6 Die Logik des statistischen Schließens

Population, Grundgesamtheit, Verteilungsformen, Normalverteilung und Standardnormalverteilung – nachdem diese Begriffe im vorangehenden Kapitel eingeführt wurden, schreiten wir jetzt weiter voran ins Zentrum der Inferenzstatistik. In diesem Kapitel geht es um Schätzen und Testen, um Regelmäßigkeiten in der Verteilung von Stichprobenparametern, um Nullhypothesen und Alternativhypothesen und um das Grundprinzip des statistischen Schließens.

Statistische Begriffe haben immer mehr in unsere Alltagswelt Einzug gehalten. Täglich sind wir mit Meldungen und Informationen konfrontiert, in denen von *signifikanten Ergebnissen in repräsentativen Studien* die Rede ist – und sogleich sind wir geneigt, dem Gesagten Glauben zu schenken, scheint es doch durch dieses Zauberwort quasi bewiesen („signifikant"!) und durch die Hinweise auf die Repräsentativität der Stichprobe auch gegen jeden Verdacht des Singulären immunisiert. Mit statistischen Schlüssen verhält es sich aber gar nicht so einfach, wie wir dies im Alltag bei der Lektüre solcher Forschungsergebnisse vielleicht gerne glauben möchten. Das statistische Schließen besitzt eine Reihe von Voraussetzungen, Stolperfallen und Unsicherheiten. Die Basis von alldem sind Grundannahmen über die Verteilung von Stichprobenkennwerten.

6.1 Die Verteilung von Stichprobenkennwerten

Das Ziel der Inferenzstatistik ist es, von den bekannten Kennwerten einer Stichprobe (z.B. Mittelwert, Varianz) auf die unbekannten Parameter einer Grundgesamtheit zu schließen. Warum stellt sich dieses Problem überhaupt? Der wichtigste Grund ist, dass es wegen des immensen Aufwandes praktisch unmöglich und unbezahlbar ist, die entsprechenden interessierenden Kennwerte der Grundgesamtheit zu ermitteln. Wir können nicht die Mathematikkompetenzen aller deutschen Schülerinnen und Schüler mithilfe eines Tests untersuchen und sie anschließend mit den Schülerinnen und Schülern der anderen 26 EU-Länder vergleichen. Selbst ein so einfach zu erhebendes Merkmal wie Körpergröße oder Gewicht würde bei 80 Millionen Probanden in Deutschland mit riesigen Erhebungskosten verbunden sein, sodass sich jeder Gedanke an eine solche Studie von vornherein verbietet. Eine Vollerhebung der Grundgesamtheit ist aber auch – der statistischen Theorie sei Dank – relativ überflüssig, weil man aus einer ge-

U. Kuckartz et al., *Statistik,* DOI 10.1007/978-3-531-19890-3_6,

nügend großen Stichprobe die gewünschten Informationen mit einer hinreichenden Genauigkeit schließen (schätzen) kann.

Um die Problematik von Stichprobe und Grundgesamtheit zu verdeutlichen, wollen wir zunächst einen kleinen virtuellen Versuch starten, und zwar nicht mit einer unbekannten, sondern einer bekannten Grundgesamtheit, nämlich den 2.034 Personen, die in der Studie „Umweltbewusstsein in Deutschland"[10] befragt wurden[11]. Diese Personen sollen für die folgenden Überlegungen die Grundgesamtheit darstellen, ihr Durchschnittsalter beträgt 48,9 Jahre. Wir ziehen nun aus dieser Grundgesamtheit eine Zufallsstichprobe von $n = 100$ Personen und ermitteln für diese Stichprobe das Durchschnittsalter. Das Ziehen einer Stichprobe könnte man nach dem Urnenmodell[12] per Hand realisieren, bequemer geschieht dies mit einem Statistikprogramm, das in der Lage ist, Zufallsstichproben gewünschter Größe zu ziehen. Für diesen Versuch haben wir die Stichprobe mithilfe von SPSS gezogen; das Ergebnis zeigt, dass die jüngste Person unserer 100er Stichprobe 19 Jahre alt ist und die älteste 77. Der Altersmittelwert beträgt 51,1 Jahre und die Standardabweichung 15,97 Jahre.

Versetzen wir uns nun in die Lage, auf der Basis dieser Stichprobe den Altersdurchschnitt aller 2.034 Befragten erraten zu müssen: Wie würden wir vorgehen? Da man keinen anderen Wert als diesen Stichprobenmittelwert von 51,1 Jahren zur Verfügung hat, ist es naheliegend, eben diesen als Schätzwert des Altersmittelwerts der Grundgesamtheit zu benutzen und den Mittelwert der Grundgesamtheit ebenfalls mit 51,1 Jahren zu veranschlagen. Was wir bis zu diesem Punkt getan haben, bezeichnet man als Parameterschätzung: Der uns unbekannte Parameter „Mittelwert der Grundgesamtheit" wurde aus einem Stichprobenmittelwert geschätzt. Bei diesem Versuch kennen wir aber den „wirklichen" Mittelwert der 2.034 Personen, der also durch die Schätzung um 1,2 Jahre verfehlt würde.

Wir starten nun einen zweiten Versuch und ziehen erneut eine Stichprobe von $n = 100$ Personen. Das minimale Alter beträgt nun laut SPSS-Ausdruck 19 Jahre, das maximale 83. Als Mittelwert wird 49,6 (Standardabweichung 17,95) berechnet. Diese Stichprobe ist also etwas jünger als beim ersten Versuch. Um die Sache an dieser Stelle abzukürzen, haben wir die Ergebnisse unserer ersten beiden und acht weiterer Versuche in Tab. 6-1 zusammengestellt.

10 Siehe Kuckartz u.a. 2006

11 Wer den Versuch selbst ausprobieren will, kann die Daten der Studie in Form einer SPSS-Datei unter www.statistik-verständlich.de herunterladen.

12 Hierzu müssten sich alle 2034 Personen ähnlich wie die 49 Lottokugeln in einer Urne befinden, aus der sie dann zufällig gezogen würden.

Tab. 6-1: Mittelwerte von Zufallsstichproben der Größe *n* = 100

Versuch	Mittelwert der Stichprobe	Abweichung vom Mittelwert der Grundgesamtheit (48,9)
1	51,1	+2,2
2	49,6	+0,7
3	46,3	−2,6
4	49,4	+0,5
5	50,0	+1,1
6	45,3	−3,6
7	47,6	−1,3
8	49,8	+0,9
9	45,9	−3,0
10	48,7	−0,2

Die Werte der zweiten Spalte beinhalten die Verteilung des Stichprobenparameters Mittelwert für eine gegebene Stichprobengröße (hier $n = 100$). Für die Streuung dieser Verteilung lassen sich auch Varianz und Standardabweichung berechnen, wobei zu beachten ist, dass es sich dabei um etwas anderes handelt als um die Streuung der 100 Messwerte. Die dritte Spalte gibt an, wie weit die jeweiligen Stichprobenmittelwerte vom Mittelwert der Grundgesamtheit abweichen. Es ist erkennbar, dass sich der Mittelwert in vier der zehn Stichproben (Nr. 1, 3, 6, 9) um mehr als 2 Jahre vom Populationsmittelwert unterscheidet.

Wir führen nun eine zweite Versuchsreihe durch, diesmal mit einem größeren Stichprobenumfang von $n = 300$ Personen. Es ergeben sich die in Tab. 6-2 dargestellten Mittelwerte und Abweichungen:

Tab. 6-2: Mittelwerte von Zufallsstichproben der Größe *n* = 300

Versuch	Mittelwert der Stichprobe	Abweichung vom Mittelwert der Grundgesamtheit (48,9)
1	48,2	−0,7
2	49,0	+0,1
3	48,6	−0,3
4	48,9	0
5	50,1	+1,2
6	49,5	+0,6
7	48,6	−0,3
8	50,3	+1,4
9	48,9	0
10	48,2	−0,7

Auf den ersten Blick ist schon zu erkennen, dass die Stichprobenmittelwerte weniger schwanken als zuvor bei der kleineren Stichprobe mit $n = 100$ und die maximale Abweichung vom Populationsmittelwert beträgt nur noch 1,4 Jahre. Würde man die Anzahl der gezogenen Stichproben nun immer weiter erhöhen, so würde sich die Verteilung der Stichprobenmittelwerte immer mehr der Normalverteilung annähern, wobei der Mittelwert dem Mittelwert der Grundgesamtheit entsprechen würde.

Diese Erkenntnis führt uns zu einem der zentralen Lehrsätze der schließenden Statistik, dem *zentralen Grenzwertsatz*, der im Kern besagt, dass Stichprobenkennwerte, z.B. Mittelwerte, normalverteilt sind.[13]

> Die Mittelwerte von hinreichend großen Stichproben ($n \geq 30$) verteilen sich normal um μ, den Mittelwert der Grundgesamtheit. Diese Verteilung ist unabhängig von der Verteilung der Werte der Grundgesamtheit, d.h. diese müssen nicht normalverteilt sein.

Für kleine Stichproben ($n < 30$) folgen Stichprobenkennwerte nicht der Normalverteilung sondern der t-Verteilung mit $df = n-1$ Freiheitsgraden.

13 Genau genommen besagt das zentrale Grenzwerttheorem, dass Summen von Zufallsvariablen normalverteilt sind. Die n Messwerte x_i einer Stichprobe werden dabei als voneinander unabhängige Realisierung einer Zufallsvariablen betrachtet.

Stichprobenkennwerte weisen wie oben gezeigt eine Streuung auf, dabei wird die Standardabweichung des Stichprobenmittelwerts als *Standardfehler* oder Stichprobenfehler des Mittelwerts bezeichnet. Der Standardfehler lässt sich nach folgender Formel aus der Varianz bzw. der Standardabweichung der Messwerte schätzen. Da die Varianz der Grundgesamtheit nicht bekannt ist, wird die Stichprobenvarianz als Schätzer benutzt. Der Buchstabe σ ist in der Formel mit einem „^" versehen, da mit dieser Schreibweise in der Statistik üblicherweise Schätzungen von unbekannten Werten gekennzeichnet werden.

$$\text{Standardfehler} \quad \hat{\sigma}_{\bar{x}} = \sqrt{\frac{s^2}{n}} = \frac{s}{\sqrt{n}}$$

s^2 = Stichprobenvarianz als Schätzer für die Populationsvarianz
n = Anzahl der Fälle in der Stichprobe
s = Standardabweichung der Stichprobe als Schätzer für die Population

Für unser allererstes Experiment mit der 100er-Stichprobe beträgt der Standardfehler also:

$$\hat{\sigma}_{\bar{x}} = \sqrt{\frac{15{,}97^2}{100}} = \frac{15{,}97}{10} = 1{,}597$$

Je größer der Standardfehler ist, desto unsicherer ist die Schätzung des Mittelwerts, d.h. die Genauigkeit der Schätzung hängt entscheidend von der Streuung der Variablenwerte ab: Je kleiner die Streuung der Messwerte, desto genauer die Schätzung, je größer die Streuung desto unpräziser.

6.2 Konfidenzintervalle

Dank des zentralen Grenzwertsatzes und der Möglichkeit, unbekannte Parameter der Grundgesamtheit zu schätzen, können wir nun die in Kapitel 5.5 dargelegten Erkenntnisse über die Eigenschaften der Normalverteilung und insbesondere der Standardnormalverteilung kreativ anwenden. Dort war gezeigt worden, wie man zunächst einen Messwert in einen z-Wert transformieren und dann mithilfe der Standardnormalverteilung bestimmen kann, wie groß der Anteil der Verteilung ist, der in einem bestimmten Messwertebereich liegt. Für jede Person der Stichprobe lässt sich auf diese Weise angeben, wie viele Einheiten (in Standardabweichungen gemessen) sie vom Mittelwert positiv oder negativ entfernt ist. In dem in Kapitel 5.5 dargestellten Beispiel hatte der vierjährige Nils für das Zusammenlegen eines Puzzles einen z-Wert von −1 erreicht und war damit eine Standardabweichung schneller als der Durchschnitt. Ähnliche Berechnungen lassen sich auch für Stichprobenkennwerte durchführen, um so bspw. zu ermitteln, innerhalb welchen Wertebereichs wie viel Prozent der Kennwerte liegen.

Bei der Schätzung von Parametern der Grundgesamtheit ist man nicht nur an der Angabe eines ganz bestimmten Wertes (Punktschätzung) interessiert, sondern man möchte auch wissen, innerhalb welchen Bereichs der Parameter mit großer Sicherheit liegt (Intervallschätzung).

Mit dem Begriff *Konfidenzintervall* bezeichnet man den Bereich, innerhalb dessen ein bestimmter Prozentsatz aller möglichen Populationsparameter liegt.[14] Am häufigsten wird das 95%-Konfidenzintervall berechnet, gelegentlich auch das 99%-Intervall. Konfidenzintervalle werden gelegentlich auch Vertrauensintervalle genannt.

Wie berechnet man nun das 95%-Konfidenzintervalls für den Mittelwert? Die allgemeine Formel lautet für die untere bzw. obere Grenze:

Untere Grenze: $\mu - z \cdot \hat{\sigma}_{\bar{x}}$
Obere Grenze: $\mu + z \cdot \hat{\sigma}_{\bar{x}}$

Im Fall unseres Beispiels kennen wir den Mittelwert der Grundgesamtheit (n = 2.034), er beträgt 48,9 Jahre. Meist sind die Parameter der Grundgesamtheit nicht bekannt, dann verwendet man die Stichprobenwerte als Schätzung. Zur Berechnung der konkreten Unter- und Obergrenze ersetzen wir z in den obigen Formeln durch die entsprechenden z-Werte der Standardnormalverteilung.

Zu diesem Zwecke betrachten wir die Tabelle der Standardnormalverteilung im Anhang B1 (S. 305). Dort informiert die auf der rechten Seite abgebildete Tabelle darüber, welcher z-Wert eine bestimmte Fläche auf der linken Seite abschneidet. Beim 95%-Konfidenzintervall werden links und rechts jeweils 2,5%, insgesamt also 5% der Fläche abgeschnitten. Wir suchen folglich nach dem z-Wert, der 97,5% der Fläche abschneidet und entnehmen der Tabelle den Wert 1,96. Das bedeutet, dass sich 95% der Werte zwischen z = −1,96 und z = +1,96 befinden, d.h. wir setzen 1,96 in die Formeln für die Grenzen des Intervalls ein. Für das Konfidenzintervall ergibt sich somit folgende Rechnung:

14 Korrekterweise müsste man eigentlich formulieren, dass in x% aller Fälle, in denen man eine Stichprobe zieht, der betreffende Parameter innerhalb des Konfidenzintervalls liegt.

Untere Grenze: $\bar{x} - 1{,}96 \cdot \hat{\sigma} = 48{,}9 - 1{,}96 \cdot 1{,}597 = 45{,}8$
Obere Grenze: $\bar{x} + 1{,}96 \cdot \hat{\sigma} = 48{,}9 + 1{,}96 \cdot 1{,}597 = 52{,}0$

In Worten bedeutet dies „In 95% aller Fälle, in denen eine Stichprobe dieser Größe ($n = 100$) aus der Grundgesamtheit gezogen wird, liegt der Mittelwert zwischen 45,8 und 52,0“.

Um ein anderes Konfidenzintervall zu berechnen, sind lediglich die der Standardnormalverteilungstabelle (vgl. Anhang B1, S. 305) zu entnehmenden z-Werte anstelle von 1,96 einzusetzen, z.B. 2,58 für die Berechnung des 99%-Konfidenzintervalls, denn hier werden links und rechts jeweils 0,5% der Fläche abgeschnitten:

Untere Grenze: $\bar{x} - 2{,}58 \cdot \hat{\sigma} = 48{,}9 - 2{,}58 \cdot 1{,}597 = 44{,}8$
Obere Grenze: $\bar{x} + 2{,}58 \cdot \hat{\sigma} = 48{,}9 + 2{,}58 \cdot 1{,}597 = 53{,}0$

Weshalb benötigt man überhaupt Konfidenzintervalle? Der entscheidende Punkt ist, dass die Inferenzstatistik kein sicheres Wissen produziert, wobei die Größe der Stichprobe, mit der man arbeitet, letzten Endes gleichgültig ist. Sei die Stichprobe nun gleich 1.000, 2.000, 5.000 oder 10.000 Personen, man kann die Prozentanteile der Parteien in der Grundgesamtheit „Wahlberechtigte in der Bundesrepublik Deutschland“ nicht *sicher* ermitteln.

Was die statistischen Verfahren aber können, ist es, anzugeben innerhalb welcher Spannweite der Wert in der Grundgesamtheit mit welcher Wahrscheinlichkeit liegt, also bspw., dass der gesuchte Wert „Prozentanteil der Männer in deutschen Pädagogik-Studiengängen“ derzeit mit 95%iger Wahrscheinlichkeit zwischen 18,0 und 20,2% und mit 99%iger Wahrscheinlichkeit zwischen 16,5 und 21,7% liegt. Diese Konfidenz- oder Vertrauensintervalle kann man ebenso wie die Wahrscheinlichkeit exakt berechnen, d.h. man kann darauf vertrauen, dass sich der interessierende Wert mit einer bestimmten Wahrscheinlichkeit in dem bezeichneten Intervall befindet. Sicher sein kann man jedoch keineswegs, denn im Umkehrschluss gibt es auch immer eine bezifferbare Wahrscheinlichkeit für den umgekehrten Fall, dass der Wert nämlich außerhalb des Vertrauensbereichs liegt. Daraus nun den Schluss zu ziehen, die Irrtumswahrscheinlichkeit möglichst klein zu halten, indem man stets das 99%- oder 99,9%-Konfidenzintervall berechnet, stellt allerdings auch keine Lösung dann, denn dies bewirkt nur, dass das Vertrauensintervall sehr breit wird und damit keinen Informationswert mehr besitzt.

6.3 Die statistische Hypothese

Die Logik des statistischen Hypothesentestens baut auf den Verteilungseigenschaften statistischer Kennwerte und dem zentralen Grenzwertsatz auf. Auch im Alltag ziehen wir Schlüsse und formulieren Hypothesen: Zum Beispiel gehen wir als Erstsemester/in um 10 Uhr in eine Vorlesung, deren Beginn für 10 Uhr angekündigt ist und stellen fest, dass der Hörsaal noch ziemlich leer ist und die Lehrkraft, ohne irgendeine Spur von Hektik zu zeigen, noch mit der Inbetriebnahme der Beameranlage beschäftigt ist. Daraus leiten wir die Hypothese ab, dass entweder unsere Uhr falsch geht oder die Vorlesung doch erst später beginnt, als es im Vorlesungsverzeichnis abgedruckt ist. Von der ersten Hypothese rücken wir aber gleich wieder ab, als wir sehen, dass die Uhr im Hörsaal exakt die gleiche Uhrzeit wie die eigene Armbanduhr anzeigt. Der verspätete Beginn hat also offenbar nichts mit einem Defekt unserer Uhr zu tun. Wir entscheiden uns deshalb für die zweite Variante und fragen die Kommilitonin in der Reihe vor uns, was hier eigentlich los sei und erfahren, dass Vorlesungen und Seminare normalerweise erst eine Viertelstunde später als angekündigt beginnen. Beim nächsten Mal werden wir also etwas später in die Vorlesung gehen.

Alltagshypothesen unterscheiden sich grundlegend von statistischen Hypothesen: Während es dort ein Alltagsproblem ist, dessen Lösung wir mit einiger Tüftelei, mit Fragen und Trial-and-error herausfinden können, handelt es sich bei statistischen Hypothesen um formalisierte Aussagen, um eine möglichst präzise Annahme, die wir mit statistischen Mitteln – und zwar einem Kalkül auf der Basis von Verteilungsannahmen – systematisch überprüfen.

Die Grundeinheit einer statistischen Hypothese ist die *Variable*. Eine mit den Mitteln der Statistik zu prüfende Hypothese ist immer als eine präzise Aussage zu formulieren, in der ein Zusammenhang zwischen mindestens zwei Variablen behauptet wird. Dabei besitzt eine Variable immer mindestens zwei Ausprägungen – ansonsten wäre es ja auch keine *Variable*, sondern eine Konstante. Jede Hypothese muss so formuliert sein, dass sie auch scheitern kann, d.h. sich empirisch im Rahmen der durchgeführten Studie und auf der Basis der dafür erhobenen Daten als falsch erweisen kann.

Eine einfache Hypothese könnte etwa lauten: Frauen sind klimabewusster als Männer. Es ergibt sich eine Art Vier-Felder-Tafel:

1 klimabewusste Männer	2 klimabewusste Frauen
3 nicht klimabewusste Männer	4 nicht klimabewusste Frauen

Trifft die Hypothese zu, so müsste sich dies auch in der Vier-Felder-Tafel niederschlagen, und zwar so, dass im Idealfall nur Felder auf der Diagonalen besetzt sind, hier also Feld 2 „klimabewusste Frauen“ und Feld 3 „nicht klimabewusste Männer“. Ist das nicht der Fall, d.h. gibt es prozentual vielleicht sogar mehr klimabewusste Männer als Frauen, so ist die Hypothese augenscheinlich gescheitert. Es gibt nun allerdings auch noch eine dritte Möglichkeit, nämlich dass die empirischen Häufigkeiten und der sich dort zeigende Zusammenhang auf die Zufälligkeit der Auswahl der Probanden zurückzuführen sind.

Beim statistischen Hypothesentest prüft man einen empirischen Sachverhalt gegen die Zufälligkeit einer solchen Verteilung. Der statistische Hypothesentest ist immer eine Entscheidung zwischen zwei Möglichkeiten: Man prüft nicht einfach, ob eine Hypothese richtig oder falsch ist, zutrifft oder nicht, sondern man stellt zwei Hypothesen einander gegenüber: die sogenannte *Alternativhypothese* und die *Nullhypothese.*

Der Begriff *Alternativhypothese*, üblicherweise mit H_1 abgekürzt, hat nichts mit der „alternativen Szene“ oder anderen Alltagsbedeutungen von „alternativ“ zu tun, sondern das Wort „alternativ“ hat hier die Bedeutung von „neu“. Die als Alternativhypothese formulierte Aussage ist es, die im Mittelpunkt des Interesses der Forschenden steht. Hiermit möchte man ein bestimmtes Phänomen erklären und Zusammenhänge offen legen. Das tut man gewöhnlich nur im Falle eines bisher nicht oder nur unzureichend erklärten Sachverhalts. Insofern ist der Begriff „alternativ“ ganz treffend, denn es ist eine neue Erklärung, alternativ und/oder ergänzend zum bisherigen Forschungsstand.

Die *Nullhypothese*, abgekürzt mit H_0, ist nun eine formale Gegenhypothese zur formulierten Alternativhypothese. Sie ist eine Negativhypothese, mit der behauptet wird, dass die zur Alternativhypothese komplementäre Aussage richtig ist. Die Nullhypothese besagt schlichtweg, dass der postulierte Zusammenhang null und nichtig ist. Null- und Alternativhypothese sind miteinander konfrontiert wie die Abgeordneten von Regierung und Opposition im englischen Parlament. Es gibt nur zwei Alternativen, entweder H_1 oder H_0 trifft zu, eine dritte Möglichkeit existiert nicht, der Kampf kann also weder unentschieden ausgehen, noch kann die Alternativhypothese „ein bisschen zutreffen“.

Das Entscheidungskalkül ist dabei konservativ, d.h. die Nullhypothese hat eigentlich die besseren Chancen, dass die Entscheidung für sie ausfällt – solange man die empirischen Gegebenheiten mit hinreichender Wahrscheinlichkeit aus dem Zufall erklären kann, solange behält man auch die H_0 bei.

Wichtig zum Verständnis der Logik des statistischen Schließens ist, dass sich die beiden gegenüberstehenden Hypothesen auf die Grundgesamtheit beziehen. Dies ist der Zusammenhang, der eigentlich interessiert: ein Zusammenhang zwischen zwei Variablen in der Grundgesamtheit. Um die Parameter der Grundgesamtheit von den Kennwerten der Stichprobe zu unterscheiden, benutzt man bei Formeln und Notationen griechische Buchstaben für die Parameter der Grundgesamtheit.

Aus einer Alternativhypothese ergibt sich qua logischer Gegenüberstellung die statistische Nullhypothese.

Alternativhypothese (H_1) und Nullhypothese (H_0) sind zueinander komplementäre Aussagen. Die Nullhypothese ist eine formale Gegenhypothese, die behauptet, dass der in H_1 formulierte Zusammenhang nicht existiert.

Wenn bspw. eine neue Methode des Schriftspracherwerbs gegenüber der herkömmlichen Methode getestet werden soll, kann man die Alternativhypothese „Die neue Methode ist besser als die alte“ aufstellen. Die Nullhypothese würde folglich lauten „Die alte Methode ist besser oder gleich gut wie die neue“. Formalisiert schreibt man folgendes, wobei μ für den Mittelwert der Grundgesamtheit steht:

$H_1: \mu_{neu} > \mu_{alt}$ → $H_0: \mu_{neu} \leq \mu_{alt}$

Andere denkbare Formulierungen für die Alternativhypothese und die komplementäre Nullhypothese lauten:

$H_1: \mu_{neu} < \mu_{alt}$ → $H_0: \mu_{neu} \geq \mu_{alt}$
$H_1: \mu_{neu} \neq \mu_{alt}$ → $H_0: \mu_{neu} = \mu_{alt}$

Arten von Hypothesen

Es lassen sich verschiedene Arten von Hypothesen unterscheiden:

- *Unterschieds- und Zusammenhangshypothesen*: Unterschiedshypothesen untersuchen, ob systematische Unterschiede zwischen zwei und mehr Gruppen bestehen. Beispiele: „Frauen sind umweltbewusster als Männer“; „Das Klimabewusstsein unterscheidet sich nach Bildungsstand“; „Kinder mit Migrationshintergrund besuchen seltener weiterführende Schulen“; „Durch ein frühpädagogisches Förderprogramm lässt sich der Spracherwerb entscheidend verbessern“. Zusammenhangshypothesen untersuchen Zusammenhänge zwischen mindestens zwei Variablen. Beispiele: „Je öfter Kinder mit dem Gameboy spielen, desto schlechter können sie sich konzentrieren“; „Je positiver eine Person zum Umweltschutz eingestellt ist, desto positiver ist auch ihr Umweltverhalten“; „Je länger die Schul- und Hochschulausbildung dauert, desto höher das Einkommen“.

- *Gerichtete und ungerichtete Hypothesen*: Sowohl Unterschieds- als auch Zusammenhangshypothesen können gerichtet oder ungerichtet formuliert werden. Die ungerichtete Hypothese behauptet, dass zwischen zwei Merkmalen ein Zusammenhang bzw. Unterschied besteht, ohne dass eine Angabe über die Richtung des Zusammenhangs bzw. Unterschieds gemacht wird. Beispiel: „Das Umweltbewusstsein differiert nach Geschlecht"; „Das Klimabewusstsein unterscheidet sich nach Bildungsstand"; „Umweltbewusstsein und Umweltverhalten hängen zusammen". Die gerichtete Hypothese beinhaltet bereits eine Angabe über die Richtung des vermuteten Zusammenhangs, z.B.: „Frauen sind umweltbewusster als Männer"; „Kinder mit Migrationshintergrund besuchen seltener weiterführende Schulen".
- *Spezifische und unspezifische Hypothesen*: Die oben formulierten Hypothesen sind allesamt unspezifische Hypothesen, denn es wird nicht angegeben, wie groß der Unterschied bzw. wie groß der Zusammenhang ist. Wenn man in der Lage ist, die Größe anzugeben (Beispiel: „Kinder mit Migrationshintergrund besuchen halb so häufig weiterführende Schulen wie Kinder ohne Migrationshintergrund"), bezeichnet man die Hypothese als spezifische Hypothese.

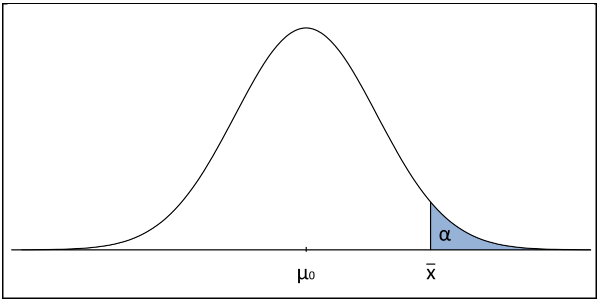
6.4 Der Hypothesentest

Wir haben oben gezeigt, dass der Stichprobenkennwert Mittelwert ($\bar{x}$) normalverteilt ist, d.h. wir können bei bekanntem Mittelwert der Grundgesamtheit für jeden beliebigen Stichprobenmittelwert berechnen, wie wahrscheinlich das Auftreten dieses Wertes oder größerer Werte ist.

Die Nullhypothese ist eine Annahme über die Wahrscheinlichkeitsverteilung einer Zufallsvariablen.

Betrachten wir hierzu die folgende Mittelwertverteilung (Abb. 6-1). In der Grafik bezeichnet μ_0 den Mittelwert der Grundgesamtheit und $\bar{x}$ den in der Stichprobe ermittelten Mittelwert. In der Grafik liegt dieser Mittelwert der Stichprobe relativ weit von μ_0 entfernt. Mit dem schraffierten Bereich α wird der Bereich unterhalb der Normalverteilungskurve bezeichnet, welcher der Wahrscheinlichkeit von $\bar{x}$ und allen größeren Werten entspricht. Es dürfte deutlich werden, dass dieser Bereich schrumpft, wenn $\bar{x}$ größer wird, d.h. die Wahrscheinlichkeit, dass der Mittelwert aus dieser Verteilung mit den Parametern μ_0 und σ_0 stammt, sinkt mit steigender Entfernung zu μ_0.

Abb. 6-1



Da die Normalverteilungskurve sich asymptotisch der x-Achse nähert, diese aber nie schneidet, gibt es keinen Punkt, keinen Schwellenwert, jenseits dessen man mit Sicherheit sagen könnte, dass der Stichprobenmittelwert nicht aus dieser Verteilung stammen kann. Man kann also bei diesem Modell des Hypothesentestens nie ausschließen, dass die Nullhypothese doch gilt, sondern lediglich eine äußerst geringe Wahrscheinlichkeit diagnostizieren. Das Modell des statistischen Schließens schafft also keine Sicherheit, sondern verspricht nur Wahrscheinlichkeiten. Das kann in Situationen und Konflikten der Lebenswelt durchaus entscheidende Bedeutung haben. Es macht eben für die gesellschaftliche Kommunikation einen Unterschied, ob Wissenschaftler/innen einen GAU in einem Atomkraftwerk ausschließen können oder ob sie feststellen, dass die Wahrscheinlichkeit kleiner als 0,0001 ist. Wenn sie dann noch einräumen müssen, dass die Aussage des Modells „einmal in 10.000 Jahren" auch bedeuten kann, dass dieses besagte Ereignis zufällig bereits im nächsten Jahr eintreten könnte, trägt dies vermutlich nicht zur Beruhigung der Bevölkerung bei.

Der Mittelwert einer Stichprobe lässt sich immer aus der mit der Nullhypothese verbundenen Verteilungsannahme erklären. Wie groß die Wahrscheinlichkeit ist, den aufgetretenen Wert unter der Annahme der Gültigkeit der Nullhypothese zu erklären, können wir relativ leicht ermitteln, in dem wir den zu $\bar{x}$ gehörenden z-Wert berechnen und in der Tabelle der Normalverteilung nachschlagen, wie groß die Wahrscheinlichkeit für das Auftreten von diesem und allen größeren Werten ist.

Würde man sich in der obigen Abbildung für die Alternativhypothese entscheiden (d.h. $\bar{x}$ entstammt nicht der abgebildeten Verteilung), so wäre diese Entscheidung mit der Irrtumswahrscheinlichkeit α belastet. Angenommen, diese betrüge 3%, ließe sich auch formulieren: „Die Irrtumswahrscheinlichkeit bei meiner Entscheidung für die Alternativhypothese beträgt 3%" oder „Lehne ich die Nullhypothese ab, so begehe ich mit 3% Wahrscheinlichkeit einen Fehler".

Wie viel Irrtumswahrscheinlichkeit ist man bereit zu akzeptieren? Die empirische Forschung hat hier gewisse Konventionen entwickelt und bezeichnet diese als *Signifikanzniveaus*. Es ist also nicht so, dass jeweils neu zu entscheiden wäre, wie viel Prozent Irrtumswahrscheinlichkeit man genau akzeptieren will, sondern die Wahl reduziert sich auf die Entscheidung zwischen verschiedenen vorgegebenen Signifikanzniveaus. Die empirische Forschung unterscheidet in der Regel das 0,1%-, das 1%- und das 5%-Signifikanzniveau, in manchen (seltenen) Fällen arbeitet man – bei kleinen Stichproben – auch mit dem 10% Niveau.

Beträgt die Wahrscheinlichkeit des gefundenen oder eines extremeren Untersuchungsergebnisses unter der Annahme, die H_0 sei richtig, unter 5%, so wird dieses Ergebnis als *signifikant* bezeichnet. Beträgt diese Wahrscheinlichkeit unter 1%, so ist das Ergebnis *sehr signifikant* (synonym: *hoch signifikant*).

Die Wahl des Signifikanzniveaus ist abhängig von der Stichprobengröße einerseits und von den Konsequenzen eines Fehlers andererseits. Bei großen Stichproben ($n > 1.000$) sollte das 1%-Niveau gewählt werden. Gleiches gilt, wenn eine Fehlentscheidung zugunsten der H_1 gravierende Folgen nach sich ziehen würde, bspw. dass eine neue Lernmethode mit neuen Schulbüchern im Schulunterricht eingeführt würde oder Präventionsprogramme implementiert würden, die sehr große Kosten nach sich ziehen.

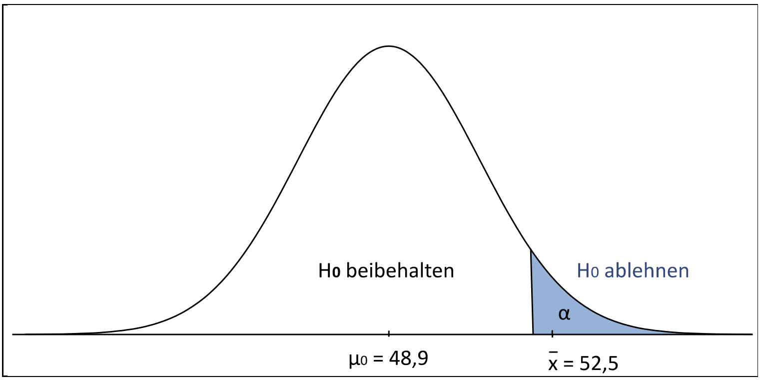
Alle mit dem statistischen Testen verbundenen Fragen sind – wie ein Blick auf die Formel zur Ermittlung des Standardfehlers zeigt – aufs engste mit der Stichprobengröße verknüpft. Je größer n, desto geringer der Standardfehler. Dieser nimmt proportional zur Quadratwurzel des Stichprobenumfanges ab. Dies bedeutet praktisch, dass man den Stichprobenumfang vervierfachen müsste, um den Standardfehler zu halbieren.

6.5 Einseitige und zweiseitige Tests

Hypothesen können wie oben dargelegt gerichtet oder ungerichtet sein. Für gerichtete Hypothesen werden einseitige, für ungerichtete zweiseitige Tests durchgeführt. Für den einseitigen Test ergibt sich folgende Situation: Überschreitet der Stichprobenmittelwert einen bestimmten Punkt, die Signifikanzschwelle oder Signifikanzhürde, so führt dies zur Ablehnung der Nullhypothese. Alle Stichprobenmittelwerte, die größer sind, liegen im *Ablehnungsbereich*, der in Abb. 6-2 blau dargestellt ist. Alle Werte im weißen Bereich links davon führen zur Beibehaltung der H_0 (*Annahmebereich*). Bei Wahl des 1%-Niveaus anstelle des 5%-Niveaus rückt der Schwellenwert auf der x-Achse weiter nach rechts und α wird entsprechend kleiner. Es kann selbstverständlich vorkommen, dass Stichprobenmittelwerte zwischen den Schwellenwerten 1%- und 5%-Niveau liegen, d.h. das eine Mal (1%-Niveau) würde man sich für die H_0, das andere Mal (5%-Niveau) für die H_1 entscheiden.

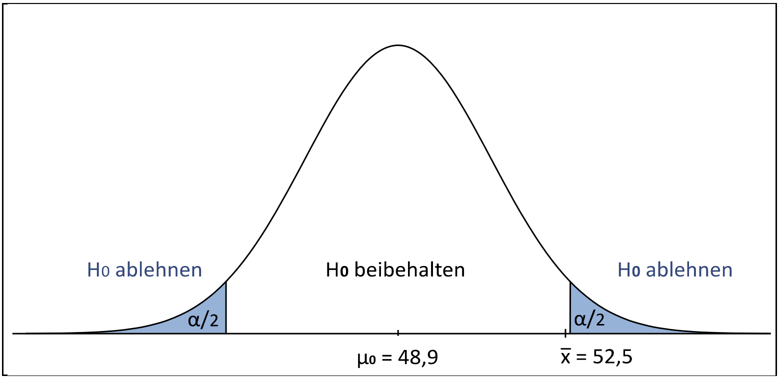
Betrachten wir das zu Beginn dieses Kapitels dargestellte Beispiel einer Population von n = 2.034 Personen, der Mittelwert des Alters μ_0 beträgt 48,9 Jahre. Nun hätten wir bei der Stichprobe von n = 100 einen Altersmittelwert von 52,5 ermittelt und unsere Hypothese wäre, dass es sehr unwahrscheinlich ist, dass diese im Mittel deutlich ältere Stichprobe aus der Grundgesamtheit mit $\mu_0 = 48{,}9$ stammt. Für die gerichtet formulierte Hypothese ergibt sich die in Abb. 6-2 dargestellte Situation. $\bar{x}$ liegt im Ablehnungsbereich, d.h. die Nullhypothese wird abgelehnt und die Entscheidung erfolgt pro Alternativhypothese.

Abb. 6-2



Bei der ungerichteten Hypothese und dem zweiseitigen Hypothesentest ergibt sich eine etwas veränderte Situation, die in Abb. 6-3 dargestellt ist. Jetzt gibt es auf beiden Seiten der Verteilung einen Ablehnungsbereich, da ja in der Hypothese nichts über die Richtung des Zusammenhangs ausgesagt wird. Die Irrtumswahrscheinlichkeit α wird nun zu gleichen Teilen auf die beiden Seiten aufgeteilt. Die Konsequenz ist folglich, dass die Signifikanzschwelle bei der gerichteten Hypothese niedriger ist, denn dann ist – bildlich gesprochen – alle Irrtumswahrscheinlichkeit an einer Seite angelagert, so dass die Chance, dass der Stichprobenmittelwert nicht mehr im Annahmebereich der Nullhypothese liegt, größer ist. Und tatsächlich liegt in unserem Beispiel der Stichprobenmittelwert von 52,5 Jahren jetzt außerhalb des verkleinerten Signifikanzbereichs und wir würden in diesem Fall die H_0 beibehalten.

Abb. 6-3



Beim Testen von Hypothesen sind zwei Punkte unbedingt zu beachten:

Erstens müssen die Hypothesen *vor* dem Test formuliert werden. Also es muss vorher klar sein, ob die Hypothese gerichtet oder ungerichtet ist und ob dementsprechend ein einseitiger oder zweiseitiger Test durchzuführen ist. Eine spätere Modifikation der Hypothese mit dem Ziel, die H_1 annehmen zu können, ist nicht statthaft.

Zweitens sollte die Zahl der Hypothesentests möglichst gering gehalten werden, denn bei einer hinreichend großen Zahl wächst aufgrund der gegebenen Irrtumswahrscheinlichkeit die Gefahr, sich irrtümlich für die Alternativhypothese zu entscheiden.

6.6 Alpha-Fehler und Beta-Fehler

Das inferenzstatistische Hypothesentesten bezieht sich immer auf die Grundgesamtheit, deren Parameter meist unbekannt sind. Ob die in den Hypothesen formulierten Zusammenhänge „stimmen“, wird immer mit Blick auf die Grundgesamtheit untersucht und nicht lediglich beschränkt auf die Stichprobe. Beim Entscheidungsprozess Nullhypothese versus Alternativhypothese existieren nun prinzipiell vier Varianten für eine richtige bzw. falsche Entscheidung:

		In der Grundgesamtheit gilt die	
		H_0	**H_1**
Entscheidung aufgrund der Stichprobe (empirische Studie)	**H_0**	richtige Entscheidung	*β-Fehler (Fehler 2. Art)*
	H_1	*α-Fehler (Fehler 1. Art)*	richtige Entscheidung

Unproblematisch ist das Feld H_0/H_0, d.h. in der Grundgesamtheit gilt die H_0 und aufgrund der Ergebnisse der Stichprobe entscheidet man sich richtigerweise ebenfalls für die H_0. Ein Beispiel: Die Ergebnisse einer Studie an der Uni Marburg zeigen keine Unterschiede zwischen Studierenden der Medizin und Ingenieurstudierenden hinsichtlich der wöchentlich für das Studium benötigten Zeit. Gleiches gilt in der Grundgesamtheit. Ähnlich verhält es sich im Fall H_1/H_1, wenn man sich aufgrund der Ergebnisse der Stichprobe für die H_1 entscheidet. Die Stichprobe zeigt: Studierende der Medizin verbringen wöchentlich mehr Zeit mit dem Studium als Studierende der Soziologie. Dies gilt auch in der Grundgesamtheit, die Entscheidung ist richtig.

Die Kombinationen H_0/H_1 bzw. H_1/H_0 stellen hingegen fehlerhafte Entscheidungen dar. Im einen Fall (H_1/H_0) entscheidet man sich aufgrund der Ergebnisse der Stichprobe für die H_1 (ja, es gibt einen Unterschied zwischen den beiden Gruppen), aber in der Grundgesamtheit gilt die Nullhypothese (also kein Zusammenhang). Diesen Fehler bezeichnet man als *Fehler erster Art* (engl. type I error) oder *Alpha-Fehler*.

Bei der Kombination H_0/H_1 gilt genau das Umgekehrte: Man entscheidet sich aufgrund der Stichprobenergebnisse für die Beibehaltung der H_0, aber in der Grundgesamtheit gilt die H_1 (entgegen den Ergebnissen der Studie unterscheiden sich Medizinstudierende und Ingenieursstudierende doch). Diesen Fehler bezeichnet man als *Fehler zweiter Art* (engl. type II error) oder *Beta-Fehler*.

Wie man die Alpha-Irrtumswahrscheinlichkeit, die Wahrscheinlichkeit für den Fehler erster Art, bestimmen kann, wurde in diesem Kapitel ausführlich beschrieben. Ein vergleichbares Verfahren für die Bestimmung des Beta-Fehlers existiert aber nicht. Warum dies so ist, lässt sich leicht nachvollziehen: Der Beta-Fehler soll ja angeben, wie wahrscheinlich es ist, dass der gefundene Mittelwert der Stichprobe aus einer (alternativen) Verteilung mit dem Mittelwert μ_1 und dem Standardfehler σ_1 stammt. Nun weiß man aber gar nichts über eine solche Verteilung und kann folglich auch keinen Verteilungsmittelwert und keinen Standardfehler schätzen. Theoretisch könnte es sich um unendlich viele alternative Verteilungen handeln. Der Beta-Fehler lässt sich nur dann bestimmen, wenn man eine spezifische Hypothese formuliert, also eine Information über μ_1 und σ_1 besitzt. Nur für diesen Fall ist man in der Lage, die Wahrscheinlichkeit zu bestimmen, dass der Mittelwert der Stichprobe aus einer Kennwerteverteilung mit eben diesen beiden Parametern stammt. Spezifische Hypothesen werden in der erziehungs- und sozialwissenschaftlichen Forschung aber nur selten formuliert und insofern findet man in der Forschungsliteratur auch nur selten Angaben zu Beta-Fehlern, während die Alpha-Irrtumswahrscheinlichkeit einen auf Schritt und Tritt begleitet.

6.7 Signifikanz – ein Begriff, der in die Irre führen kann

Signifikant, das wissen diejenigen die in der Schule Latein lernen konnten (mussten), bedeutet *entscheidend, wichtig, bedeutsam*. So könnte man denn auch annehmen, dass ein signifikantes Ergebnis eben ein *bedeutsames* ist. Manche schließen aus dem Begriff „signifikant“ sogar, dass damit ein Ergebnis der empirischen Studie *bewiesen* sei.

Das ist aber gerade nicht die Bedeutung, wenn ein Forschungsresultat als statistisch signifikant bezeichnet wird. Beim statistischen Hypothesentest bedeutet die Formulierung, dass ein Ergebnis signifikant ist, lediglich, dass die Wahrscheinlichkeit, dass dieses zufällig entstanden ist, also aus der Nullhypothese erklärt werden kann, gering ist.

Diese reduzierte Bedeutung des Untersuchungsbefunds „signifikant“ mag auf den ersten Blick enttäuschend klingen, denn es ist eben nicht so, dass man durch einen statistischen Test die „Wahrheit“ finden könnte oder ihr doch zumindest immer näher käme. Nein, man findet keine Wahrheiten, sondern man muss immer mit einem Alpha-Fehler rechnen.

Die Art und Weise, wie die Entscheidung für die Alternativhypothese gefällt wird, ist alles andere als einfach und direkt – für Außenstehende ist sie nicht leicht verständlich. Die Entscheidung für die Alternativhypothese erfolgt immer nur indirekt, nämlich als Umkehrschluss. Umgekehrt: Auch ein nicht-signifikantes Ergebnis ist keineswegs ein Beleg für die Richtigkeit der Nullhypothese. Sie ist damit keineswegs bewiesen und aus dem Tatbestand, dass die Alpha-Irrtumswahrscheinlichkeit bspw. 6% beträgt, lässt sich keineswegs der Schluss ziehen, die Nullhypothese gelte mit 94% Sicherheit. Auch aus einer hohen Irrtumswahrscheinlichkeit und der daraus folgenden klaren Entscheidung pro H_0 lässt sich keine Aussage über die Wahrscheinlichkeit der Gültigkeit der H_0 ableiten. Es lässt sich lediglich behaupten, dass das empirische Ergebnis gut mit der H_0 vereinbar ist, aber es ist ein Fehlschluss anzunehmen, dass die H_0 umso richtiger sei, je höher die Alpha-Irrtumswahrscheinlichkeit ist. Gleiches gilt für die H_1: Hat man sich bspw. aufgrund von $\alpha < 1\%$ für die H_1 entschieden, so kann man deshalb noch lange nicht behaupten, diese gelte mit einer Wahrscheinlichkeit größer 99%.

Generell gilt, dass sich die Irrtumswahrscheinlichkeit in folgenden Fällen vermindert:

- mit größer werdender Differenz von Stichprobenmittelwert und Mittelwert der Grundgesamtheit,
- mit kleiner werdender Populationsstreuung und
- mit vergrößertem Stichprobenumfang.

Vor allem in der Nähe der Signifikanzschwelle befindet man sich also in einem recht unsicheren Terrain, weiß man doch, dass eine Erhöhung der Stichprobengröße in einer zukünftigen Untersuchung höchstwahrscheinlich zu einem signifikanten Ergebnis führen würde. Zudem ist es so, dass bei genügend großer Stichprobe hypothesenkonforme Unterschiede immer signifikant werden. Auch dies stellt in der öffentlichen Kommunikation über Ergebnisse wissenschaftlicher Forschung ein Problem dar, weil hierdurch auch substantiell ziemlich irrelevante Ergebnisse als signifikant (gleich bedeutsam) dargestellt werden.

Die inhaltliche Bedeutsamkeit ist eine Frage der absoluten Differenz von Stichprobenmittelwert und Mittelwert der Grundgesamtheit und nicht eine Frage, die durch den Signifikanztest beantwortet wird.

6.8 Effektgröße

Die Effektgröße (engl. effect size), häufig auch Effektstärke genannt, ist ein Maß für die praktische Bedeutsamkeit eines gefundenen Zusammenhangs. Die Signifikanz ist, wie dargestellt, abhängig von der Stichprobengröße und insofern kein geeignetes Maß, um einschätzen zu können, ob ein Ergebnis wirklich praktisch bedeutsam ist. Ob Mittelwertunterschiede zwischen Experimentalgruppe und Kontrollgruppe – oder allgemein formuliert zwischen zwei Gruppen – von praktischer Bedeutung sind, ist natürlich zuallererst eine Frage der theoretischen Bewertung. Wenn ein pädagogisches Frühförderprogramm bei dreijährigen Kindern bspw. dazu führt, dass nach einer sechsmonatigen Intervention bei einem kognitiven Test im Mittel 102 anstelle von 98 Punkten erreicht werden, so bedarf dies vorrangig einer pädagogischen Bewertung, bei der sicherlich auch die Kosten eines solchen Programms nicht unberücksichtigt bleiben können. Die Effektgröße als Maßzahl liefert für eine solche Bewertung eine wertvolle Grundlage, denn sie setzt die Unterschiede zwischen den beiden Gruppenmittelwerten ins Verhältnis zur Standardabweichung der Population, d.h. es findet eine Art Normierung statt, welche die Vergleichbarkeit von Resultaten ermöglicht. Vergleichen lassen sich mittels der Maßzahl „Effektgröße" nicht nur verschiedene empirische Studien, sondern auch Ergebnisse, die mit verschiedenen Messinstrumenten erhoben wurden. Insofern ist die Effektgröße insbesondere für Metaanalysen, in der Resultate unterschiedlicher empirischer Studien miteinander verglichen werden, von besonderer Bedeutung.

Die Effektgröße für die standardisierte Mittelwertdifferenz (in Kapitel 7 wird hierüber mehr ausgeführt) wird nach folgender Formel bestimmt:

$$\text{Effektgröße für Mittelwertdifferenzen} = \frac{\mu_1 - \mu_2}{\sigma}$$

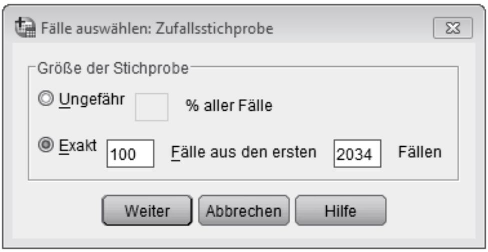
Die griechischen Buchstaben lassen erkennen, dass sich alle Größen der Gleichung auf Parameter der Grundgesamtheit beziehen. In den meisten Fällen liegen diese nicht vor, sondern müssen aus den empirisch ermittelten Kennwerten der beiden Gruppen geschätzt werden. Cohen (1988) unterscheidet zwischen drei Effektstufen bei Mittelwertdifferenzen: kleine (0,2), mittlere (0,5) und große Effekte (0,8).

Effektstärkenmaße existieren in zahlreichen Formen und für unterschiedliche Anwendungen, also nicht nur für Mittelwertdifferenzen. In Kombination mit Überlegungen zu Alpha- und Beta-Irrtumswahrscheinlichkeiten, die man zu akzeptieren bereit ist, gestatten solche Maße nicht nur eine a-posteriori-Bestimmung der Effektgröße, sondern mit ihrer Hilfe lassen sich auch für geplante Untersuchungen optimale Stichprobengrößen bestimmen, um so eine angestrebte Effektgröße auch nachweisen zu können.

6.9 So geht es mit SPSS/MYSTAT

Eine Zufallsstichprobe in SPSS ziehen

Um aus einem Datensatz eine Zufallsstichprobe zu ziehen, muss die Funktion „Daten > Fälle auswählen > Zufallsstichprobe“ gewählt werden. Im folgenden Dialogfeld kann man festlegen, wie groß die zu ziehende Stichprobe sein soll.



Zur Durchführung der oben beschriebenen Zufallsexperimente sind 100 bzw. 300 Fälle zufällig auszuwählen. In der Datenansicht von SPSS lässt sich erkennen, welche Fälle ausgewählt wurden und welche nicht – die Datensatznummer nicht ausgewählter Fälle in der am weitesten links stehenden Spalte ist durchgestrichen. Die getroffene Auswahl bleibt so lange erhalten, bis sie explizit durch „Daten > Fälle auswählen > Auswählen > Alle Fälle“ aufgehoben wird.

Standardfehler in SPSS berechnen

Der Standardfehler lässt sich mithilfe von „Analysieren > Deskriptive Statistiken > Häufigkeiten“ berechnen. Dort ist auf der Registerkarte „Statistiken“ bei der Option „Standardfehler“ ein Häkchen zu setzen.

Zum gleichen Resultat führt die SPSS-Prozedur „Analysieren > Deskriptive Statistiken > Deskriptive Statistik“, wo aus den Optionen „Standardfehler“ angeklickt werden muss.

Konfidenzintervalle in SPSS

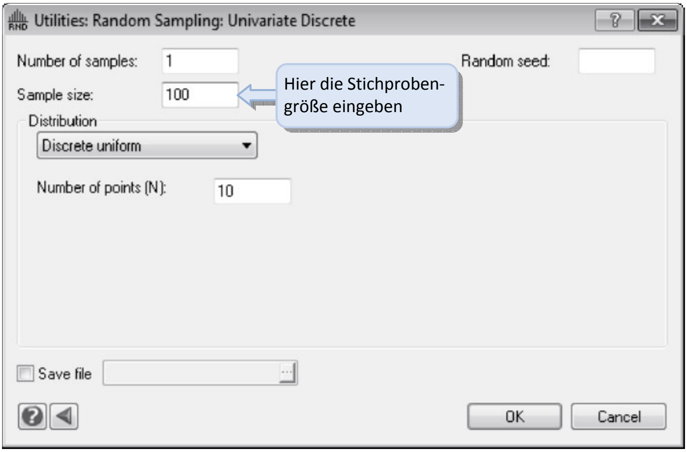
Interessierende Konfidenzintervalle für Populationsmittelwerte können entweder nach den oben beschriebenen Formeln selbst berechnet werden, nachdem man mit SPSS den Mittelwert und den Standardfehler berechnet hat. Alternativ und zeitsparender ist der Weg über „Analysieren > Mittelwerte vergleichen > T-Test bei einer Stichprobe“. Dort kann im Fenster Optionen das gewünschte Konfidenzintervall, z.B. 95% oder 99%, eingestellt werden. Zudem gibt SPSS bei vielen anderen Berechnungen automatisch das Konfidenzintervall für den jeweiligen Parameter mit aus.

Einseitige und zweiseitige Tests in SPSS

Bei vielen der in den folgenden Kapiteln näher beschriebenen statistischen Verfahren sieht das SPSS-Programm keine explizite Auswahl für einseitige oder zweiseitige Tests vor. In der Regel wird die ausgegebene Irrtumswahrscheinlichkeit für die zweiseitige Fragestellung berechnet, so dass diese für den einseitigen Test halbiert werden muss.

Eine Zufallsstichprobe in MYSTAT ziehen

Zufallsstichproben lassen sich in MYSTAT über „Utilities > Random Sampling > Univariate Discrete“ ziehen. Im Dialogfeld lässt sich bei „Number of Samples“ die Anzahl der gewünschten Stichproben und bei „Sample size“ die gewünschte Stichprobengröße angeben. Die resultierende Zufallsstichprobe kann direkt als neue Datei gespeichert werden.



Standardfehler und Konfidenzintervall in MYSTAT

Im Bereich „Analyze > Basic Statistics" findet man unter Optionen „SE of AM" (dies ist der Standardfehler des Mittelwerts) und „CI of AM", das Konfidenzintervall für den Populationsmittelwert. Dort kann die Art des zu berechnenden Konfidenzintervalls angegeben werden. Für das 95%-Konfidenzintervall ist .95 anzugeben und für das 99%-Intervall .99.

Einseitige und zweiseitige Tests in MYSTAT

MYSTAT sieht bei allen Hypothesentests („Analyze > Hypothesis Testing") die explizite Auswahl für einseitige oder zweiseitige Tests vor. Im Dialogfeld für den jeweiligen Hypothesentest lässt sich im Feld „Alternative type" die Auswahl „not equal" (zweiseitig) sowie „greater than" bzw. „less than" (einseitig) treffen. MYSTAT gibt dann die entsprechende Irrtumswahrscheinlichkeit aus.

Lernfragen zu diesem Kapitel

1. Welche Bedeutung hat der Standardfehler?
2. Ist das 99%-Konfidenzintervall genauer als das 95%-Konfidenzintervall? Empfiehlt es sich, möglichst immer das 99%-Konfidenzintervall zu benutzen?
3. Warum kann ein Alpha-Fehler nicht gleichzeitig mit einem Beta-Fehler auftreten?
4. Aufgrund einer Irrtumswahrscheinlichkeit von 5,8% entscheidet man sich nicht für die Alternativhypothese. Bedeutet dies, dass die Nullhypothese mit 94,2%iger Sicherheit gilt?
5. Können sich Situationen ergeben, in denen keine Entscheidung zwischen Nullhypothese und Alternativhypothese möglich ist?
6. Wie groß darf die Alpha-Irrtumswahrscheinlichkeit höchstens sein, wenn man von einem signifikanten Ergebnis spricht?