)*100),
             round(sex[2]/sum(sex)*100))
```

Erstellen Sie nun ein weiteres Objekt, mit dem Sie das Kreisdiagramm beschriften werden. Zunächst geben Sie die beiden Geschlechter an, zuerst „männlich“ danach „weiblich“. Mit der Funktion `paste()` werden nun die Beschriftungen entsprechend zusammengefügt.

```
names <- c("männlich", "weiblich")
names <- paste(names, percent)
names <- paste(names, "%", sep = "")
```

Die obigen drei Zeilen lassen sich folgendermaßen übersetzen: Zunächst wird die Beschriftung erstellt. In der zweiten Zeile soll der Beschriftung der dazugehörige Prozentwert folgen, wobei auf die Reihenfolge zu achten ist, bei der Erstellung der Variablen „percent“ und „names“. In der letzten Zeile soll nach der Prozentangabe ein Prozentzeichen folgen und mit dem Argument `sep` wird angegeben, dass dazwischen kein Leerzeichen sein soll. Beachten Sie, dass die Variable „names“ sukzessive überschrieben wird.

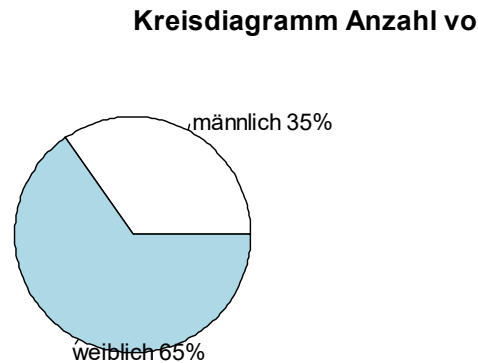
Mit dem Argument `labels` können Sie nun die Variable „names“ in dem Kreisdiagramm eingeben.

<https://lehrbuch-psychologie.springer.com/content/zusatztexte-mit-anleitungen-zu-spss-r-und-gpower-sowie-datensätze>

Aus: Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 1*, 5. Auflage. Heidelberg: Springer.

```
pie(summary(beispiel$sex),  
     labels = names,  
     main = "Kreisdiagramm Anzahl von Geschlecht")
```

Sie erhalten folgendes Diagramm:



Statistische Kennwerte

Die folgenden Ausführungen beschreiben die Berechnung der in Kapitel 1.3 besprochenen statistischen Kennwerte. Dabei handelt es sich um drei Maße der zentralen Tendenz (Modus, Median, Mittelwert) sowie um drei Dispersionsmaße (Variationsbreite, Varianz und Streuung).

Die Berechnung der gewünschten statistischen Kennwerte mit R für die Variable „Gesamtzahl erinnelter Adjektive“ verläuft über die Funktion `describe()` des Pakets `psych`.

```
install.packages("psych")  
library(psych)  
describe(beispiel$ges)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
x1	1	150	10.07	4.37	10	9.83	4.45	1	26	25	0.65	0.59	0.36

Die Varianz erhalten Sie, indem Sie die Standardabweichung quadrieren oder Sie können auch die Funktion `var()` verwenden. Für den Modalwert verwenden Sie die Funktion `Mode()` des Pakets `DescTools`.

```
install.packages("DescTools")  
library(DescTools)  
Mode(beispiel$ges)  
[1] 7
```

<https://lehrbuch-psychologie.springer.com/content/zusatztexte-mit-anleitungen-zu-spss-r-und-gpower-sowie-datensätze>

Aus: Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 1*, 5. Auflage. Heidelberg: Springer.

Um die Bedeutung der einzelnen Werte zu verstehen, können Sie `??Funktionsname` in der Konsole eingeben und darüber hinaus verweisen wir auf Bortz und Schuster (2010), sofern sie nicht selbsterklärend sind.

Im Mittel haben die 150 Versuchspersonen ungefähr 10 Wörter erinnert. Dieser Wert bildet gleichzeitig auch den Median, also den Wert, der die Verteilung in zwei Hälften teilt. Am häufigsten wurden sieben Worte erinnert. Die Dispersionsmaße Standardabweichung und Varianz sind zunächst unanschaulich und eine Interpretation darüber fällt schwer, ob die gefundenen Werte von 4,37 für die Standardabweichung bzw. 19,08 für die Varianz eher als groß oder klein zu gelten haben. Die Spannweite oder Variationsbreite dagegen ist sehr anschaulich zu interpretieren: Die Person mit der besten Erinnerungsleistung hat 25 Wörter mehr erinnert als die Person mit der geringsten Erinnerungsleistung.

z-Standardisierung

Die z-Standardisierung einer oder mehrerer Variable(n) erledigt R über die Funktion `scale()`. Wir führen die Berechnung beispielhaft mit der Variable „Gesamtzahl erinnertes Adjektive“ durch.

```
zges <- scale(beispiel$ges)
```

Geben Sie `zges` in die Konsole ein. Sie erhalten den z-Wert jeder Person. Ihr Datensatz bleibt jedoch unverändert. Wenn Sie diese neue Variable dem Datensatz hinzufügen möchten geben Sie folgendes ein.

```
beispiel <- cbind(beispiel, zges)
```

Geben Sie `view(beispiel)` in die Konsole ein oder klicken Sie oben rechts im Fenster „Environment“ auf „beispiel“, um sich den Datensatz anzuschauen. Die Variable `zges` befindet sich in der letzten Spalte rechts.

Jede Person hat einen Wert auf dieser neuen Variablen. Wenn Sie sich von dieser neuen Variable Mittelwert und Standardabweichung über `describe(zges)` anzeigen lassen, sehen Sie, dass diese wie in Kapitel 1.4 dargestellt bei 0 bzw. 1 liegen. Ein negativer z-Wert zeigt also an, dass die betreffende Person insgesamt weniger Adjektive als der Durchschnitt der Stichprobe erinnert hat. Ein positiver z-Wert steht für eine deskriptiv überdurchschnittliche Gedächtnisleistung.

```
vars  n mean sd median trimmed mad  min  max range skew kurtosis  se
x1    1 150  0  1 -0.02  -0.05 1.02 -2.08 3.65  5.72 0.65    0.59 0.08
```

Hinweis: Wenn Sie den Mittelwert über `mean(zges)` berechnen, werden Sie `-1.790008e-16` erhalten. Dies hängt mit der Art und Weise zusammen, wie R den Mittelwert berechnet, weshalb es zu Ungenauigkeiten kommen kann.

<https://lehrbuch-psychologie.springer.com/content/zusatztexte-mit-anleitungen-zu-spss-r-und-gpower-sowie-datensätze>

Aus: Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 1*, 5. Auflage. Heidelberg: Springer.

R-Ergänzungen

Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 1* (5. Auflage). Heidelberg: Springer.

Literatur

Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Heidelberg: Springer.

<https://lehrbuch-psychologie.springer.com/content/zusatztexte-mit-anleitungen-zu-spss-r-und-gpower-sowie-datensätze>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 1*, 5. Auflage. Heidelberg: Springer.