

Bevor wir uns mit den in Kap. 1 eingeführten Fragestellungen beschäftigen, brauchen wir noch ein Werkzeug, um die Datensätze später wirklich auswerten zu können. Sicher lässt sich das in den meisten Fällen auch per Hand durchführen. Aber insbesondere für große Datensätze und spätestens, wenn Zufallsexperimente simuliert werden sollen, stoßen wir hier an unsere Grenzen. Daher soll im Folgenden eine Statistiksoftware eingeführt werden, die wir zur Auswertung und Simulation nutzen werden. Für alle hier im Buch vorgestellten Methoden werden auch immer die jeweils auszuführenden Schritte in der Software aufgeführt. Wir haben uns für die freie Software R, siehe R Core Team (2013), entschieden. Neben der freien Verfügbarkeit hat die Software R den Vorteil, dass sie immer häufiger verwendet wird und dadurch immer mehr Routinen im Internet dazu zu finden sind. Hier wird daher nur diese Software benutzt. Damit haben Sie die Möglichkeit, nach der Installation von R alle Verfahren direkt an den zur Verfügung gestellten Datensätzen selbst auszuprobieren. Natürlich kann aber für die vorgestellten Verfahren auch jede andere Statistiksoftware benutzt werden. In diesem Kapitel wird nun eine kurze Einführung in die grundlegende Bedienung der Software R gegeben. Eine ausführliche Einführung geben zum Beispiel Groß (2010), Wollschläger (2012), Verzani (2004) oder Venables und Ripley (2003), tiefere Einblicke bieten Ligges (2009) und Faraway (2006). Die Bücher DIALEKT-Projekt (2002) und Schlittgen (2005) stellen eine Oberfläche der hier behandelten Software vor, die noch einfacher zu bedienen ist. Allerdings hat diese Oberfläche einige Nachteile, weshalb hier die direkte Anwendung der Software erlernt werden soll.

2.1 Herunterladen der freien Software R

Unter <http://cran.r-project.org/> finden Sie die neuesten R-Versionen für Windows, Mac und Linux. Windows-Nutzer klicken nach der Auswahl des Betriebssystems auf base. Hier finden Sie auch Installationsanleitungen. Nun kann R gestartet werden. Dazu ist es sinnvoll R mit einem eigenen Arbeitsverzeichnis zu verknüpfen, da ansonsten von R erstell-

te Dateien im Programm-Ordner abgelegt werden. Unter dem Menü-Punkt Hilfe, findet sich unter anderem eine ausführliche Einführung in R auf Englisch. Kennen Sie den Namen der R-Funktion, können Sie eine Hilfe zu der entsprechenden Funktion direkt über die Eingabe *?Name der Funktion* aufrufen. Hier finden Sie auch die Liste der zu übergebenden Argumente und die Ausgaben sowie Beispiele für die Anwendung der Funktion.

2.2 Grafische Darstellungen mit R

Eindimensionale Grafiken können einfach mit der Funktion `plot` erzeugt werden. Dies wird hier anhand der Gaußschen Normalverteilung, die auf dem 10 DM-Schein abgebildet war, demonstriert. Die Dichte dieser Normalverteilung mit Parameter $\text{mean} = \mu$ und $\text{sd} = \sigma$ ist eine Funktion von \mathbb{R} nach \mathbb{R} und ist gegeben durch

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\},$$

mehr dazu finden Sie auch im Abschn. 12.2. Möchte man zum Beispiel die Dichte der sogenannten Standardnormalverteilung ($\mu = 0, \sigma = 1$) im Intervall $[-4, 4]$ darstellen, so erzeugt man zuerst einen Vektor, der z. B. x genannt werden kann, der in aufsteigender Reihenfolge Werte aus dem Intervall $[-4, 4]$ enthält und damit das Intervall $[-4, 4]$ rastert. Je feiner die Rasterung, desto glatter wird die Funktion. Anschließend wendet man die Funktion, die die Dichte der Standardnormalverteilung liefert, `dnorm(x, mean=0, sd=1)`, auf x an.

```
> x<-seq(-4,4,0.1)
> plot(x, dnorm(x, mean=0, sd=1), type="l", ylab="f(x)",
+ main="Dichte der Standardnormalverteilung f(x)")
```

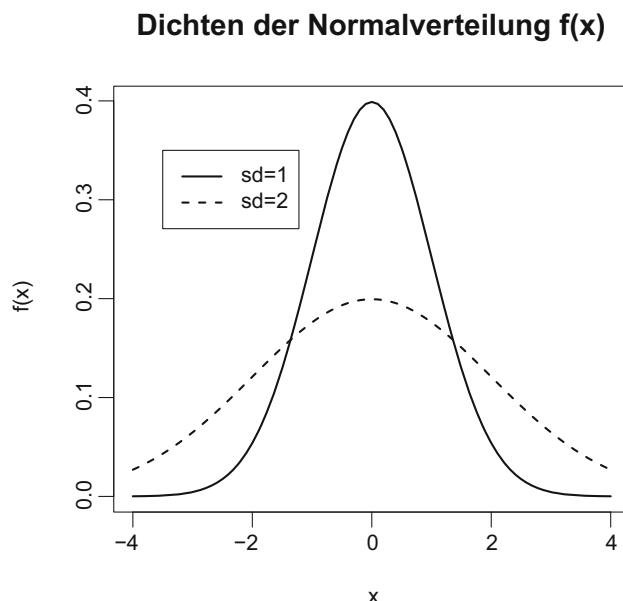
Statt `dnorm(x, mean=0, sd=1)` kann auch `dnorm(x)` geschrieben werden, da `mean=0` und `sd=1` die Voreinstellungen sind. Lässt man `type="l"` weg, so werden nur Punkte $(x, f(x))$ gezeichnet. Das Argument `ylab="f(x)"` ergibt eine Beschriftung für die y -Achse und mit `main="..."` wird die Überschrift bestimmt.

Will man mehrere Dichten von Normalverteilungen in einer Grafik darstellen, können die weiteren Dichten mit `lines` hinzugefügt werden:

```
> x<-seq(-4,4,0.1)
> plot(x, dnorm(x, mean=0, sd=1), type="l", ylab="f(x)",
+ main="Dichten der Normalverteilung f(x)")
> lines(x, dnorm(x, mean=0, sd=2), lty=2)
> legend(-3.5, 0.35, legend=c("sd=1", "sd=2"), lty=c(1, 2))
```

Damit erhalten wir Abb. 2.1. Mit dem Argument `lty` wird der Linientyp bestimmt, z. B. `lty=1` für eine durchgezogene Linie, `lty=2` für eine gestrichelte Linie. Eine Legende kann mit der R-Funktion `legend` hinzugefügt werden. Grundsätzlich können, wie

Abb. 2.1 Zwei Dichten der Normalverteilung



schon oben erwähnt, die Argumente und die Ausgaben einer R-Funktion mit dem Befehl `?Funktionsname` abgefragt werden. Mit

```
> ?legend
```

werden zum Beispiel die Argumente der R-Funktion `legend` angezeigt.

2.3 Übungsaufgaben

Übung 2.1 In R können die Logarithmen zu verschiedenen Basen mit der Funktion `log` berechnet werden, wobei im Argument `base` die Basis festgelegt wird. Plotten Sie die Logarithmusfunktion zu den Basen 2, 3, e (in `R exp(1)`) und 10. Stellen Sie alle vier Logarithmusfunktionen in einer Grafik dar. Wählen Sie dazu einen geeigneten Bereich für die x -Achse.

Übung 2.2 Stellen Sie die folgenden Dichten der Normalverteilung jeweils in einer Grafik dar. Wählen Sie dabei geeignete Bereiche für die x -Achse:

1. Die Dichte der Normalverteilung mit $\mu = \text{mean} = 1$, $\sigma = \text{sd} = 1$ und die Dichte der Normalverteilung mit $\mu = \text{mean} = 0$, $\sigma = \text{sd} = 1$ in einer Grafik.
2. Die Dichte der Normalverteilung mit $\mu = \text{mean} = 1$, $\sigma = \text{sd} = 1$ und die Dichte der Normalverteilung mit $\mu = \text{mean} = 2$, $\sigma = \text{sd} = 1$ in einer Grafik.

3. Die Dichte der Normalverteilung mit $\mu = \text{mean} = 1$, $\sigma = \text{sd} = 1$ und die Dichte der Normalverteilung mit $\mu = \text{mean} = 1$, $\sigma = \text{sd} = 2$ in einer Grafik.

Welche Schlüsse ziehen Sie daraus über die Auswirkung der Änderung der beiden Parameter?

Literatur

- DIALEKT-Projekt (2002). Statistik interaktiv. Deskriptive Statistik. Springer, Berlin.
- Faraway, J.J. (2006). Extending the linear model with R. Chapman & Hall/ CRC, Boca Raton.
- Groß, J. (2010). Grundlegende Statistik mit R. Eine anwendungsorientierte Einführung in die Verwendung der Statistik Software R. Vieweg + Teubner, Wiesbaden.
- Ligges, U. (2009). Programmieren mit R. Springer, Heidelberg.
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Schlittgen, R. (2005). Das Statistiklabor. Einführung und Benutzerhandbuch. Springer, Berlin.
- Venables, W.N. und Ripley, B.D. (2003). Modern applied statistics with S. Springer, New York.
- Verzani, J. (2004). Using R for introductory statistics. Chapman & Hall /CRC, London.
- Wollschläger, D. (2012). Grundlagen der Datenanalyse mit R. Eine anwendungsorientierte Einführung. Springer, Heidelberg.

Stochastik in den Ingenieurwissenschaften

Eine Einführung mit R

Müller, C.; Denecke, L.

2013, XVI, 273 S. 91 Abb., 12 Abb. in Farbe., Softcover

ISBN: 978-3-642-38959-7