utb.

Elisabeth Steiner
Michael Benesch

Der Fragebogen

Von der Forschungsidee zur SPSS-Auswertung

5. Auflage

Leseprobe

Vorwort zur 5., aktualisierten und überarbeiteten Auflage

Als wir vor mittlerweile zehn Jahren im Oktober 2008 mit der Erstauflage starteten, konnten wir den Erfolg unseres Konzepts für dieses Lehrbuch nicht vorhersehen. Umso erfreulicher ist es nun, die 5., aktualisierte und überarbeitete Auflage vorliegen zu haben, bedeutet dies doch für uns eine hohe Akzeptanz des gesetzten Ziels des Lehrbuches, mit intuitiven Zugängen und einem gesunden Maß an „Alltagsverständnis" an die Aufbereitung bzw. Vermittlung von statistischen Grundkenntnissen an EinsteigerInnen heranzugehen. Unterstützt wird dieser Eindruck auch durch Rückmeldungen von Studierenden unterschiedlicher Studienrichtungen und von KollegInnen, die in der Praxis der Vermittlung von quantitativen Forschungsmethoden stehen.

Inhaltlich wurden nur minimale sprachliche Veränderungen vorgenommen und die Beispiele mit der aktuellen Version von SPSS, der Version 24, dargestellt.

Es bleibt uns nun wieder die Hoffnung, dass das vorliegende Buch auch weiterhin Unterstützung in verschiedensten Studienrichtungen bei der Erstellung wissenschaftlicher Arbeiten bzw. Abschlussarbeiten bietet und vielleicht auch ein wenig Lust auf die Generierung von Daten und deren statistische Auswertung macht. Zumindest wäre es schon ein wesentlicher Schritt in die richtige Richtung, wenn Ängste bzw. Vorbehalte, die quantitativen Zugängen oft entgegengebracht werden, reduziert bzw. ein wenig verändert werden könnten.

Wien, im April 2018

Elisabeth Steiner
Michael Benesch

Vorwort

An Universitäten, Fachhochschulen sowie Studien- bzw. Ausbildungslehrgängen an den unterschiedlichsten Einrichtungen wird Wissen in Statistik und empirischer Methodik an EinsteigerInnen vermittelt. Diese haben oft aufgrund entsprechender Erfahrungen während der Schulzeit keinen bedenkenlosen Zugang zur Materie und „schalten" allzu gerne ab, wenn sie mit formalen Herleitungen, „Formelwerk" oder abstraktem statistischen Gedankengut konfrontiert werden. Dieses Wissen wird jedoch auch in Gesundheits- und sozialen Berufen auf akademischem Niveau zunehmend benötigt. Der deutliche Trend hin zu empirisch fundierten Studien resultiert aus dem Wunsch der Etablierung eigener wissenschaftlicher Zugänge und stellt neue Herausforderungen an die Studierenden.

Das vorliegende Buch versteht sich jedoch nicht nur als Unterstützung für Personen aus diesen Bereichen, sondern ist ein generelles Angebot an EinsteigerInnen, in die vorliegende Thematik einzusteigen.

Die Idee zu diesem Buch entstand im Zuge unserer langjährigen Lehrtätigkeit in den unterschiedlichsten Bereichen und der dabei gewonnenen Erfahrungen in der Vermittlung statistischer Grundkenntnisse an EinsteigerInnen. Dabei haben wir immer wieder eine wesentliche Beobachtung machen können, nämlich die, dass eine eher intuitive, auf „Alltagsverständnis" aufbauende Herangehensweise, welche auf formalistische Zugänge weitestgehend verzichtet, von den Studierenden sehr geschätzt wird und das Interesse am Fach fördert. Dieses Buch möchte also eine didaktische Lücke schließen: Es möchte keine Formeln, keine abstrakten Herleitungen, sondern eine „sanfte" Hinführung zur empirischen Methodik für EinsteigerInnen in das Feld bieten. Es soll der Forschungsprozess von der Idee bis zur statistischen Auswertung und Berichterstellung vermittelt werden, um eine Grundlage für die weitere Beschäftigung mit dem Thema zu schaffen.

Methodisch fortgeschrittenere LeserInnen mögen uns deshalb bitte verzeihen, wenn manches zugunsten einer für EinsteigerInnen verständlicheren Darstellung formal-methodisch vereinfachend beziehungsweise „sparsam" transportiert wird – wenn die Grundlagen erst einmal geschaffen sind, wird den Interessierten die Lektüre weiterführender Werke empfohlen.

Alle im Buch angeführten Beispiele können mithilfe von SPSS selbst nachgerechnet werden – das entsprechende Datenfile finden Sie auf www.utb-mehr-wissen.de. Die Daten sind fiktiv und beziehen sich auf den im Anhang abgebildeten Übungsfragebogen.

Wir empfehlen auch die selbstständige Bearbeitung der jedem Kapitel angehängten Übungsbeispiele – zu Ihrer Kontrolle finden Sie Musterlösungen auf den Seiten 175 bis 186.

Wir wünschen Ihnen jedenfalls viel Spaß beim Einstieg in die empirische Forschung!

Wien, im Oktober 2008

Elisabeth Raab-Steiner
Michael Benesch

Inhaltsverzeichnis

7 Schluss von der Stichprobe auf die Population

Normalerweise ist die bloße **Beschreibung einer Stichprobe** nicht das, was wirklich interessiert. Nehmen Sie manchmal, wenn Sie Spaghetti kochen, eine einzelne Nudel aus dem Topf und probieren, ob sie schon weich genug ist? Und sind Sie daran interessiert, zu erfahren, ob diese einzelne Nudel essfertig ist, oder benutzen Sie sie nicht vielmehr als Stichprobe, um auf den Rest der Teigwaren im Topf – die Grundgesamtheit Ihrer Nudeln – zu schließen?

Dabei könnte es geschehen, dass Sie zufälligerweise eine Nudel herausfischen, die als einzige ausreichend gar ist, aus welchen Gründen auch immer. Nehmen wir an, Sie haben die im Topf befindlichen Nudeln immer ausreichend umgerührt, das Wasser im Topf kocht gleichmäßig, die einzelnen Nudeln sind in puncto Zusammensetzung der Teigmasse und hinsichtlich ihrer Stärke vergleichbar usw., dann ist es allerdings äußerst unwahrscheinlich, dass Ihre „Stichprobennudel“ gar ist, die anderen aber noch nicht. Diese eine Nudel wird wohl repräsentativ für die Grundgesamtheit sein.

7.1 Alltags- und statistische Hypothesen

Im Alltagsleben stellen wir mehr oder minder ständig **Hypothesen** über das, was wir erleben und wie wir es erklären, auf, ohne uns dessen bewusst zu sein. „Wenn meine Stichprobennudel gar ist, werden wohl alle anderen auch gar sein“, oder „Wer Armani-Anzüge trägt, verdient viel Geld“, oder „Wer Universitätsprofessor ist, verfügt über einen hohen Intelligenzquotienten“. Beispiele für derartige Alltagshypothesen sind leicht zu finden, und wir suchen und finden Erklärungsmodelle für unsere Entscheidungen und Bewertungen zumeist implizit. Wir legen nicht aufgrund transparenter, nachvollziehbarer Kriterien fest, was und aus welchem Grund wir dies für dermaßen richtig halten, dass wir unsere Entscheidungen darauf aufbauen.

Auch in der Statistik werden Entscheidungen aufgrund von Hypothesen getroffen. Allerdings unterscheiden sich diese Hypothesen in einigen wesentlichen Punkten gravierend von den eben besprochenen **Alltagshypothesen**. **Statistische Hypothesen** werden stets als „Hypothesenpaar“ formuliert: Die sogenannte **„Nullhypothese“** steht der **„Alternativhypothese“** (auch Forschungshypothese genannt) gegenüber, und es ist die Aufgabe der Signifikanztests (Kapitel 8), diese Hypothesen zu überprüfen.

Die Nullhypothese „behauptet“ meistens, dass es zwischen Gruppen oder Variablen keine Zusammenhänge oder Unterschiede gibt, die Alternativhypothese „behauptet“ meistens, dass es Zusammenhänge oder Unterschiede gibt. Deshalb steckt das, was der Untersucher/die Untersucherin glaubt, normalerweise in der Alternativhypothese. Die meisten Alternativhypothesen machen Aussagen über Zusammenhänge, Unterschiede oder Veränderungen.

In unserem Fragebogen werden die Personen u. a. gefragt, wie viele Stunden pro Woche sie sich körperlich betätigen (C1.4) und für wie sportlich sie sich halten (C1.5). Nun könnte man die Forschungshypothese aufstellen, dass körperliche Bewegung und subjektive Einschätzung der eigenen Sportlichkeit zusammenhängen. Das wäre in diesem Fall also die Alternativhypothese. Die dazugehörende Nullhypothese behauptet inhaltlich das Gegenteil: Körperliche Bewegung und subjektive Einschätzung der eigenen Sportlichkeit hängen nicht zusammen.

Wichtig ist: Diese Hypothesen beziehen sich nicht auf die Stichprobe, sondern auf die dahinterstehende **Grundgesamtheit (Population)**, die in der Regel unbekannt ist – denn wären die Verhältnisse in der Grundgesamtheit bekannt, bräuchte man über sie auch keine hypothetischen Überlegungen („Hypothesen" eben ...) anzustellen! Denken Sie an das einführende Nudelbeispiel: Kontrolliert ein Koch jede einzelne Nudel aus dem Topf (= Population), braucht er keine Hypothesen über diese Population aufzustellen – er kennt sie ja in diesem Fall. Zieht er aber nur eine Stichprobe von sagen wir zwei Nudeln (das ist dann seine Stichprobe) und sind diese zwei Nudeln gar, so weiß er nicht mit Sicherheit, ob alle Nudeln ebenso wie diese zwei gezogenen Nudeln gar sind – hier würde er dann hypothesengeleitet vorgehen, indem er die Alternativhypothese „Die Nudeln der Population sind gar" versus die Nullhypothese „Die Nudeln der Population sind nicht gar" aufstellt.

Bisher haben wir uns allerdings noch keine Gedanken darüber gemacht, wie wir die „Garheit" messen können. Dazu könnten wir eine subjektive Skala von 0 („hart und ungenießbar") bis 5 („weich und ungenießbar") aufstellen, wobei ein Wert von 3 optimal wäre. Wir könnten also zehn Nudeln ziehen (Stichprobe), den Mittelwert der „Garheit" ausrechnen und hätten so die Garheit operationalisiert.

Operationalisierung
Möchte man etwas messbar machen, muss zunächst definiert werden, was gemessen werden soll. Bei der Körpergröße etwa ist das leicht machbar. Aber wie definiert man das Konstrukt „romantische Liebe"? Man könnte sie beispielsweise als das Bedürfnis definieren, seinem Partner/seiner Partnerin rote Rosen zu schenken (ob das eine sinnvolle Definition ist, sei dahingestellt ...). Aber das allein reicht nicht aus. Es muss auch angegeben werden, durch welche beobachtbaren Ereignisse dieses Konstrukt erfasst, also gemessen werden kann. Etwa: Häufigkeit, mit der man der betreffenden Person rote Rosen schenkt. Hat man sich also auf eine Definition und messbare Ereignisse geeinigt, ist der Begriff „romantische Liebe" messbar gemacht, er ist operationalisiert. Der Begriff „romantische Liebe" wird operationalisiert durch die Häufigkeit, mit der man rote Rosen schenkt ...

Mittels der **Inferenzstatistik** werden also konkurrierende Hypothesen, die Null- und die Alternativhypothese, geprüft. Die Alternativhypothese kann auf zwei Arten formuliert werden: gerichtet oder ungerichtet. Ungerichtet ist eine Alternativhypothese, wenn keine Richtung des Zusammenhangs oder Unterschieds vorgegeben wird.

SPSS geht grundsätzlich von der Ungerichtetheit der Hypothesen aus.

Alternativhypothese (H_1): „Männer und Frauen unterscheiden sich hinsichtlich ihres durchschnittlichen wöchentlichen Bewegungspensums."

Nullhypothese (H_0): „Männer und Frauen unterscheiden sich *nicht* hinsichtlich ihres durchschnittlichen wöchentlichen Bewegungspensums."

Gerichtet ist eine Alternativhypothese, wenn etwas über die Richtung des erwarteten Zusammenhangs oder Unterschieds ausgesagt wird, wie folgendes Beispiel zeigt.

Alternativhypothese (H_1): „Männer machen durchschnittlich mehr Bewegung als Frauen."

Nullhypothese (H_0): „Männer machen durchschnittlich weniger oder höchstens gleich viel Bewegung wie Frauen."

7.2 Statistischer Test

Wenn Hypothesen formuliert sind, muss ein Weg gefunden werden, um zu entscheiden, welche der beiden konkurrierenden Hypothesen für „wahr" gehalten wird. „Wahr" heißt: Was wird wohl in der Population gelten? In der Population, die wir nicht kennen! Da wir die Population also nicht kennen, können Aussagen über sie **nicht mit Sicherheit** getroffen werden. Man könnte sagen: Ein Befund, der an einer Stichprobe erhoben wurde, soll auf seine **Allgemeingültigkeit** – über die konkrete Stichprobe hinaus – überprüft werden.

Ein statistischer Test ist das Mittel, um diese **Prüfung auf Allgemeingültigkeit** vorzunehmen. Es gibt mehrere Arten statistischer Tests. Nachdem statistische Hypothesen formuliert und ein Untersuchungsdesign festgelegt wurden, bietet sich folgende Grundstruktur an:
1. Erhebung empirischer Daten, eben z. B. durch Ausfüllen von Fragebogen
2. Berechnung von Statistiken aus diesen Daten (Stichprobenkennwerte), wie beispielsweise Mittelwerte
3. „Verpackung" dieser Stichprobenkennwerte nach bestimmten Regeln in bestimmte Formeln. Was dabei herauskommt, ist eine sogenannte „Teststatistik".
4. Berechnung, wie wahrscheinlich diese oder eine extremere Teststatistik ist, unter der Annahme, dass in der Population die Nullhypothese gilt
5. Wenn diese Wahrscheinlichkeit gering ist, „glaubt" man an die Alternativhypothese (signifikant); wenn diese Wahrscheinlichkeit groß ist, „glaubt" man weiterhin an die Nullhypothese (insignifikant).

In unserem Fragebogen wird neben dem Geschlecht (C1.1) auch erhoben, wie viele Stunden die Befragten mit körperlicher Bewegung verbringen (C1.4). Wir vermuten im Vorfeld, dass sich Männer und Frauen durchschnittlich hinsichtlich des Bewegungspensums unterscheiden, und formulieren deshalb ein Hypothesenpaar.

Alternativhypothese: „Männer und Frauen unterscheiden sich hinsichtlich ihres durchschnittlichen Bewegungspensums."
Nullhypothese: „Männer und Frauen unterscheiden sich *nicht* hinsichtlich ihres durchschnittlichen Bewegungspensums."

Als Untersuchungsdesign ergibt sich die Fragebogenerhebung.

1. Die Daten werden erhoben, indem Personen gebeten werden, den Bogen auszufüllen.
2. Für die Gruppe der Männer und Frauen werden jeweils Mittelwerte berechnet, was ihr wöchentliches Bewegungspensum betrifft.
3. Aus diesen Mittelwerten wird eine *Teststatistik* berechnet. *Im Wesentlichen* werden die Mittelwerte der beiden Gruppen voneinander subtrahiert, und wenn *in der Population die Nullhypothese gelten* sollte (kein Unterschied zwischen Männern und Frauen), erwarten wir eigentlich, dass bei dieser Subtraktion der Stichprobenmittelwerte der Wert „null" herauskommt – Werte weit entfernt von null wären in diesem Fall doch unwahrscheinlich ... Wichtig ist, dass die so berechnete Teststatistik eine gewisse Auftretenswahrscheinlichkeit hat, die berechenbar ist. Genau das macht das Statistikprogramm. Diese Auftretenswahrscheinlichkeit heißt „Signifikanz" oder „p-Wert" (von lat. probabilitas: Wahrscheinlichkeit).
4. Dieser p-Wert gibt also die Wahrscheinlichkeit an, diese aus den empirischen Daten berechnete Teststatistik zu erhalten, wenn – und das ist wesentlich! – in der Population, die wir ja nicht kennen, die Nullhypothese gilt. Mit anderen Worten: „Der P-Wert ist die Wahrscheinlichkeit, mit der man sich irrt, wenn man die Nullhypothese ablehnt" (Sachs, 1999, S. 188).
5. Wenn diese Wahrscheinlichkeit gering ist, entscheidet man sich für die Alternativhypothese. Was heißt aber „gering"? Dafür wurden willkürliche Grenzen festgelegt, man spricht vom Signifikanzniveau „Alpha" (α). Üblich sind Signifikanzniveaus von 5 %, 1 % und 0,1 % (0,05; 0,01; 0,001).

Wurde also vor der Untersuchung ein **Signifikanzniveau** von 5 % (oder: 0,05) festgelegt und beträgt der ermittelte p-Wert 0,02, so ist das Ergebnis auf dem 5 %-Niveau signifikant, was bedeutet, dass die Alternativhypothese angenommen wird. Mit anderen Worten: Höchstwahrscheinlich wird in der Population die Alternativhypothese gelten! Aber sicher wissen können wir es nicht, weshalb eine statistische Hypothese auch nicht streng bewiesen werden kann. Sicher wissen könnten wir es nur, wenn wir die Population kennen würden, aber dann bräuchten wir auch keine Hypothesen und somit auch keinen statistischen Test zur Überprüfung dieser Hypothesen!

Ermitteln wir einen p-Wert von $p = 0{,}02$, so ist das Ergebnis auf dem 5 %-Niveau signifikant, auf dem 1 %-Niveau nicht mehr. Mit anderen Worten: Nehmen wir bei $p = 0{,}02$ (2 %) also die Alternativhypothese an, besteht ein zweiprozentiges Risiko, eine falsche Entscheidung getroffen zu haben.

Ermitteln wir einen p-Wert von $p = 0{,}06$, so ist das Ergebnis weder auf dem 5 %- noch auf dem 1 %- noch auf dem 0,1 % -Niveau signifikant (signifikant wäre es auf dem 10 % -Niveau – ein derart hohes Signifikanzniveau wird allerdings selten gewählt, da hier die Wahrscheinlichkeit, einen Fehler erster Art zu begehen, zu groß wäre).

7.3 Fehler erster und zweiter Art und die Macht eines Tests

In der „wirklichen“ Welt gibt es also zwei Möglichkeiten: Es ist in der Population ein Effekt vorhanden, d. h. beispielsweise, dass sich die Männer und Frauen hinsichtlich ihres Bewegungspensums unterscheiden (dann gilt in der Population die Alternativhypothese), oder es ist kein Effekt vorhanden (Männer und Frauen unterscheiden sich nicht – es gilt die Nullhypothese). Die Inferenzstatistik ermöglicht uns eine Aussage darüber, wie wahrscheinlich die aus den Daten berechnete Teststatistik ist, unter der Annahme, dass in der Population die Nullhypothese gilt – oder, nicht ganz korrekt und etwas „salopp“ formuliert: welcher der beiden Fälle (Null- oder Alternativhypothese) vermutlich in der Population gelten wird („wahr“ ist). Da wir aber aufgrund der Stichprobenergebnisse nicht sicher wissen können, was „wahr“ ist, kann die Annahme der Alternativhypothese entweder richtig oder falsch sein (und umgekehrt gilt das natürlich auch für die Nullhypothese). Somit können zwei Fehler passieren:

Ein **Fehler erster Art** (oder Alphafehler) passiert, wenn wir an einen Unterschied (oder allgemein: Effekt) in der Population glauben, also die Alternativhypothese annehmen, obwohl sie (in der Population) nicht gilt.

Ein **Fehler zweiter Art** (oder Betafehler) passiert, wenn wir annehmen, es gäbe keinen Effekt in der Population, also die Nullhypothese beibehalten, obwohl sie (in der Population) nicht gilt.

Diese Möglichkeiten korrekter und falscher Entscheidungen sind gut in einem Vierfelderschema darstellbar (Tabelle 7.1).

Tab. 7.1: Testentscheidung und Wirklichkeit

		In der „Wirklichkeit“ gilt:	
		Nullhypothese	Alternativhypothese
Testentscheidung	Nullhypothese	Korrekte Entscheidung	Betafehler
	Alternativhypothese	Alphafehler	Korrekte Entscheidung

Es ist wichtig, zwischen einem konkreten, aus den Daten berechneten **p-Wert** und dem vor der Datenerhebung festgelegten **Signifikanzniveau Alpha** zu unterscheiden. Letzteres ist eine willkürlich bestimmte Grenze, die angibt, wie hoch die Wahrscheinlichkeit dafür höchstens sein darf, sich zu irren, wenn man die Alternativhypothese annimmt. Für diese Festlegung benötigt man noch keine konkreten Daten! Der konkrete p-Wert wird aus den Stichprobendaten berechnet und gibt die Wahrscheinlichkeit an, sich zu irren, wenn man die Alternativhypothese annimmt. Diese Wahrscheinlichkeit darf das vorher festgelegte Signifikanzniveau nicht überschreiten. Je mehr man sich davor schützen möchte, eine falsche

Alternativhypothese anzunehmen, desto strenger muss das Signifikanzniveau gewählt werden (man setzt also die Grenze dann nicht bei 5 %, sondern bei 1 % oder gar 0,1 % an). Der Nachteil dabei ist, dass mögliche **Effekte** größer sein müssen, damit ein konkreter p-Wert erreicht wird, um diese dann strengeren Signifikanzgrenzen zu „unterschreiten". Auf der anderen Seite kann das Signifikanzniveau auch nicht allzu streng gewählt werden, weil dann ein Effekt, sollte er in der Population tatsächlich vorhanden sein, leicht unentdeckt bleiben kann. Per Konvention sollte der Fehler zweiter Art nicht größer sein als 20 %. Diese Grenze bezeichnet man als Betafehler-Niveau.

Mit anderen Worten:

Angenommen, in der Population existiert kein Effekt (es gilt also die Nullhypothese): Wenn die gleiche Untersuchung hundertmal unabhängig voneinander durchgeführt wird, können wir erwarten, dass wir uns – bei einem Alphafehler-Niveau von 5 % – in circa fünf von diesen hundert Untersuchungen für die Annahme der Alternativhypothese entscheiden, obwohl in der Population die Nullhypothese gilt, d. h., wir finden einen Effekt, der in der Population nicht vorhanden ist. Mit dieser Fehlerwahrscheinlichkeit müssen wir leben ...

Angenommen, in der Population existiert ein Effekt: Wenn die gleiche Untersuchung hundertmal unabhängig voneinander durchgeführt wird, können wir erwarten, dass wir uns – bei einem Betafehler-Niveau von 20 % – in circa zwanzig von diesen hundert Untersuchungen für die Beibehaltung der Nullhypothese entscheiden, obwohl in der Population die Alternativhypothese gilt. Auch mit dieser Fehlerwahrscheinlichkeit müssen wir leben ... Das Betafehler-Niveau ist allerdings nicht so leicht festzulegen wie das Alphafehler-Niveau, da es von einer Reihe zumeist unbekannter Faktoren (wie etwa dem Populationsunterschied) abhängt.

Das Betafehler-Niveau ist eng mit der sogenannten **Macht eines Tests**, seiner **„Power"**, verbunden. Unter der Power versteht man das Vermögen, die „Kraft" eines statistischen Tests, eine in der Population gültige Alternativhypothese, auch zu erkennen. Im eben genannten Beispiel heißt das: Wenn wir in etwa zwanzig von hundert unabhängigen Untersuchungen die Nullhypothese fälschlich annehmen, nehmen wir in den restlichen achtzig Fällen eine korrekte Alternativhypothese an. Das ist die Power: In 80 % der Fälle entscheiden wir uns korrekterweise für die (richtige) Alternativhypothese.

Wenn das Ergebnis eines statistischen Tests, der zwei verschiedene Gruppen von Personen miteinander vergleicht (z. B. Männer und Frauen hinsichtlich ihrer Einstellung zu einem bestimmten Thema), signifikant ausfällt, heißt das also: Der Stichproben-Unterschied zwischen diesen beiden Gruppen wird höchstwahrscheinlich dadurch verursacht, dass es Unterschiede in den Populationen gibt!

7.4 Zusammenfassung des Kapitels

Die Inferenzstatistik verfolgt das Ziel, ausgehend von einer Stichprobe auf Verhältnisse in der Grundgesamtheit (Population) zu schließen. Dazu werden statistische Hypothesen formuliert: Die Nullhypothese behauptet zumeist, dass in der Population zwischen Gruppen oder Variablen keine Zusammenhänge bzw. Unterschiede bestehen. Die Alternativ- oder Forschungshypothese behauptet zumeist, dass es Zusammenhänge bzw. Unterschiede gibt.

Um eine Entscheidung darüber zu treffen, welche der beiden konkurrierenden Hypothesen für „wahr“ gehalten wird (es handelt sich stets um Wahrscheinlichkeitsaussagen, da die Verhältnisse in der Population nicht bekannt sind), werden statistische Tests eingesetzt. Ein statistischer Test ist also das Mittel, um diese Prüfung auf Allgemeingültigkeit vorzunehmen. Das Ergebnis eines statistischen Tests ist der sogenannte „p-Wert“. Dieser wird mit einem vorher festgesetzten Signifikanzniveau „Alpha“ verglichen, das zumeist 5 % beträgt. Ist der aus den Daten ermittelte p-Wert kleiner oder gleich Alpha, wird zugunsten der Alternativhypothese entschieden, andernfalls wird die Nullhypothese beibehalten.

Da die Verhältnisse in der Population also unbekannt sind und nur aufgrund von Stichprobenresultaten entschieden wird, sind die getroffenen Aussagen unsicher – es handelt sich um Wahrscheinlichkeitssaussagen. Deshalb kann es zu Fehlentscheidungen kommen. Der Fehler erster Art oder Alphafehler resultiert daraus, dass aufgrund des statistischen Tests zugunsten der Alternativhypothese entschieden wird, obwohl in der Population die Nullhypothese gilt. Der Fehler zweiter Art oder Betafehler resultiert daraus, dass aufgrund des statistischen Tests zugunsten der Nullhypothese entschieden wird, obwohl in der Population die Alternativhypothese gilt.

Die Macht eines Tests, die Power, bezeichnet sein „Vermögen“, einen in der Population bestehenden Effekt auch tatsächlich „aufzuspüren“.

7.5 Übungsbeispiele

Überprüfen Sie Ihr Wissen und versuchen Sie, die fünf Übungsbeispiele zu lösen:

1. Erklären Sie den Unterschied zwischen Deskriptiv- und Inferenzstatistik.
2. Erklären Sie die Begriffe „Nullhypothese“ und „Alternativhypothese“.
3. Was ist der Unterschied zwischen dem Signifikanzniveau Alpha und dem p-Wert?
4. Was versteht man unter einem statistischen Test?
5. Was versteht man unter Fehlern erster und zweiter Art und unter der Power?

Die Lösungen zu den Übungsbeispielen finden Sie im Anhang auf Seite 184 f.

Für die Wahl des statistischen Tests haben auch das **Skalenniveau** der interessierenden Variablen und die Frage, ob sie normalverteilt sind, Bedeutung. Sind die interessierenden Variablen intervallskaliert und normalverteilt (bei einigen Testverfahren wird auch noch Varianzgleichheit vorausgesetzt), kommen sogenannte **parametrische Verfahren** zum Zug, andernfalls wählt man **parameterfreie Verfahren**, welche nur die Ranginformation verwenden (wie bei der Berechnung des Medians). In diesem Buch werden die in Tabelle 8.3 angeführten Verfahren besprochen.

Tab. 8.3: Statistische Tests

Anzahl der Stichproben	**Art der Abhängigkeit**	**Skalen-niveau**	**Normal-verteilung**	**Verfahren**
2	unabhängig	metrisch	ja	t-Test für unabhängige Stichproben
2	abhängig	metrisch	ja	t-Test für abhängige Stichproben
2	unabhängig	ordinal	nein	U-Test nach Mann & Whitney
2	abhängig	ordinal	nein	Wilcoxon-Test
> 2	unabhängig	metrisch	ja	einfaktorielle Varianzanalyse (Kapitel 10)
> 2	abhängig	metrisch	ja	einfaktorielle Varianzanalyse mit Messwiederholung (Kapitel 10)
> 2	abhängig	ordinal	nein	Friedman-Test

Ein weiteres Testverfahren, das am Ende dieses Kapitels besprochen wird, ist der Chi-Quadrat-Test. Dieser wird eingesetzt, wenn es sich um Häufigkeitsdaten handelt.

8.1 T-Test für unabhängige Stichproben

Es sollen zwei unabhängige Stichproben hinsichtlich der abhängigen Variablen „Körpermasseindex“ (BMI) verglichen werden. Als unabhängige Variable (**Gruppierungsvariable**) dient die Einschätzung der eigenen Sportlichkeit (C1.5). Da diese in den Ausprägungen von 0 % bis 100 % vorliegt, wird die Variable zunächst dichotomisiert. Wer sich als bis zu 40 % sportlich einschätzt, kommt in die Gruppe 0, der Rest in die Gruppe 1. Die so gebildete neue Gruppierungsvariable wird „Sportlichkeit dichotom“ genannt und erhält den Variablennamen C1.5_1. Es muss also im ersten Schritt eine neue Variable C1.5_1 erstellt werden (das Umkodieren wurde bereits unter 5.2.6 erklärt – der folgende Punkt dient der Wiederholung dieser wichtigen SPSS-Funktion):

Transformieren – Umkodieren in andere Variable

Bringen Sie die Variable C1.5 in das Feld *Numerische Var. → Ausgabevar.*, geben Sie der neuen Variablen den Namen *C1.5_1* und klicken Sie auf *Ändern*. Dann klicken Sie auf *Alte und neue Werte*, markieren das Feld *Bereich: Kleinster Wert bis* und geben den Wert *40* ein. Bei „neuer Wert“ tragen Sie *0* ein und klicken auf *Hinzufügen*. Abschließend markieren Sie *Alle* *anderen Werte*, geben als neuen Wert *1* ein und klicken auf *Hinzufügen*. Schließen Sie mit einem Klick auf *Weiter* ab und starten Sie die Umkodierung mit einem Klick auf *OK* (Abb. 8.1). Es wird eine neue Variable C1.5_1 angelegt, die allen Fällen, welche einen Wert bis 40 angegeben haben, den neuen Wert *0* und allen anderen den Wert *1* zuweist. Nun haben wir eine dichotome Gruppierungsvariable erstellt. Vergeben Sie als Variablenlabel „Sportlichkeit dichotom" und als Wertelabel für *0*: „gering" und für *1*: „hoch".

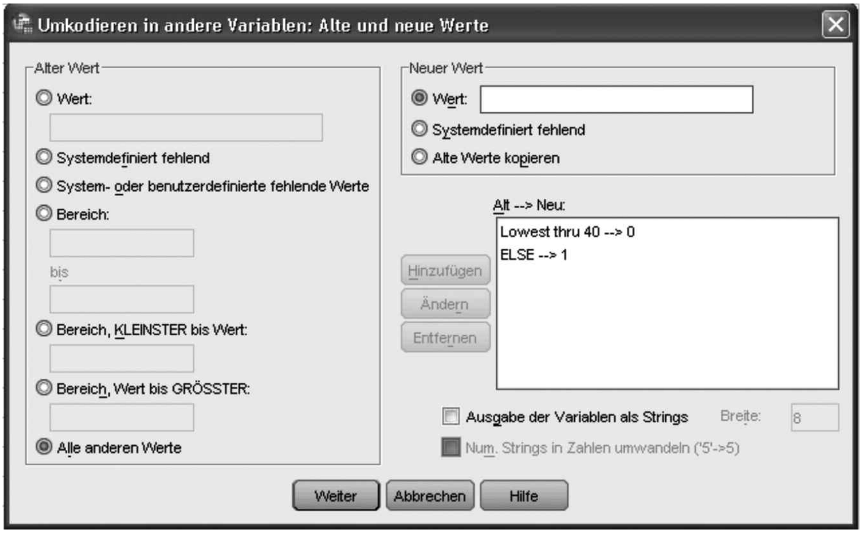

Abb. 8.1a und Abb. 8.1b: Erstellen einer dichotomen Variablen

Die abhängige Variable ist der Body-Mass-Index (Körpergewicht in Kilogramm, dividiert durch Größe in Meter zum Quadrat). Diesen berechnen Sie in SPSS folgendermaßen (Abb. 8.2):

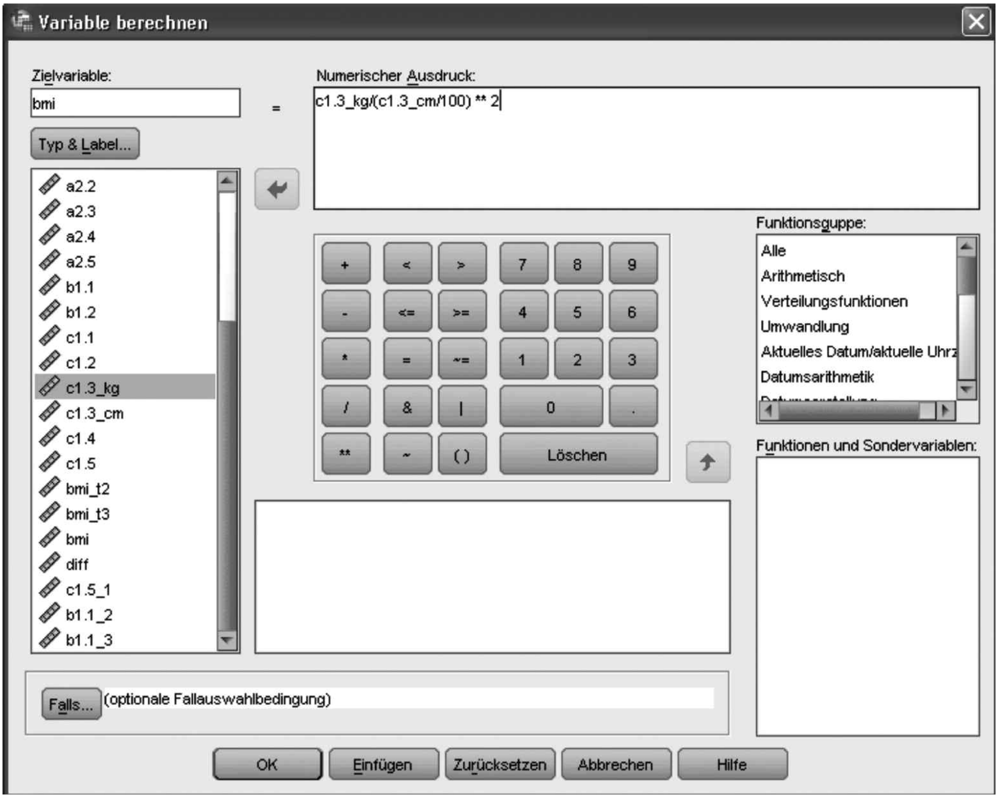

Abb. 8.2: Berechnen des BMI

Transformieren – Variable berechnen (bis Version 17: *Berechnen*)

Geben Sie in das Fenster den numerischen Ausdruck wie in Abbildung 8.2 ein und klicken Sie auf *OK* (die Variable C1.3_cm enthält die Körpergröße in cm, die Variable C1.3_kg das Körpergewicht in kg). Da die Körpergröße im Datenfile in Zentimetern vorhanden ist, für die Berechnung des BMI aber die Körpergröße in Meter notwendig ist, muss durch 100 dividiert werden.

Der t-Test für unabhängige Stichproben vergleicht die Mittelwerte zweier Stichproben. Die Messwerte müssen normalverteilt und die Varianzen in den beiden Stichproben homogen sein, d. h., die Varianzen dürfen sich **nicht signifikant** voneinander unterscheiden. Die ungerichtete Nullhypothese des unabhängigen t-Tests lautet: „Die Mittelwerte in den Populationen unterscheiden sich nicht." Und die dazu korrespondierende Alternativhypothese lautet: „Die Mittelwerte in den Populationen unterscheiden sich."

Aufgrund der strengen Voraussetzungen müssen die Daten vor Berechnung des t-Tests auf **Normalverteilung** (bei kleineren Stichproben vorzugsweise mit dem Kolmogorov-Smirnov-Test; siehe beispielsweise Bortz, 2005) geprüft werden (die Prüfung auf Normalverteilung muss für beide Gruppen erfolgen, weshalb die Datei zunächst nach der neuen Variablen „Sportlichkeit dichotom“ [C1.5_1] geteilt werden muss). Dies geschieht folgendermaßen:

Der **Kolmogorov-Smirnov-Test** (auch in der Schreibweise „Kolmogoroff-Smirnoff-Test“ zu finden) ist ein verteilungsunabhängiger Test, der besonders bei kleinen Stichprobenumfängen Abweichungen von der Normalverteilung entdeckt. Er setzt stetige Verteilungen voraus, kann jedoch auch bei diskreten Verteilungen angewandt werden und prüft die Nullhypothese, dass die Stichprobe einer normalverteilten Grundgesamtheit entstammt, im Gegensatz zur Alternativhypothese, dass die Stichprobe nicht einer normalverteilten Grundgesamtheit entstammt. Klicken Sie in SPSS auf:

Daten – Datei aufteilen

Markieren Sie *Ausgabe nach Gruppen aufteilen* und bringen Sie die Variable *C1.5_1* in das Feld *Gruppen basierend auf*. Nach Klick auf *OK* werden nun alle folgenden Berechnungen für die beiden Gruppen getrennt ausgeführt (rechts unten im SPSS-Fenster erscheint der Hinweis *Datei aufteilen an* – bis Sie die Aufteilung wieder aufheben).

Um den Kolmogorov-Smirnov-Test durchzuführen, gehen Sie so vor:

SPSS: *Analysieren – Nichtparametrische Tests – Alte Dialogfelder – K-S bei einer Stichprobe*

Bringen Sie die Variable *BMI* in das Variablenfeld und klicken Sie auf *OK* (der K-S-Test ist voreingestellt). Tabelle 8.4 zeigt den Output für die Gruppe „Sportlichkeit gering“.

Tab. 8.4: Output Kolmogorov-Smirnov-Test

Kolmogorov-Smirnov-Anpassungstest[c]

		BMI zum Zeitpunkt Fragebogen-vorgabe
N		14
Parameter der Normalverteilung[a, b]	Mittelwert	25,1705
	Standardabweichung	1,22169
Extremste Differenzen	Absolut	,155
	Positiv	,155
	Negativ	-,110
Kolmogorov-Smirnov-Z		,581
Asymptotische Signifikanz (2-seitig)		,888

a. Die zu testende Verteilung ist eine Normalverteilung.
b. Aus den Daten berechnet
c. Sportlichkeit = gering

Die **Nullhypothese des K-S-Tests** lautet, dass die Variable in der Grundgesamtheit normalverteilt ist, weshalb hier ein nichtsignifikantes Ergebnis wünschenswert ist. Der p-Wert beträgt p = 0,888, sodass die Nullhypothese beibehalten werden kann. Für die Gruppe „Sportlichkeit hoch" beträgt der p-Wert p = 0,952. In beiden Gruppen kann von Normalverteilung in der Grundgesamtheit ausgegangen werden, womit auch diese Voraussetzung für den t-Test erfüllt ist. Heben Sie nun die Aufteilung der Datei wieder auf, indem Sie im Menü *Daten – Datei aufteilen* auf *Zurücksetzen* und anschließend auf *OK* klicken. Der Hinweis *Datei aufteilen an* verschwindet nun.

Jetzt kann der t-Test berechnet werden:

Analysieren – Mittelwerte vergleichen – T-Test bei unabhängigen Stichproben

Die Gruppenvariable ist in unserem Beispiel die Variable *Sportlichkeit dichotom* (*C1.5_1*) – ziehen Sie diese also in das Feld *Gruppenvariable*. Wenn Sie auf *Gruppen definieren* klicken, kommen Sie zu einem Dialogfeld, in dem die Werte für die jeweilige Gruppe angegeben werden müssen: *0* und *1*. Schließen Sie nun das Dialogfenster. Die Testvariable ist *BMI*. Klicken Sie auf *OK*, um den t-Test-Output zu erhalten (Tab. 8.5 und 8.6 – der Output wurde aus Gründen der Übersichtlichkeit reduziert).

Tab. 8.5: Output t-Test: Gruppenstatistiken

Gruppenstatistiken

	Sportlichkeit	N	Mittelwert	Standard-abweichung
BMI zum Zeitpunkt Fragebogenvorgabe	gering	14	25,1705	1,22169
	hoch	6	22,1192	1,41096

Tab. 8.6: Output t-Test: Levene-Test und t-Test für die Mittelwertgleichheit

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit		
		F	Signifikanz	T	df	Sig. (2-seitig)
BMI zum Zeitpunkt Fragebogen-vorgabe	Varianzen sind gleich	,102	,754	4,897	18	,000
	Varianzen sind nicht gleich			4,608	8,396	,002

Die Tabelle **„Gruppenstatistiken"** liefert deskriptivstatistische Resultate: Sie können hier ablesen, welche Mittelwerte und Standardabweichung die beiden Gruppen aufweisen. In der Gruppe „Sportlichkeit gering" beträgt der Mittelwert des BMI 25,17, in der Gruppe „Sportlichkeit hoch" beträgt er 22,12. Das eigentliche Ergebnis des t-Tests zeigt die Tabelle 8.6. Zuerst müssen Sie allerdings den Levene-Test der Varianzgleichheit betrachten, da er Information darüber liefert, ob davon ausgegangen werden kann, dass in der Grundgesamtheit der „Sportlichen“ und „Nichtsportlichen“ (diese Grundgesamtheiten werden durch unsere Dichotomisierung sozusagen „künstlich“ erzeugt) die Varianzen gleich sind. Die Hypothesen lauten:

Nullhypothese: „Die Varianzen in den beiden Grundgesamtheiten sind gleich (homogen).“
Alternativhypothese: „Die Varianzen in den beiden Grundgesamtheiten sind nicht gleich (inhomogen).“

Wir hoffen, dass die Nullhypothese beibehalten werden kann, da andernfalls die Ergebnisse des t-Tests nicht interpretiert werden dürfen. Der p-Wert, der das Entscheidungskriterium dafür liefert, ist in der Spalte *Signifikanz* zu finden. Ist er größer als 0,05 (bzw. 5 %), wird die Nullhypothese beibehalten und der t-Test darf interpretiert werden, was hier der Fall ist: Der p-Wert des Levene-Tests beträgt p = 0,754. Deshalb schauen wir in der Zeile *Varianzen sind gleich* zum eigentlichen t-Test.

Für den t-Test wurde ebenfalls ein Hypothesenpaar formuliert, das sich auf mögliche Populationsunterschiede den mittleren BMI betreffend bezieht (jetzt geht es nicht mehr um das Hypothesenpaar des Levene-Tests, sondern um jenes des t-Tests!).

In der Spalte **Signifikanz** (das ist der p-Wert) ist abzulesen, ob die Stichproben-Differenz „überzufällig“ ist oder nicht. Das heißt: Beträgt der p-Wert höchstens 5 %, ist die Stichproben-Differenz wahrscheinlich nicht mehr nur durch Zufall entstanden, sondern resultiert daher, dass die „Sportlichen“ und „Nichtsportlichen“ eigene Populationen bilden, die sich hinsichtlich des BMI tatsächlich unterscheiden. In unserem Beispiel ist das der Fall (p = 0,000), wir entscheiden uns also für die Annahme der Alternativhypothese.

8.2 T-Test für abhängige Stichproben

Nehmen wir an, alle Personen hätten nach dem Ausfüllen des Fragebogens ein Ernährungs- und Bewegungsprogramm durchgeführt und nach acht Wochen wäre ihr Gewicht ein zweites Mal erfasst und der BMI berechnet worden – wir hätten also einen „Vorher-nachher-Vergleich“ und somit eine „abhängige“ Stichprobe (diese Daten sind in der Spalte *bmi_t2* des Datenfiles enthalten).

Es kann also eine **Differenz** berechnet werden (BMI zu den Zeitpunkten 1 und 2), und diese Differenz über alle Personen hat natürlich auch einen Mittelwert (den Mittelwert der Differenzen) (Tab. 8.7). Wenn das Ernährungs- und Bewegungsprogramm keine Auswirkungen hätte, müsste diese Differenz eigentlich 0 sein: Das würden wir unter der Nullhypothese erwarten! Doch auch in diesem Fall werden wir nicht genau 0 beobachten, denn selbst wenn das Programm nichts bewirkt, bleiben die BMIs der Personen nicht zu den beiden Zeitpunkten identisch, Abweichungen von 0 wird es geben. Das (zweiseitig formulierte) Hypothesenpaar des abhängigen t-Tests lautet:

Nullhypothese: „Der ‚wahre‘ Mittelwert der Differenzen ist Null.“
Alternativhypothese: „Der ‚wahre‘ Mittelwert der Differenzen ist ungleich Null.“

utb. Schlüsselkompetenzen

Wie plant man eine empirische Erhebung? Wie setzt man die Idee um? Wie werden die Daten analysiert und interpretiert? In gut nachvollziehbaren Schritten bietet dieses Lehrbuch in der 5., aktualisierten und überarbeiteten Auflage einen Leitfaden für die Umsetzung wissenschaftlicher Erhebungen: von der Forschungsidee über die Konstruktion eines Fragebogens bis hin zu den wichtigsten Auswertungsschritten mit dem Statistikprogramm SPSS, Version 24, in unterschiedlichen Schwierigkeitsgraden. Ein durch viele Beispiele sehr anschauliches und gut verständliches Lehrbuch!

Dies ist ein utb-Band aus dem Verlag facultas. utb ist eine Kooperation von Verlagen mit einem gemeinsamen Ziel: Lehrbücher und Lernmedien für das erfolgreiche Studium zu veröffentlichen.

ISBN 978-3-8252-8727-6

9 783825 287276

QR-Code für mehr Infos und Bewertungen zu diesem Titel

utb-shop.de