IN DIESEM KAPITEL

- Variablen und Messniveaus verstehen
- Deskriptive und inferentielle Statistik einordnen
- Forschungsdesigns unterscheiden

Kapitel 1
Statistik? Ich dachte, es geht um Psychologie!

Wenn wir unseren Erstsemesterstudierenden in ihre anfangs neugierigen und begeisterten Gesichter blicken und ihnen dann mitteilen, dass Statistik ein wesentlicher Bestandteil ihres Studiengangs ist, reagiert etwa die Hälfte von ihnen richtig erschrocken. »Wir wollen Psychologie studieren und nicht Statistik«, rufen sie. Wahrscheinlich dachten sie, während ihres Studiums verwirrten Leuten zu helfen, die sich auf die Couch legen und über ihre Mütter erzählen. Wir erklären ihnen dann, dass die Statistik zu allen Psychologievorlesungen im Bachelor- und Masterstudium gehört und dass sie ohne Statistik nicht auskommen werden, wenn sie vorhaben, psychologisch zu arbeiten. Dann jammern sie: »Aber ich bin doch kein Mathematiker! Ich interessiere mich für Menschen und ihr Verhalten.« Wir erwarten nicht, dass unsere Studierenden Mathematikerinnen sind oder werden. Wenn Sie dieses Buch einmal grob durchblättern, werden Sie sofort sehen, dass es (fast) keine furchterregenden Gleichungen enthält. Heute gibt es zum Glück Computerprogramme wie beispielsweise SPSS, die Programmiersprache R oder die Software RStudio, die die komplizierten Berechnungen für uns erledigen können.

Psychologie ist eine wissenschaftliche Disziplin. Wenn Sie mehr über Menschen lernen wollen, müssen Sie objektiv Informationen sammeln, diese zusammenfassen und analysieren. Nur so können Sie Informationen interpretieren und ihnen eine theoretische wie praktische Bedeutung geben. Die Zusammenfassung und Analyse von Informationen ist aber nichts anderes als Statistik! Statistik ist also ein grundlegender Bestandteil der Psychologie.

Dieses Buch möchte Ihnen einen Überblick über die wichtigsten statistischen Konzepte geben, denen Sie im Bachelorstudium Psychologie begegnen werden. Sie werden auch Verweise auf Kapitel finden, in denen Sie mehr zu den jeweiligen Themen erfahren.

Machen Sie sich ein Bild von Ihren Variablen

Für alle quantitativen Forschungen in der Psychologie müssen Informationen (sogenannte *Daten*) gesammelt werden. Diese können durch Zahlen dargestellt werden. Beispielsweise lassen sich Depressionsstufen durch Depressionspotenziale darstellen, die man anhand eines Fragebogens ermittelt. Auch das Geschlecht einer Person kann durch eine Zahl dargestellt werden (zum Beispiel 1 für männlich, 2 für weiblich, 3 für divers). Die von Ihnen gemessenen Eigenschaften werden als *Variablen* bezeichnet, weil sie variieren! Sie können für eine Person über die Zeit variieren (Depressionspotenziale können über die Lebensdauer einer Person variieren) oder zwischen verschiedenen Personen (Personen können beispielsweise als männlich, weiblich oder divers klassifiziert werden, aber nach erfolgter Klassifikation ändert sich diese Variable im Allgemeinen nicht mehr).

Mit den Variablen in einer Datenmenge sind verschiedene Namen und Eigenschaften verknüpft, mit denen Sie sich vertraut machen sollten. Variablen können stetig oder diskret sein, sie können unterschiedliche Maße haben und sie können unabhängig oder abhängig sein. Wir werden in Kapitel 2 genauer darauf eingehen. Anfänglich werden all diese Begriffe recht verwirrend sein, aber Sie müssen sie unbedingt gut verstehen, weil alle wichtigen statistischen Analysen davon Gebrauch machen. Beispielsweise kann es sinnvoll sein, ein mittleres Depressionspotenzial von 32,4 für eine bestimmte Teilnehmergruppe zu nennen, während ein mittleres Geschlechtsmaß von 1,6 für dieselbe Gruppe wenig sinnvoll ist! Um die Mittelwerte wird es in Kapitel 4 gehen.

Variablen sind *diskret*, wenn sie Kategorien abbilden (zum Beispiel männlich, weiblich oder divers). Man nennt sie *stetig*, wenn die Werte überall innerhalb eines vorgegebenen Wertebereichs liegen können. Beispielsweise können Depressionspotenziale Zahlenwerte zwischen 0 und 63 annehmen, wenn sie anhand des Beck-Depressions-Inventars gemessen werden. Außerdem unterscheiden sich Variablen in ihren Messeigenschaften. Es gibt vier Messniveaus:

- ✔ **Nominal**: enthält die wenigste Information aller Niveaus. Auf nominalem Niveau wird ein numerischer Wert willkürlich zugeordnet. Das Geschlecht ist ein Beispiel für ein *nominales Messniveau* (zum Beispiel 1 für »männlich«, 2 für »weiblich«, 3 für »divers«). Es ist nicht sinnvoll zu sagen, eine Ausprägung sei größer oder kleiner als die andere, lediglich über Häufigkeiten (beispielsweise 35 Teilnehmende sind weiblich, 47 männlich und 15 divers) oder über Anteilswerte (beispielsweise »in der Stichprobe waren 35 % weibliche, 40 % männliche, 20 % diverse Teilnehmende, 5 % machten keine Angaben«) können bei diesem Skalenniveau Aussagen getroffen werden.
- ✔ **Ordinal**: Die Noten bei einer Klassenarbeit sind ein Beispiel für ein *ordinales Messniveau*. Wir können Teilnehmende von der besten bis zur schlechtesten Note einordnen, aber wir wissen nicht, um wie viel besser die eine Person im Vergleich zur zweiten Person war. Wie groß zum Beispiel der Unterschied zwischen »gut« und »ausreichend« ist, bleibt meist unklar.

- ✔ **Intervall:** Beispielsweise Intelligenzquotienten werden auf *Intervallniveau* gemessen, das heißt, wir können die Punktezahlen sortieren und der Abstand zwischen zwei benachbarten Punktzahlen ist gleich groß. Der Unterschied zwischen Intelligenzquotienten von 95 und 100 ist daher also – statistisch – derselbe wie der Unterschied zwischen Intelligenzquotienten von 115 und 120.
- ✔ **Verhältnis- oder Ratioskala**: Bei einem *Verhältnis- oder Ratio-Messniveau* können die Punktwerte sortiert werden. Der Unterschied zwischen den einzelnen Punkten auf der Skala ist gleich und – das kommt zum vorherigen Skalenniveau hinzu – die Skala hat einen echten *absoluten Nullpunkt*. Beispielsweise wird das Gewicht auf Verhältnisskalenniveau gemessen. Das Vorliegen eines echten Nullpunkts sagt aus, dass ein Gewicht von null dem Fehlen von Gewicht entspricht. Außerdem können Sie proportionale Aussagen treffen, zum Beispiel »10 kg sind halb so schwer wie 20 kg, 15 kg dreimal so schwer wie 5 kg«.

Intervall- und Verhältnisskala werden auch als »metrische Skalen« oder »Kardinalskalen« bezeichnet.

Darüber hinaus sollten Sie die Variablen in Ihren Daten als *unabhängig* oder *abhängig* klassifizieren können. Diese Klassifizierung hängt von der aktuellen Fragestellung ab. Wenn Sie beispielsweise den Unterschied in Depressionspotenzialen zwischen männlichen, weiblichen und sich als divers bezeichnenden Personen untersuchen, ist die unabhängige Variable das Geschlecht, das heißt die Variable, für die Sie eine Änderung prognostizieren möchten. Das Depressionspotenzial ist hingegen die abhängige Variable, das heißt die Ergebnisvariable, deren Ergebnisse von der unabhängigen Variablen abhängen.

In den Beispielen in diesem Buch werden aus Gründen der Anschaulichkeit der Berechnungen in Beispielen häufig nur zwei Geschlechterbezeichnungen (männlich, weiblich) verwendet. Das soll keinesfalls bedeuten, dass es nur diese beiden Geschlechter gibt.

Was ist SPSS?

Dieses Buch geht davon aus, dass Sie Ihre Daten mit SPSS oder mit R unter Zuhilfenahme der Software RStudio analysieren wollen. Die beiden letztgenannten Hilfsmittel werden in einem späteren Abschnitt vorgestellt, zunächst geht es um SPSS. Die Abkürzung SPSS steht für *Statistical Package for the Social Sciences* (deutsch: Statistikpaket für die Sozialwissenschaften). Es ist ein Computerprogrammpaket, mit dem Sie Daten speichern, aufbereiten und analysieren können. SPSS ist ein gebräuchliches Statistikpaket in der Sozialwissenschaft, aber natürlich gibt es auch noch viele andere, vergleichbare Programme für speziellere Analysen.

In SPSS gibt es drei Hauptansichten oder Fenster. Die erste ist die »Variablenansicht«. Hier ordnen Sie den Variablen, mit denen Sie arbeiten möchten, *Label* oder *Etiketten* zu und codieren sie. Beispielsweise könnten Sie die beiden Variablen »Geschlecht« und »Depression« anlegen. Die »Datenansicht« ist wie ein großes Datenblatt, in das Sie alle Ihre Daten eingeben. Das Normalformat für die Dateneingabe verwendet für jede Variable eine

Spalte (zum Beispiel Geschlecht oder Depression). Jede Zeile stellt in der Regel eine Person dar, die an der Untersuchung teilgenommen hat. Wenn Sie also Informationen über das Geschlecht (männlich, weiblich, divers) und das Depressionspotenzial (nach dem Beck-Depressions-Inventar) von 10 Teilnehmenden sammeln und eingeben, hätten Sie 3 Spalten und 10 Zeilen in Ihrer SPSS-Datenansicht. SPSS gestattet die Eingabe von numerischen Daten, nicht-numerischen Informationen, wie beispielsweise Namen, und die Zuordnung von Codes (zum Beispiel 1 für »männlich«, 2 für »weiblich«, 3 für »divers«).

Im Unterschied zu Tabellenkalkulationsprogrammen, wie zum Beispiel Microsoft Excel, enthalten die Zellen keine Formeln oder Gleichungen.

Mit SPSS oder anderen Statistikprogrammen können Sie die verschiedensten Datenanalysen durchführen. Hierbei helfen Ihnen meist Dropdown-Menüs. Es gibt Hunderte verschiedener Analysen und Optionen, aus denen Sie auswählen können. In diesem Buch erklären wir nur diejenigen statistischen Verfahren, die Sie für einführende Veranstaltungen auf Bachelorniveau brauchen. Wenn Sie eine Analyse ausgewählt haben, werden Ihre Ergebnisse in einem separaten »Ausgabefenster« angezeigt. Sie können dann alle relevanten Informationen ablesen und interpretieren.

Neben der Verwendung von Dropdown-Menüs können Sie SPSS auch mithilfe einer einfachen Syntaxsprache programmieren. Das kann praktisch sein, wenn Sie immer wieder dieselben Analysen für viele verschiedene Datensätze durchführen müssen. Eine ausführliche Erklärung dieser Programmierung ist jedoch im Rahmen dieses einführenden Buchs nicht möglich.

Dennoch müssen Sie nicht vollständig darauf verzichten: Sie können über den Button `Einfügen` die über ein Dropdown-Menü ausgewählte Funktion in die Syntax einfügen und den SPSS-Code darüber ausführen lassen. Dies kann sinnvoll sein, wenn Sie zum Beispiel Auswertungen mehrfach durchführen wollen. Das »Syntaxfenster« ist ähnlich aufgebaut wie ein Texteditor. Die Syntax-Befehle können entsprechend verändert und angepasst werden.

SPSS wurde 1968 veröffentlicht. Es gibt inzwischen zahlreiche Versionen und Upgrades. In diesem Buch wurde SPSS in der Version 27 verwendet. Sollten Sie bei der Arbeit mit diesem Buch eine neuere Version verwenden, können also mittlerweile kleinere Änderungen an den Menüs und Funktionen vorgenommen worden sein. Zwischen 2009 und 2010 wurde SPSS kurzzeitig als »Predictive Analytics SoftWare« bezeichnet und war auf Ihrem Computer unter dem Namen *PASW* zu finden. 2010 wurde es von IBM gekauft und ist seither unter dem einheitlichen Namen *IBM SPSS Statistics* bekannt. (Und nein, wir wissen auch nicht, warum hinten noch einmal ein »Statistics« steht!).

Was sind R und RStudio?

R ist eigentlich eine Programmiersprache, mit der die Rechenarbeit für statistische Analysen erledigt werden kann. Außerdem kann R Grafiken, Tabellen, Diagramme und so weiter erstellen, mit denen diese Berechnungen veranschaulicht werden können. Ursprünglich entwickelt wurde R bereits in den 1990er-Jahren durch Ross Ikaha und Robert Gentleman

an der Universität von Auckland, Neuseeland. Inzwischen kümmert sich der Verein »*The R Foundation for Statistical Computing*« maßgeblich um die Weiterentwicklung der inzwischen sehr mächtigen Erweiterungen, die über das *CRAN* (Abkürzung für »*The Comprehensive R Archive Network*«) zu beziehen sind. Abbildung 1.1 zeigt die Startoberfläche von R.

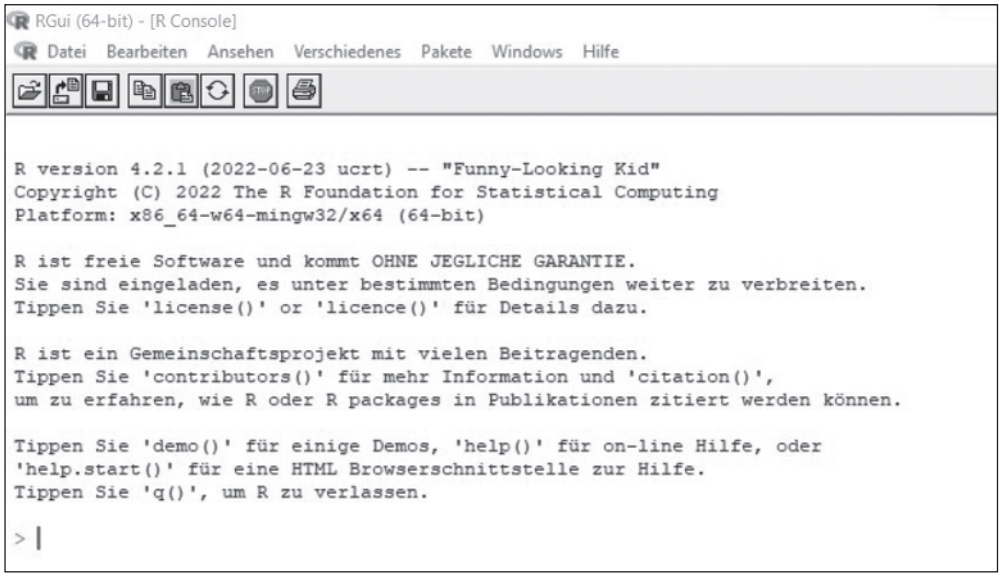

Abbildung 1.1: Die Arbeitsoberfläche von R beim ersten Start

Wie kommen Sie an das Programm heran? Sie können es sowohl für die Betriebssysteme Windows und macOS als auch für Linux zum Beispiel über die CRAN-Website der Westfälischen Wilhelms-Universität Münster CRAN.UNI-MUENSTER.DE (oder vieler anderer Hochschulen) herunterladen (siehe Abbildung 1.2).

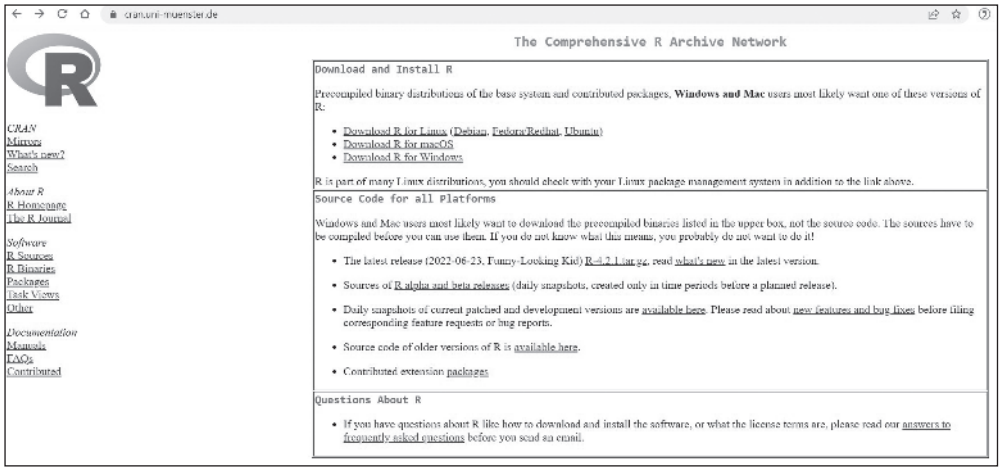

Abbildung 1.2: Download von R über die CRAN-Website der Westfälischen Wilhelms-Universität Münster

Nicht jede Person kann mit der Zeilenlogik einer Programmiersprache wie R umgehen. Komplexe Analysen werden außerdem mit der Zeit ziemlich unübersichtlich und die Befehlsketten (sogenannte *R-Codes*) kompliziert. Für eine einfachere Erstellung und Ausführung von R-Code wurde daher die Software *RStudio* entwickelt. 2022 hat sich das Unternehmen, das RStudio vertreibt, in »posit« umbenannt. Angezielt wird, dass auch die Integration von anderen Anwendungen durch Programmiersprachen wie Python erleichtert wird. RStudio Desktop gibt es als Open-Source-Variante (kostenfrei für akademische Anwendungen zum Stand Februar 2023) und als kostenpflichtige kommerzielle Version für verschiedene Betriebssysteme. Beide Varianten erfordern allerdings die vorherige Installation von R.

RStudio Desktop kann über die Website `https://posit.co/downloads/` heruntergeladen werden.

RStudio in der Version 2022.07.1+554 sieht dann beim ersten Start so aus wie in Abbildung 1.3 gezeigt (wahrscheinlich gibt es schon eine neuere Version mit vielen zusätzlichen Erweiterungen, wenn Sie diese Zeilen lesen).

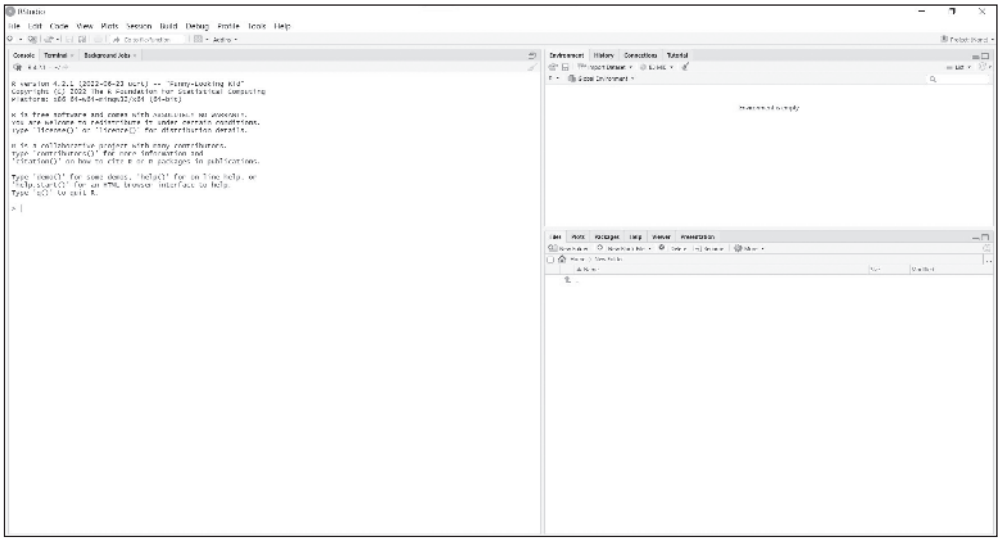

Abbildung 1.3: Oberfläche von R-Studio beim ersten Start

Die SPSS-Screenshots in diesem Buch wurden zu einem kleinen Teil auf Apple-macOS-Geräten aufgenommen, zum größeren Teil auf einem Microsoft-Windows-Rechner. Die R- und RStudio-Screenshots wurden mit einem Microsoft-Windows-Rechner erstellt. Abgesehen von den typischen betriebssystemabhängigen Aspekten ist die Bedienung von SPSS, R und RStudio auf den verschiedenen Betriebssystemen weitgehend gleich, unabhängig davon, ob ein Apple-, Linux- oder Windows-Rechner verwendet wird. Allerdings können sich zum Beispiel Schriftarten, das Öffnen oder Schließen von Fenstern und Shortcuts unterscheiden. Lassen Sie sich davon nicht verwirren.

Hin und wieder wird bemängelt, dass SPSS und R zu (leicht) unterschiedlichen Ergebnissen kämen. Dies ist in der Regel auf bestimmte (Vor-)Einstellungen in Untermenüs in SPSS beziehungsweise unterschiedlich verwendete R-*Argumente* (das sind die Ausdrücke in Klammern oder hinter einem Gleichheitszeichen nach einem R-Befehl) zurückzuführen. Schauen Sie sich daher möglichst die vollständigen Einstellungsmöglichkeiten bei SPSS beziehungsweise die Hilfe bei R an. (Wie Sie sich bei R Hilfe holen können, erläutern wir später.)

Wir haben uns in diesem Buch bemüht, in R die gleichen Verfahren einzusetzen wie in den SPSS-Beispielen. Der häufig geäußerte Wunsch, dass SPSS und R exakt gleiche Ergebnisse erzielen sollen, widerspricht allerdings oft der Idee von Paketentwickelnden in R. (Was ein Paket oder Package in R ist, erfahren Sie in Kürze.) Bei der Paketentwicklung sind statistische Fachartikel die Leitlinie mit dem Ziel, ein Verfahren bestmöglich in R zu implementieren. Meistens gibt es einige Freiheiten bei der Entwicklung, sodass ziemlich häufig unterschiedliche Ergebnisse zwischen verschiedenen Paketen im Vergleich mit SPSS entstehen. Die Paketentwickelnden haben in der Regel nicht den Anspruch, die gleichen Ergebnisse einer anderen Software »nachzubauen«, sondern eine bestmögliche Implementierung hinzubekommen. Manchmal enthalten SPSS-Algorithmen oder R-Codes aber auch einfach nur Fehler, die erst in neueren Versionen korrigiert werden.

Im vorliegenden Buch geben wir Ihnen meist diejenige R-Funktion an, die zu denselben oder zumindest ähnlichen Ergebnissen kommt wie eine Auswertung mit SPSS. Das bedeutet nicht, dass es sich jeweils um den bestmöglichen Analyseweg handelt. Sprechen Sie mit Ihrer betreuenden Lehrkraft oder einer statistikkundigen Person über den richtigen Auswertungsweg in Ihrem konkreten Anwendungsfall.

Wenn Sie in RStudio einen Befehl eingeben, diesen mit dem Cursor markieren und die Funktionstaste `F1` drücken, erhalten Sie im Fenster rechts unten unter dem Reiter *Help* ausführliche Angaben zu diesem Befehl. In der obersten Zeile finden Sie noch einmal den Befehlsnamen und in geschweiften Klammern das Paket (Package), aus dem der Befehl stammt. Dann folgen eine Beschreibung, die Verwendung des Befehls mit den möglichen Argumenten, eine Angabe zu den ausgegebenen Ergebnissen, Quellenangaben, R-Code-Beispiele und so weiter.

Wie in nahezu allen Wissenschaften gibt es auch in der Psychologie Konventionen, wie Daten berichtet werden (sollen). Üblicherweise werden Zahlen in der angloamerikanischen Notation berichtet, das heißt mit einem Dezimalpunkt statt eines Dezimalkommas. In diesem Buch verwenden wir allerdings die klassische »deutsche« Notation mit dem Komma als Dezimalzeichen. Um später in Ihrem eigenen Bericht nichts falsch zu machen, fragen Sie unbedingt die Person, die Ihre Arbeit betreut, welche Notationsform gewünscht ist! Wollen Sie Ihre Studie in einem Fachjournal veröffentlichen, gibt es in der Regel bestimmte Notationsvorgaben. Wir orientieren uns in diesem Buch am Publikationsmanual der American Psychological Association (APA).

Darstellungen mit vielen Nachkommastellen werden schnell unübersichtlich, weshalb die Zahlen in der Regel gerundet werden. Generell werden Zahlen größer als 100 ohne Nachkommastelle berichtet (zum Beispiel Mittelwert M = 1533). Das gilt auch für Werte, die keine Nachkommastellen haben können (zum Beispiel Teilnehmendenzahlen N = 435). Zahlen von 10 bis100 werden meist mit einer Nachkommastelle dargestellt (zum Beispiel M = 15,3). Zahlen zwischen 0,10 und 10 erhalten zwei Nachkommastellen (zum Beispiel Standardabweichung SD = 0,54). Werte unterhalb von 0,10 werden mit drei Nachkommastellen berichtet. p-Werte (was das ist, werden Sie später noch erfahren) werden bis p = 0,001 möglichst exakt berichtet. Üblicherweise wird bei p-Werten, Korrelationskoeffizienten, partiellem Eta-Quadrat η_p^2 und anderen Parametern, die nicht größer als 1 werden können, die führende Null weggelassen. Da wir in diesem Buch bei der deutschen Notation bleiben, lassen wir die führende Null stehen. Und keine Angst, falls Sie die vielen Bezeichnungen verwirren: Mit allen werden wir uns noch ausführlich beschäftigen.

In der Regel werden Abkürzungen mit lateinischen Buchstaben (zum Beispiel M für den Mittelwert oder SD für die Standardabweichung) kursiv gesetzt, griechische Buchstaben dagegen nicht (zum Beispiel η_p^2). Vor und nach Gleichheitszeichen sollte ein Leerzeichen gesetzt werden.

Deskriptive Statistik

Wenn Sie Daten sammeln, müssen Sie Ihre Erkenntnisse anderen Menschen mitteilen (zum Beispiel Ihrer Professorin oder Ihrem Chef). Angenommen, Sie sammeln Daten von 100 Menschen hinsichtlich ihres Grades an Coulrophobie (das ist die Angst vor Clowns). Wenn Sie einfach dafür eine Liste mit 100 Werten in SPSS, einem Online-Survey-Tool oder einem Tabellenkalkulationsprogramm erstellen, dann ist das nicht besonders aussagekräftig. Kaum jemand kann alle diese Daten gleichzeitig im Kopf verarbeiten und brauchbare Schlussfolgerungen daraus ziehen. Stattdessen brauchen Sie eine Möglichkeit, Ihren Datensatz knapp, präzise und zusammenfassend zu beschreiben. Im Normalfall geben Sie dafür (mindestens) zweierlei Informationen an – Lagemaße und Streuungen.

Lagemaße

Es gibt verschiedene Arten von Lagemaßen. Alle versuchen, eine Zahl zu erzeugen, die Ihre Daten im Mittel genauer erklärt, und alle beginnen im Deutschen mit dem Buchstaben »M«. Das gebräuchlichste Maß ist der *Mittelwert*. Korrekt heißt dieses Maß *arithmetisches Mittel*. Vermutlich sind Sie schon mit diesem Maß vertraut. Um den Mittelwert zu erhalten, addieren Sie einfach alle Werte für eine Variable und dividieren anschließend durch die Anzahl der betrachteten teilnehmenden Personen oder Fälle.

Eine der Stärken des Mittelwerts als Lagemaß ist, dass er alle Ihre Daten repräsentiert. Das ist jedoch auch gleichzeitig eine Schwäche, weil er von Extremwerten beeinflusst wird und dadurch Verzerrungen unterliegen kann. Er stellt Ihre Daten also nicht immer auf eine wirklich repräsentative Weise dar. Der *Median* oder *Zentralwert* als mittlerer Wert sortierter

Werte ist gut geeignet, wenn Ihre Variable auf dem Ordinalniveau der Messung bewertet wird. Der *Modalwert* oder auch *Modus* genannt) ist der am häufigsten auftretende Wert in einem Datensatz. Er eignet sich als Lagemaß, wenn Ihre Variable auf Nominalniveau bewertet wird. Lagemaße werden in Kapitel 4 genauer beschrieben.

Streuung

Maße für die Streuung sind Zahlen, die die Standardabweichung oder Variabilität Ihrer Variablen zeigen. Je größer der Streuungswert für eine Variable, desto größer ihre Variabilität, zum Beispiel wie stark die Noten von Studierenden variieren. Ein kleiner Streuungswert deutet darauf hin, dass die Noten sehr wenig variieren und die Teilnehmenden tendenziell dieselbe Note erhalten. Die gebräuchlichsten Maße für die Streuung sind die Spannweite, die Varianz und die Standardabweichung. Sie sind eine Schätzung der Variabilität Ihrer Variablen. Kapitel 5 beschreibt die Standardabweichung sowie andere wichtige Maße für die Streuung.

Diagramme

Eine weitere Möglichkeit, Ihre Daten übersichtlich zu präsentieren, besteht in der visuellen Darstellung in Form eines Diagramms. Diagramme sind auch aus einem anderen Grund wichtig. Die Art der statistischen Analyse, die Sie mit Ihren Variablen durchführen können, hängt nämlich von der konkreten Verteilung Ihrer Variablen ab. Diese können Sie unter Verwendung von Diagrammen am einfachsten beurteilen. In Kapitel 6 lernen Sie einige der in der Psychologie gebräuchlichsten Diagramme kennen (Histogramm, Balkendiagramm, Kreisdiagramm und Box-Whisker-Plot) und erfahren, wie Sie sie in SPSS oder R erstellen können.

Standardisierte Messwerte

Eine einfache Mitteilung von Rohdatenwerten ist häufig nicht besonders informativ. Sie müssen vielmehr in der Lage sein, diese Werte mit den Werten anderer Personen zu vergleichen. Sie müssen auch Messwerte vergleichen können, die unter Verwendung unterschiedlicher Skalen ermittelt wurden. Stellen Sie sich beispielsweise vor, Sie haben das Extraversionsniveau eines Freundes unter Verwendung des revidierten NEO-Persönlichkeitsinventars gemessen und ihm mitgeteilt, dass er den Wert 164 erzielt hat. Sehr wahrscheinlich will er wissen, was dieser Wert im Vergleich zu den Werten anderer Personen bedeutet. Ist er hoch oder niedrig? Außerdem will er vielleicht auch wissen, wie der Wert sich zu der Psychotizismus-Bewertung von 34 verhält, die Sie ihm in der letzten Woche entsprechend dem Eysenck-Persönlichkeits-Inventar zugeordnet hatten.

Die gute Nachricht: Es ist ziemlich einfach, Ihren Rohdatenwert in einen standardisierten Wert umzuwandeln, anhand dessen solche Vergleiche möglich sind. Ein standardisierter Messwert wird in Standardabweichungen seiner eigenen Verteilung angegeben (so können Sie ihn mit standardisierten Messwerten anderer Zufallsvariablen vergleichen) und gestattet einen unmittelbaren Vergleich mit dem Mittelwert für eine Variable (so können Sie einer

Person mitteilen, ob ihr Wert höher oder niedriger als der Mittelwert ist). Wir werden in Kapitel 10 genauer über die Standardisierung sprechen.

Inferenzstatistik

Deskriptive Statistik ist praktisch, um die Eigenschaften Ihrer Stichprobe (also der Teilnehmenden, über die Sie Daten gesammelt haben) zu verdeutlichen. In den meisten Fällen sind Sie jedoch an den Eigenschaften einer ganzen Population interessiert, das heißt an den Eigenschaften aller möglichen Teilnehmenden. Vielleicht möchten Sie Ihre Aussagen auch auf eine größere Population (Grundgesamtheit) generalisieren können. Wenn Sie beispielsweise Haltungsunterschiede gegenüber Sekten zwischen männlichen, weiblichen und sich als divers bezeichnenden Schülern in Deutschland ermitteln wollen, besteht Ihre Grundgesamtheit aus allen deutschen Schulkindern. Da es aber unrealistisch ist, alle Kinder in Deutschland zu befragen, messen Sie die Tendenz bezüglich Sekten in einer kleinen Untermenge oder *Stichprobe* der Kinder. In Kapitel 7 erfahren Sie mehr über die Unterschiede zwischen Stichproben und Populationen.

Die *inferentielle Statistik* oder *Inferenzstatistik* gestattet Ihnen, basierend auf Ihrer Stichprobe Schlussfolgerungen für eine größere Population und evtl. die Grundgesamtheit zu ziehen. Wenn Sie beispielsweise in Ihrer Stichprobe Unterschiede zwischen Jungen, Mädchen und sich als divers bezeichnenden Schulkindern gefunden haben, können Sie diesen Unterschied auf alle Schulkinder übertragen. Natürlich ist nicht komplett sicher, dass diese Unterschiede in der Grundgesamtheit tatsächlich auch bestehen, weil Sie nicht jedes Kind in der Population getestet haben. Aber unter bestimmten Bedingungen können Sie mit statistischen Mitteln mit einem festgelegten Prozentniveau (meist 95 % oder 99 %) Vertrauen haben, dass die in Ihrer Stichprobe festgestellten Unterschiede auch in der Population vorliegen, über die Sie eigentlich eine Aussage treffen wollen. Mehr zu Inferenzstatistik und den gerade genannten Bedingungen erfahren Sie ebenfalls in Kapitel 7.

Die von Ihnen erstellte inferentielle Statistik gibt Auskunft über die Wahrscheinlichkeit, dass Ihr Ergebnis in der Population auftritt, das heißt, ob in Ihrer Stichprobe festgestellte Unterschiede tatsächlich auch in der Population vorhanden sind oder ob das Ergebnis aufgrund von Zufallsprozessen zustande kommt. Sie sagt Ihnen allerdings nichts über die Größe dieses Unterschieds. Wenn Sie beispielsweise feststellen, dass Männer wahrscheinlicher an Sekten interessiert sind als Frauen oder sich als divers bezeichnende Personen, ist dies nicht besonders interessant, wenn die Unterschiede sehr gering sind und keine praktische Bedeutung haben. *Effektgrößen* können diesen Mangel ausgleichen, denn sie geben die Stärke der Beziehung zwischen Ihren Variablen an. Um Effektgrößen wird es in Kapitel 11 gehen. Sie sollten Effektgrößen immer in Verbindung mit dem Wahrscheinlichkeitsmaß einer Inferenzstatistik angeben. Im Laufe des Buchs werden Sie verschiedene Effektgrößen kennenlernen.

Hypothesen

Bevor Sie mit einer Studie beginnen, brauchen Sie eine Hypothese, das heißt eine spezifische und überprüfbare Aussage bezüglich des Zieles Ihrer Studie. Aus Anschaulichkeitsgründen

reduzieren wir die Anzahl der betrachteten Gruppen in diesem Buch meist auf zwei. Im obigen Beispiel wäre eine Hypothese »Es gibt keinen Unterschied bei der Sektenempfindlichkeit zwischen Jungen und Mädchen an deutschen Schulen« denkbar. Wir beschreiben den Hypothesentest in Kapitel 8. Dort erklären wir Ihnen auch, warum Sie mit der Annahme beginnen sollten, dass Ihre Daten *keinen* Effekt, *keinen* Unterschied oder *keine* Beziehung aufweisen.

Parametrische und nicht parametrische Tests

Wenn Sie an eine Hypothese herangehen, können Sie zweierlei statistische Analysen durchführen: einen *parametrischen* Test oder ein *nicht-parametrisches* Verfahren. Parametrische Statistiken gehen davon aus, dass sich die Daten einer bestimmten Verteilung wie beispielsweise der in Kapitel 9 erklärten Normalverteilung annähern. Daraus ergeben sich Schlussfolgerungen, die diese Art Statistik sehr leistungsstark (»*teststark*«) machen und präzise Ergebnisse erzeugen. Eine genauere Beschreibung der Teststärke finden Sie in Kapitel 11. Da parametrische Statistiken jedoch auf bestimmten Annahmen basieren, müssen Sie Ihre Daten im Vorfeld bezüglich dieser Annahmen überprüfen (Wie Sie das tun, erklären wir für jede einzelne Statistik in diesem Buch). Wenn Sie die Annahmen nicht überprüfen, riskieren Sie, dass Ihre Ergebnisse und damit auch Ihre Schlussfolgerungen fehlerhaft sind.

Die nicht-parametrischen Statistiken hingegen treffen weniger Annahmen über Ihre Daten, haben also in der Regel weniger Voraussetzungen für ihre Anwendung. Entsprechend können Sie diese Statistiken nutzen, um einen vielfältigeren Datenbereich zu analysieren. Nicht-parametrische Tests sind im Allgemeinen weniger teststark als ihre parametrischen Äquivalente, deshalb sollten Sie immer versuchen, parametrische Tests zu verwenden – es sei denn, die Daten verletzen die Annahmen für solch einen Test. (Welche das sind, erfahren Sie im jeweiligen Kapitel.)

Forschungsdesigns

Welche statistischen Analysen Sie durchführen sollten, hängt von mehreren Faktoren ab. Um sich für die richtige »Testfamilie« zu entscheiden, müssen Sie zunächst über das Design Ihrer Studie nachdenken. Die Auswahl des Designs hängt wesentlich von Ihrer Hypothese ab. Ein Forschungsdesign kann ganz allgemein entweder als *korrelatives* oder *experimentelles Design* aufgebaut werden.

Korrelatives Design

Beim korrelativen Design sind Sie an den *Beziehungen* zwischen zwei oder mehr Variablen interessiert. Dieses Design unterscheidet sich vom experimentellen Design insofern, als dabei keine Variablen manipuliert werden. Stattdessen untersuchen Sie vorhandene Beziehungen zwischen den Variablen. Beispielsweise könnten Sie eine Studie durchführen, um die Beziehung zwischen einem gelegentlichen Drogenkonsum und visuellen Halluzinationen

zu untersuchen. In diesem Fall müssen Sie Teilnehmende mit unterschiedlich hohem Drogenkonsum anwerben und deren Erfahrungen zu Halluzinationen messen. Vermutlich werden Sie auf erhebliche Bedenken treffen, wenn Sie eine experimentelle Studie in diesem Bereich durchführen und illegale Drogen in unterschiedlicher Dosierung verteilen, um Halluzinationen bei Ihren Teilnehmenden zu verursachen. Teil III des Buchs beschäftigt sich mit der Inferenzstatistik, den Beziehungen oder Assoziationen zwischen Variablen, die normalerweise aus einem korrelativen Design stammen. (Bitte achten Sie auf unsere Verwendung des Wortes »normalerweise«: Es gibt fast immer Ausnahmen!)

Korrelationskoeffizienten zeigen die Stärke einer (in der Regel) linearen Beziehung und deren Richtung. Liegt eine starke positive Korrelation vor, korrelieren hohe Werte für eine Variable im Allgemeinen mit hohen Werten für die andere Variable (zum Beispiel tendieren Teilnehmende mit hohem Drogenkonsum auch zu einem hohen Maß an Halluzinationen) linear. Eine starke negative Korrelation gibt an, dass hohe Werte für eine Variable im Allgemeinen mit niedrigen Werten für die andere Variable linear korrelieren (zum Beispiel könnten Teilnehmende mit hohem Drogenkonsum auch zu einem niedrigen Maß an Halluzinationen tendieren). Es gibt viele verschiedene Arten von Korrelationskoeffizienten. Sie müssen sich jeweils entscheiden, welcher für Ihre Daten am besten geeignet ist. Wie Sie Korrelationskoeffizienten ermitteln und interpretieren, erklären wir in Kapitel 12.

Die Regression treibt das Konzept der Beziehungsuntersuchung noch einen Schritt weiter. Sie gestattet Ihnen zu testen, ob eine oder mehrere Variablen eine Ergebnisvariable oder ein Kriterium vorhersagen können. Beispielsweise könnten Sie die Regression verwenden, um zu untersuchen, ob anhand des Konsums illegaler Drogen oder anhand von Angststörungen oder anhand des Alters visuelle Halluzinationen vorhergesagt werden können. Im Vergleich zur Korrelation bietet Ihnen diese Technik mehr Informationen, zum Beispiel welche Variablen signifikante Prädiktoren für Halluzinationen sind oder die relative Stärke jeder prädiktiven Variablen. Um die einfachste Form der Regression, die lineare Regression, geht es in Kapitel 13 dieses Buchs.

Falls Sie die Verbindung zwischen zwei diskreten Variablen untersuchen möchten, zum Beispiel, ob es eine Verbindung dazwischen gibt, dass jemand überhaupt illegale Drogen nimmt und dass er gelegentlich Halluzinationen hat oder nicht, muss dieser Datentyp in einer bestimmten Art und Weise untersucht werden. In Kapitel 14 erklären wir, wie Sie hierzu Kontingenztabellen verwenden (zum Beispiel den Chi-Quadrat-Test und den McNemar-Test), um diskrete Variablen zu analysieren.

Experimentelles Design

Experimentelle Designs unterscheiden sich von korrelativen Designs insofern, als dabei die unabhängige Variable *manipuliert* und der Einfluss der unabhängigen auf die abhängige Variable genauer untersucht werden kann. (Eine Beschreibung der Variablentypen finden Sie in Kapitel 2.) Korrelative Studien konzentrieren sich auf die Beziehung zwischen vorhandenen Variablen, während beim experimentellen Design eine Variable (direkt oder indirekt) verändert wird und deshalb beeinflussende von beeinflussten Variablen unterschieden werden können. Sie bewerten dann, ob sich Einflüsse der unabhängigen Variablen auf Ihre abhängige Ergebnisvariable zeigen. Beispielsweise könnten Sie die Hypothese aufstellen, dass

Ergophobie (die Angst vor der Arbeit) bei Psychologiestudierenden im Verlauf ihres Studiums zunimmt. Es gibt zwei experimentelle Designs, wie Sie diese Hypothese testen könnten.

Im ersten Design könnten Sie den Grad an Ergophobie bei Studierenden in verschiedenen Phasen ihres Studiums testen. Eine Studie, bei der Sie separate unabhängige Teilnehmendengruppen testen, wird als *Design mit unabhängigen Gruppen* bezeichnet. (Die statistischen Analysen für dieses Design werden in Teil IV des Buchs betrachtet.) Beim zweiten Design wird der Grad an Ergophobie bei allen Erstsemesterstudierenden gemessen, anschließend werden dieselben Teilnehmenden dann mehrfach im Verlauf ihres Studiums getestet. Solch eine Studie, bei der Sie an Änderungen innerhalb derselben Teilnehmendengruppe interessiert sind, wird auch als *Design mit wiederholten Messungen* oder *Messwiederholungsdesign* bezeichnet. Die statistischen Analysen für dieses Design werden in Teil V des Buchs besprochen.

Design mit unabhängigen Gruppen

Wenn Sie ein Design mit unabhängigen Gruppen anwenden, suchen Sie nach Unterschieden bei einer Variablen zwischen separaten Personengruppen. Wenn Sie Unterschiede zwischen zwei separaten Gruppen in einer Variablen betrachten (zum Beispiel den Grad an Ergophobie von Psychologiestudierenden im ersten und im zweiten Jahr), ist es wahrscheinlich, dass sich die beiden Wertemengen in gewissem Ausmaß unterscheiden. Eine Inferenzstatistik schätzt, mit welcher Wahrscheinlichkeit es diesen Unterschied auch in der Population gibt. Sie gibt Ihnen ein Maß dafür an, wie wahrscheinlich Sie das Ergebnis zufällig erhalten haben könnten (bestimmte Bedingungen vorausgesetzt). In diesem Szenario können Sie beispielsweise entweder den parametrischen unabhängigen *t*-Test verwenden oder den nicht-parametrischen Mann-Whitney-U-Test. Mehr über diese Tests erfahren Sie in Kapitel 15.

Wenn Sie den Unterschied zwischen mehr als zwei Gruppen untersuchen möchten (zum Beispiel Psychologiestudierenden in ihrem ersten, zweiten und dritten Studienjahr), ist die parametrische Varianzanalyse (ANOVA) zwischen Gruppen am besten geeignet. Das nicht-parametrische Äquivalent ist der Kruskal-Wallis-Test. All diese Analysen werden in Kapitel 16 genauer beschrieben. Wenn Sie feststellen, dass es einen statistisch signifikanten Unterschied in den Ergophobie-Graden zwischen Psychologiestudierenden im ersten, zweiten und dritten Jahr gibt, müssen Sie herausfinden, wo genau diese Unterschiede liegen (zwischen dem ersten und dem zweiten Jahr, zwischen dem ersten und dem dritten Jahr und so weiter). Post-hoc-Tests, die Sie dafür benötigen, sind in Kapitel 17 genauer beschrieben. Auch geplante Vergleiche (Kontraste) werden erwähnt.

Wenn Sie die Wirkung zweier unabhängiger Variablen auf eine abhängige Variable untersuchen, könnten Sie auch ein etwas komplexeres Design verwenden. Wirken sich zum Beispiel das Studienjahr und das Geschlecht auf den Grad an Ergophobie aus? Der Vorteil bei diesem Design ist, dass Sie die Interaktion zwischen Ihren beiden unabhängigen Variablen untersuchen können. Zum Beispiel könnten die Ergophobie-Grade der Frauen konstant bleiben, während die Werte für die Männer in den drei Jahresgruppen stetig zunehmen. Solch ein zweifaktorielles ANOVA-Design zwischen Gruppen ist in Kapitel 16 genauer beschrieben.

Design mit wiederholten Messungen

Wenn Sie ein Design mit wiederholten Messungen anwenden, suchen Sie nach Unterschieden für eine Variable innerhalb derselben Personengruppe. Beispielsweise könnten Sie Ergophobie-Grade messen, wenn die Studierenden mit ihrer ersten Psychologielehrveranstaltung beginnen, und dann 12 Monate später bei derselben Gruppe prüfen, ob sich die Werte verändert haben. Nachdem Sie Ihre Teilnehmenden zweimal getestet haben (das heißt, die unabhängige Variable hat zwei Stufen), könnten Sie den paarweisen *t*-Test oder den nicht-parametrischen Wilcoxon-Test anwenden, um herauszufinden, ob die Unterschiede in den Werten statistisch signifikant sind. Wir besprechen diese Tests in Kapitel 18. Wenn Sie dieselbe Teilnehmendengruppe mehr als zweimal testen, sind die parametrische ANOVA innerhalb von Gruppen oder der nicht-parametrische Friedman-Test am besten geeignet. Diese Analysen werden in Kapitel 19 beschrieben. Wenn Sie feststellen, dass es einen statistisch signifikanten Unterschied in den Ergophobie-Graden zwischen den Testsitzungen gibt, müssen Sie genau analysieren, wo diese Unterschiede zustande kommen (zwischen dem ersten oder zweiten Jahr der Studierenden, oder ist die Änderung zwischen dem zweiten und dritten Jahr der Studierenden aufgetreten und so weiter). Die dazugehörigen Post-hoc-Tests sind in Kapitel 20 beschrieben.

Wenn Sie ein komplexeres Design entwerfen wollen, können Sie die Wirkung von zwei wiederholten Messvariablen auf eine abhängige Variable untersuchen. Man spricht auch von einer zweifaktoriellen ANOVA innerhalb von Gruppen (oder einer zweifaktoriellen ANOVA mit Messwiederholungen), wie in Kapitel 19 beschrieben. Alternativ können Sie sich für ein Design mit wiederholten Messungen und einer unabhängigen Gruppe für eine abhängige Variable entscheiden. Dies ist ein gemischtes ANOVA-Design, das in Kapitel 21 genauer betrachtet wird.

Darüber hinausgehende multivariate Analysen (MANOVA) werden in diesem Buch nicht besprochen.

Die ersten Schritte

Die kritischste Phase jeder Forschungsstudie ist der Anfang. Sie müssen eine Hypothese formulieren, die Sie testen können und die sich wirklich auf die Frage bezieht, an der Sie interessiert sind. (Weitere Informationen über Hypothesen finden Sie in Kapitel 8.) Ihre Hypothese muss durch die Theorie und frühere Forschungen gestützt werden. Außerdem müssen Sie sich überlegen, wie Sie Ihre Daten analysieren möchten, und sich für eine geeignete Statistik entscheiden. Diese Faktoren spielen eine wichtige Rolle für die Wahl und Messung Ihrer Variablen. Nicht, dass Sie irgendwann feststellen, dass Sie Ihre Hypothese gar nicht widerlegen können, weil Sie für die gesammelten Daten die gewünschte Analyse überhaupt nicht ausführen können!

Nach der Entscheidung für eine geeignete statistische Analyse können Sie berechnen, welche Stichprobengröße Sie benötigen (siehe Kapitel 11). Wenn Sie nicht genügend Teilnehmende befragen, erkennen Sie wahrscheinlich keinen signifikanten Effekt in Ihren Daten, selbst wenn es innerhalb der Population einen solchen gibt. Ihre Bemühungen sind dann

reine Zeitverschwendung. Wenn Sie Ihre SPSS- oder R-Datei vorbereiten, nehmen Sie sich die Zeit, um Ihre Daten angemessen zu beschriften und einfach zu lesende Werte zuzuweisen, die ihren Sinn auch dann noch verraten, wenn Sie sie Monate später wieder ansehen. Dazu ist es hilfreich, die verwendeten Bezeichnungen und Werte auch handschriftlich in einem *Laborbuch* festzuhalten, um bei der Durchführung von Analysen darauf zurückgreifen zu können. Trotz inzwischen fortgeschrittener automatischer Datensicherung: Denken Sie bei der Eingabe Ihrer Daten daran, Ihre Dateien regelmäßig zu speichern!

Wenn der Tag der Wahrheit kommt und Sie beginnen, Ihre Daten zu analysieren, atmen Sie tief durch und entspannen sich. Nehmen Sie sich ausreichend Zeit, machen Sie immer wieder Notizen, speichern Sie regelmäßig Ihre Ausgabedateien. Bereiten Sie alles sorgfältig vor, aber gestatten Sie sich auch, Fehler zu machen! SPSS oder R führen statistische Berechnungen so gut wie unmittelbar aus. Wenn Sie also einen Fehler machen (und das wird mit hoher Wahrscheinlichkeit passieren), können Sie einfach nachjustieren und die Berechnung neu beginnen.

Wir empfehlen Ihnen, schon beim Entwerfen Ihrer Studie eine statistische Beraterin oder einen statistischen Berater hinzuzuziehen. Die fachkundige Person kann Sie dabei beraten, welche Art Daten Sie sammeln sollten, welche Analysen sinnvoll sind und welche Stichprobengröße Sie benötigen. Wenn Sie die Daten erst einmal gesammelt haben, könnte es für eine Bitte um Hilfe zu spät sein (und damit das Risiko erhöht sein, dass ein sogenanntes Garbage-in-Garbage-out-Phänomen auftritt, das heißt, dass Sie Schlussfolgerungen aus unzureichenden Daten ziehen …).