Technische Universität Wien

Vienna University of Technology

Diplomarbeit

Grundlagen der Statistik und deren Anwendung im Bereich der Psychologie

Ausgeführt am Institut für
Statistik und Wahrscheinlichkeitstheorie
der Technischen Universität Wien

unter der Anleitung von
Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Heinz Stadler

durch

Wolfgang Schneßl
Schulgasse 9
2640 Gloggnitz

Datum

Unterschrift

Vorwort

Der deutsche Statistiker Walter Krämer eröffnet sein Buch *Statistik verstehen* mit folgendem Zitat des englischen Schriftstellers H. G. Wells: „A basic literacy in statistics will one day be as necessary for efficient citizenship as the ability to read and write.“ Daß eben dieser gewisse Bildungsgrad wichtig und notwendig ist, zeigt er anhand vieler praktischer Beispiele auf. Aufgrund meines Studiums der Unterrichtsfächer Mathematik und Psychologie / Philosophie wollte ich für meine Diplomarbeit ein fächerübergreifendes Themengebiet finden. Die Wahl fiel auf Statistik, da diese in der Anwendung für Psychologen fast schon ihr täglich Brot darstellt.

Für die Zwecke einer (knappen) Einführung in die Methoden der Statistik und deren Anwendung im Bereich der Psychologie kam nur das Prinzip der exemplarischen Darstellung von Verfahren in Betracht. Jedes Bemühen um einen systematischen Überblick oder um Vollständigkeit auch nur in Teilbereichen muß unterbleiben.

Teils um zu zeigen, wie Statistik in der Praxis wirken kann, und teils aus persönlichem Interesse habe ich eine kleine Untersuchung aus dem Themenbereich „Alkohol, Nikotin, Sport und Ernährung“, also im weitesten Sinne „Sport und Gesundheit“, durchgeführt.

Danken möchte ich meinem Betreuer Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Heinz Stadler für die angenehme Zusammenarbeit und zu guter Letzt, ja eigentlich in erster Linie meiner Familie, allen voran meinen Eltern. Ohne ihre Unterstützung über die gesamte Zeitdauer meines Studiums hinweg wäre der Weg um einiges schwieriger gewesen.

Wolfgang Schneßl

April 2007

PS: Noch eine kurze Anmerkung zum leidigen (und nicht befriedigend lösbaren) Problem der Verwendung geschlechtsspezifischer Sprachformen. Ich habe mich dafür entschieden, aus Gründen der besseren Lesbarkeit immer nur eine Form (weiblich, männlich) zu verwenden. Selbstverständlich ist auch immer die andere Form mitgedacht.

Inhaltsverzeichnis

Tabellenverzeichnis

1 Einleitung und Überblick

Zu Beginn meiner Arbeit wird die Rolle der Statistik in der Psychologie näher betrachtet. Was leistet Statistik? Was sind die Ziele? Warum ist Statistik untrennbar mit der empirischen Wissenschaft der Psychologie verbunden?

In den einzelnen Kapiteln werden die Begriffe des *Messens* und der *Skalenniveaus* von Variablen geklärt. Wichtige Lokalisations- sowie Dispersionsparameter werden besprochen. Zufallsvariablen und Verteilungen werden behandelt.

Die Grundlagen der analytischen Statistik, verschiedene relevante Tests und schließlich die Kapitel der Korrelation und der Regression werden näher ausgeführt.

Am Ende der Arbeit steht die Untersuchung „Leben sportliche Kinder gesünder“, in der versucht wird, die positiven Aspekte des Sports dahingehend darzustellen, daß sie dem Einfluß der in Österreich legalen Drogen Alkohol und Nikotin entgegenwirken. Der für diese Untersuchung verwendete Fragebogen ist im Anhang zu finden.

2 Aufgaben und Ziele der Psychologie

Gegenstand der modernen Psychologie, wie sie seit etwa 100 Jahren besteht, sind Verhalten, Erleben und Bewußtsein des Menschen, deren Entwicklung während seines gesamten Lebens und deren innere und äußere (von der Umwelt kommende) Bedingungen und Ursachen.

Dieser Gegenstand scheint auf den ersten Blick recht weit gefaßt zu sein. Da wir aber hier von der Wissenschaft der Psychologie im allgemeinen sprechen, muß das so sein. Der Forschungsgegenstand einzelner Untergruppen (wie Sozialpsychologie, klinischer Psychologie etc ...) wird klarer eingegrenzt.

Nun lassen sich dennoch recht gut wenige grundlegende *Aufgaben und Ziele* dieser Wissenschaft anführen:

- Beschreiben
- Erklären
- Vorhersagen

Teilweise wird in diese Liste der Ziele der Psychologie noch die *Kontrolle des Verhaltens* aufgenommen. Für unsere Zwecke wird dies nicht notwendig sein. Bei eingehender Betrachtung dieser drei Hauptziele läßt sich bereits vermuten, wie und in weiterer Folge wieso ein sinnvoller Zusammenhang zwischen der Psychologie und der Statistik besteht, man könnte auch behaupten: bestehen muß. Genaueres soll im folgenden dargestellt werden.

Betrachtet man nun den Entwicklungsprozeß der Psychologie, so läßt sich feststellen, daß die Verwendung der Mathematik wesentliche Vorteile erbracht hat, was die Objektivität und Präzision von Aussagen betrifft. Hat man in der Vergangenheit (vor 1900) Verhalten noch eher im mündlichen Stil beschrieben, so geht der Trend ab dem 20. Jahrhundert dahin, daß auch in der Psychologie *Experimente* ihren Einzug gefunden haben. Werden Ergebnisse solcher wissenschaftlicher Vorgehensweisen unter Verwendung von Zahlen objektiv festgehalten, kann in Summe viel besser und allgemeiner oder aus unserer modernen Sicht überhaupt erst damit gearbeitet werden.

Nun liegt es in der Natur des Gegenstandes der Psychologie, daß ihre Theorien mehr oder weniger unvollständig sind und daß Verhaltensvorhersagen nur mit einer gewissen Wahrscheinlichkeit, nicht jedoch mit Sicherheit zutreffen. Die Aussagen, die in der Psychologie gemacht werden können, sind also nicht deterministisch, es handelt sich um Wahrscheinlichkeitsaussagen.

3 Die Rolle der Statistik in der Psychologie

Die Statistik als Bereich der angewandten Mathematik liefert Denkmodelle und Verfahren, die es ermöglichen, Entscheidungen und Aussagen im Fall von Unsicherheit zu treffen. Statistische Methoden stellen daher in der psychologischen Forschung und Praxis ein unerläßliches Hilfmittel dar, weil eben alle möglichen Aussagen immer mit einem gewissen Grad von Unsicherheit verbunden sind (und immer sein werden). Diese Methoden ermöglichen es, das Risiko von Fehlentscheidungen beziehungsweise Falschaussagen oder Prognosen zu quantifizieren. Keine Aussage kann mit Sicherheit gemacht werden. Es besteht immer eine bestimmte (aber angebbare) Wahrscheinlichkeit, daß eine Aussage auch falsch sein kann.

In der Psychologie spielt die *Empirie* die tragende Rolle. Empirie ist im eigentlichen Sinne nur wissenschaftlich gewonnene Erfahrung. Diese stellt in der Psychologie (wie in anderen Realwissenschaften auch) den eigentlichen Weg zur Erkenntnis dar. Wenn ganz allgemein über das Verhalten einer bestimmter Gruppe von Menschen (wie beispielsweise der „17-jährigen Burschen, die eine allgemeinbildende höhere Schule besuchen") etwas ausgesagt werden soll, so ist es in der Praxis nicht möglich, alle existierenden, zu dieser bestimmten Gruppe gehörenden Menschen zu beobachten. Man untersucht immer nur einen Teil (eine *Stichprobe*) dieser ausgesuchten Gruppe, um aufgrund der so gewonnenen Daten gültige Aussagen über die gesamte Gruppe (= *die Grundgesamtheit*) machen zu können[1].

Zwei unterschiedliche Bedeutungen des Wortes Statistik

Unter einer Statistik (im Plural: Statistiken) versteht man im Alltagsgebrauch gerne eine geordnete Zusammenstellung von Zahlen, die etwas mit der Wirklichkeit zu tun haben. Die Zuständigkeit für Statistik in diesem Sinne liegt bei Institutionen wie statistischen Ämtern. In einem anderen Sinn, und so wollen wir das Wort im folgenden immer verwenden, spricht man von Statistik, wenn nicht mehr nur Inhalte, sondern die Methoden, mit deren Hilfe man zu Daten gelangen kann, die etwas über die Realität aussagen, gemeint sind. Da der Umfang dieser statistischen Methoden weit größer ist

[1] Es sei angemerkt, daß für den rein theoretischen, in der Praxis völlig unmöglichen Fall der Untersuchung aller zu einer bestimmten Gruppe gehörenden Menschen die Statistik selbstverständlich genauso wichtig bliebe, da durch Statistik ja unter anderem eine übersichtliche Darstellung der gewonnenen Daten überhaupt erst möglich wird.

als das, was man für die Erarbeitung von Statistiken benötigt, zieht man es innerhalb der Psychologie manchmal vor, von quantitativen oder überhaupt von mathematischen Methoden zu sprechen.

Was leistet die Statistik nun konkret in Bezug auf die vorab beschriebenen Ziele der Psychologie?

3.1 Die Aufgabe der deskriptiven Statistik

Mit Hilfe der deskriptiven Statistik (auch *beschreibende* oder *empirische* Statistik) werden vorliegende Daten in geeigneter Weise beschrieben und zusammengefaßt. Aus einer großen Datenmenge wird durch geeignete Methoden eine Tabelle oder eine graphische Darstellung erstellt oder eine passende Einzelkennzahl berechnet.

Beispiel: Umschulung

Nehmen wir an, jemand kann aus irgendeinem Grund (beispielsweise nach einem Verkehrsunfall o. ä.) seinen Beruf nicht mehr ausüben und muß umgeschult werden. Nun hat diese Person einen bestimmten Wunsch, was ihre künftige Tätigkeit angeht. Für deren Ausübung liege ein Eignungstest vor, bei dem Punktwerte von 0 bis 20 erreicht werden können. Als Norm zum Vergleich dienen die Testpunktwerte einiger langjährig in diesem Beruf tätiger Menschen. Es läßt sich also aus diesen Werten eine Häufigkeitsverteilung erstellen (rein tabellarisch oder in weiterer Folge auch graphisch). Sei nun diese Verteilung (graphisch) in etwa eine Glockenkurve, so daß der Durchschnittswert etwa bei 10 Punkten liegt, so kann allein aufgrund des Vergleichs eines neuen Punktwertes (nämlich dem unseres Anwerters), grob der Grad der Eignung dieser Person für die bestimmte Berufsgruppe angegeben werden. Aufgrund der relativen Einfachheit des Verfahrens könnte man es in der Praxis schnell zur Anwendung bringen. Erreicht nun unser Umschulungskandidat etwa 17 von 20 Punkten, so würde man ihn adhoc als eher geeignet für die besagte Tätigkeit bezeichnen. Daß solche Verfahren in der Praxis, die sicherlich nach mehr Informationen verlangt, oft nicht ganz so einfach strukturiert sind, soll hier kurz erwähnt werden. Grundsätzlich jedoch hilft hier bereits eine einfache Methode aus dem Bereich der Deskriptivstatistik, etwas Klarheit zu schaffen.

3.2 Die Aufgabe der Inferenzstatistik

In der Inferenzstatistik (auch *mathematische, schließende* oder *induktive* Statistik) leitet man aus den Daten einer Stichprobe Eigenschaften der Grundgesamtheit ab. Hierzu werden zuerst bestimmte Hypothesen aufgestellt und anschließend überprüft. Die Grundlagen für die erforderlichen Schätz- und Testverfahren liefert die Wahrscheinlichkeitstheorie.

Beispiel: Unterricht in Mengenlehre

Angenommen, ein Schulpsychologe erhält den Auftrag festzustellen, ob Unterricht in Mengenlehre das Abstraktionsvermögen der Kinder fördert. Bei einer Untersuchung dieser Art können natürlich nicht alle Kinder eines Jahrgangs einbezogen werden. Man wird deshalb eine Stichprobe von Schülern, die möglichst repräsentativ für alle Schüler eines Jahrgangs ist, an der Untersuchung teilnehmen lassen. Nehmen wir der Einfachheit halber an, daß zwei Klassen die Untersuchungsstichprobe darstellen. Klasse I und Klasse II werden zu Beginn des Schuljahres mit einem Test zur Erfassung des Abstraktionsvermögens untersucht. Klasse I erhält Unterricht in Mengenlehre (stellt also die sogenannte *Versuchsgruppe* dar), Klasse II erhält „normalen" Mathematikunterricht während des Schuljahres (und stellt somit die *Vergleichsgruppe* dar). Am Ende des Schuljahres wird erneut ein Test zur Erfassung der Abstraktionsfähigkeit durchgeführt um festzustellen, wie sich gegebenenfalls die Fähigkeit der Schüler verändert hat.

Nehmen wir an, diese Untersuchung hätte folgende Ergebnisse gezeigt:

(a) In den Klassen ergab sich zu Beginn des Schuljahres ein Durchschnittswert von 34,0 (Klasse I) beziehungsweise 34,2 (Klasse II).

(b) Nach Ablauf eines Schuljahres stellte der Schulpsychologe in Klasse I einen Durchschnittswert von 40,4 und in Klasse II einen durchschnittlichen Wert in dem Test zur Erfassung des Abstraktionsvermögens von 37,5 fest.

Bei der Interpretation dieser Ergebnisse ist folgendes zu beachten: Wir haben festgestellt, daß das durchschnittliche Abstraktionsvermögen, das zu Beginn der Untersuchung in beiden Klassen annähernd gleich war, nach Ablauf eines Jahres in der Klasse, die Mengenlehreunterricht erhielt, höher ist. Diese Aussage stellt zunächst nur eine schlichte Beschreibung dar: wir betrachten lediglich die Mittelwerte, um uns davon zu überzeugen, daß diese Aussage richtig ist. Wir können nun jedoch noch nichts darüber aussagen, ob Mengenlehreunterricht allgemein im Vergleich zum normalen Mathematikunterricht das Abstraktionsvermögen fördert. Wir können also keine verallgemeinernde Schlußfolgerung anstellen. Genau das ist jedoch das Ziel dieser Untersuchung, denn es geht uns nicht um eine Aussage über den Effekt des Mengenlehreunterrichts in dieser einen Klasse, sondern um eine generelle Aussage. Entscheidend ist bei diesem Anliegen die Frage, ob der Unterschied im Durchschnittswert von Klasse I und Klasse II auf Zufallseinflüsse zurückzuführen ist oder ob ein systematischer Einfluß des Mengenlehreunterrichts angenommen werden kann. Diese Frage ist nicht unmittelbar beantwortbar.

Offenkundig befinden wir uns, wenn wir eine Aussage über den Effekt des Mengenlehreunterrichts machen wollen, in einer Situation der Unsicherheit: Einerseits möchten wir verallgemeinern, andererseits haben wir nur einen vergleichsweise geringen Anteil der Schüler eines Jahrganges in die Untersuchung einbezogen.

In Situationen dieser Art können wir uns der Denkmodelle und Methoden der Inferenzstatistik bedienen. Sie erlauben uns nämlich die Bestimmung des Risikos, das wir bei einer Aussage eingehen. Wir können dieses Risiko in einer quantifizierten Irrtumswahrscheinlichkeit ausdrücken und gelangen dann zu verallgemeinernden Aussagen wie etwa der folgenden: „Es konnte gezeigt werden, daß Mengenlehreunterricht im Vergleich zum normalen Mathematikunterricht das durchschnittliche Abstraktionsvermögen von Schülern erhöht.“ Allerdings kann diese Aussage nicht (und das niemals) mit Sicherheit gemacht werden.

Die Verfahren der Inferenzstatistik erlauben es uns also, unsere Entscheidungen auf eine rationale Basis zu stellen. Die Entscheidungen selbst werden uns damit allerdings nicht abgenommen, da alle Aussagen im Bereich der empirischen Wissenschaften mit einem Fehlerrisiko verbunden sind. Wenn dieses auch beliebig klein wird, so bleibt es doch immer von Null verschieden.

3.3 Die Aufgabe der explorativen Statistik

Die explorative Statistik (auch *hypothesen-generierende* Statistik) befaßt sich mit Zusammenhängen und Unterschieden in vorhandenen Datenbeständen. Methodisch gesehen, stellt sie eine Zwischenform der beiden bisher genannten Teilbereiche der Statistik dar. Mit Hilfe deskriptiver Verfahren und Test-Methoden aus der Inferenzstatistik werden Unterschiede oder Zusammenhänge zwischen Daten lokalisiert. Gleichzeitig wird die Stärke der Ergebnisse bewertet. Diese Ergebnisse können schließlich in Anwendungen dazu verwendet werden, Prognosen, die ähnlich geartete Fälle betreffen, aufzustellen.

3.4 Fazit: Aufgaben und Ziele der Statistik

Wir wollen nun die letzten drei Kapitel zusammenfassen. Auf die Frage, was die Aufgaben und Ziele der Statistik seien, läßt sich grundsätzlich die prägnante Antwort formulieren:

- Beschreiben
- Erklären
- Vorhersagen

Genau dieselbe Auflistung begegnete uns bereits auf Seite 9. Das bedeutet, Psychologie und Statistik wollen, wenn man ihre eigentliche Struktur betrachtet, bereits dasselbe. Auf der einen Seite stellt die Statistik für viele verschiedene Anwendungen des täglichen Lebens ein Werkzeug dar und wird klarerweise nicht nur im Bereich der Psychologie genutzt. Auf der anderen Seite hingegen, und das ist eine Tatsache, die man sich vergegenwärtigen muß, gilt: Wenn von Psychologie die Rede ist, so spielt dabei automatisch Statistik eine wichtige Rolle. Nur die Verwendung statistischer Methoden ermöglicht es überhaupt, Aussagen über bestehende psychologische Sachverhalte zu treffen, sowie Prognosen für künftige, ähnlich beschaffene Fälle abzugeben. Innerhalb der Psychologie sind statistische Methoden ein Teilgebiet der Methodenlehre. Neben diesen geht es dabei noch hauptsächlich um die Planung und Durchführung von Datenerhebungen. Auf die Probleme, die im Zusammenhang damit auftreten können, soll hier nur hingewiesen werden. Diese Bereiche werden später behandelt, da der verwendete Fragebogen (siehe Anhang) ja auch eine mögliche Art der Datenerhebung darstellt.

4 Begriffsklärungen

4.1 Das Messen

Im umgangssprachlichen Sinn spricht man von *Messen* dann, wenn durch diesen Vorgang bestimmte Zahlenwerte festgestellt werden, die dann Eigenschaften wie Größe, Gewicht o. ä. bestimmter Objekte angeben. Hat man also eine bestimmte Variable wie beispielsweise die Variable „Körpergewicht", so deckt sich unsere Alltagsauffassung des Begriffs mit der in der Psychologie gebräuchlichen Verwendung. Allerdings gibt es auch Variablen, die ihrer inneren Struktur nach anders beschaffen sind. Es wird sicherlich öfters eine Rolle spielen, wie viele Versuchspersonen weiblich oder männlich sind. Die Variable „Geschlecht" kann ganz offensichtlich nur diese beiden Werte annehmen. Wollen wir sie statistisch erfassen, so müssen auch ihnen Zahlenwerte zugeordnet werden, zum Beispiel „1" für „männlich" und „2" für „weiblich".

Eine für uns geeignetere Definition des Begriffs des Messens lautet daher:

Definition 4.1. *Das Messen einer Variablen ist die Zuordnung von Zahlen zu den einzelnen Fällen.*

Wie wir gesehen haben, können Variablen ganz unterschiedlich strukturiert sein. In der Psychologie üblich ist deren Einteilung in vier verschiedene Skalenniveaus, die im nächsten Punkt besprochen werden.

4.2 Skalenniveaus

Variablen gehören verschiedenen Skalenniveaus an. Diese werden auch als *Meßniveaus* bezeichnet. Für die Auswahl eines geeigneten statistischen Verfahrens ist die folgende Einteilung maßgebend.

4.2.1 Nominalniveau

Ist eine Variable nominalskaliert, so bedeutet dies, es gibt eine endliche Anzahl von Möglichkeiten, die sie annehmen kann. In einem Fragebogen würde sich ein solcher Fall beispielsweise durch Auswahlmöglichkeiten (*Kategorien*) realisieren lassen, von denen nur eine gewählt werden kann. In Tabelle 1 sind Beispiele angeführt.

Geschlecht			**Familienstand**			**Beruf**		
1	=	männlich	1	=	ledig	1	=	Angestellter
2	=	weiblich	2	=	verheiratet	2	=	Beamter
			3	=	geschieden	3	=	Arbeiter
			4	=	verwitwet	4	=	Selbständiger
						5	=	Hausfrau
						6	=	in Ausbildung
						7	=	Pensionist

Tabelle 1: Beispiele für nominalskalierte Variablen

In allen diesen Fällen hätte die Zuordnung der Zahlen auch anders vorgenommen werden können. Sie ist im Falle einer nominalskalierten Variablen völlig beliebig - ganz im Gegensatz zum Fall des Ordinalniveaus, auf das im folgenden Kapitel eingegangen wird. In dem Spezialfall einer nominalskalierten Variablen mit nur zwei Kategorien spricht man auch von einer *dichotomen Variable*. Typische Beispiele hierfür sind wie in Tabelle 1 das Geschlecht oder auch Zuordungen wie ja/nein, richtig/falsch, trifft zu/trifft nicht zu etc.

Nominalskalierte Variablen sind in ihrer Auswertungsmöglichkeit sehr eingeschränkt. Beispielsweise wäre hier die Berechnung eines Mittelwerts sinnlos. Sie können allerdings einer *Häufigkeitsauszählung* unterzogen werden.

4.2.2 Ordinalniveau

Im Gegensatz zum Nominalniveau wird hier eine Ordnungsrelation wiedergegeben.

Schulbildung		
1	=	Hauptschule
2	=	Berufsschule
3	=	mittlere Schule
4	=	Matura
5	=	Hochschule

Tabelle 2: Beispiel für eine ordinalskalierte Variable

Es läßt sich allerdings keine Angabe darüber machen, wie groß die Unterschiede von einem Wert der Variablen (wie in Tabelle 2 der Schulbildung) zum nächsten sind. Die einzelnen Werte werden einzig geordnet.

4.2.3 Intervallniveau

Wie auch in den ersten beiden Fällen, weist auch hier der Name bereits auf die Bedeutung dieser Einteilung hin. Hier sei mit einem verständlichen Beispiel begonnen: dem Körpergewicht als intervallskalierter Variable. Bezüglich des Körpergewichts geben die entsprechenden Werte nicht nur eine Rangordnung wieder, wie dies bei Schulbildung (und allen anderen ordinalskalierten Variablen) der Fall ist (vgl. Tabelle 2). Auch den Differenzen zweier Werte kommt eine empirische Bedeutung zu, und deshalb auch der Name.

Wiegt nun eine bestimmte Person (Anton) 70 kg, eine weitere (Berta) 80 kg und eine dritte (Cäsar) 90 kg, so kann man bezüglich der Differenzen sagen, daß Berta um gleich viel schwerer ist als Anton verglichen mit Cäsar, der um gleich viel mehr wiegt als Berta. Genau diese Eigenschaft war beim vorher genannten Ordinalniveau eben nicht gegeben.

Intervallskalierte Variablen sind in der Praxis die wichtigste Form. Ihre Bearbeitung unterliegt keinen Einschränkungen. Der Mittelwert wäre beispielsweise ein sinnvoller Wert zur Beschreibung solcher Variablen.

Neben dem Körpergewicht sei als typisches Beispiel noch das Alter genannt. Streng genommen gehören alle Größen, die im Alltag gemessen werden (wie Körpergröße, Gewicht, Alter), bereits der folgenden Kategorie des Verhältnisniveaus an. Dort kommt noch eine Voraussetzung hinzu, so daß alle verhältnisskalierten Variablen per definitionem auch als intervallskaliert aufgefaßt werden können.

4.2.4 Verhältnisniveau

Wie auch hier der Name verrät, kommt nicht nur der Differenz zweier Werte empirische Bedeutung zu, sondern auch deren Verhältnis. Ist Jaquline 30 Jahre alt und Karl bereits 60, so stimmt es definitionsgemäß zu sagen, daß Karl doppelt so alt ist wie Jaquline. Solche Variablen nennt man verhältnisskaliert. Es sind dies alle intervallskalierten Variablen, die den Wert Null annehmen können, wobei dieser gleichzeitig der niedrigste denkbare Wert ist.

Hier kommt diesem Niveau keine allzu große Bedeutung zu. Mit Ausnahme des später vorgestellten geometrischen Mittels (vgl. Kapitel 5.1.2), behandeln wir keine Verfahren, die Verhältnisniveau voraussetzen.

4.2.5 Klassifikation relevanter Variablen

Für die Auswahl geeigneter statistischer Tests empfiehlt es sich, vor einer Auswertung einer Datenmenge eine Klassifikation aller relevanten Variablen vorzunehmen. Diese gedankliche Arbeit kann von keinem noch so speziellen Computerprogramm übernommen werden. In Tabelle 3 wird eine solche Klassifikation vorgenommen. Alle wichtigen Möglichkeiten betreffend der Skalenniveaus sind aufgeführt.

Stufe	**Skalenniveau**
1	nominalskaliert mit mehr als zwei Kategorien
2	nominalskaliert mit zwei Kategorien
3	ordinalskaliert
4	intervallskaliert und nicht normalverteilt
5	intervallskaliert und normalverteilt

Tabelle 3: Variablenklassifikation

Bezüglich der Überprüfung auf Normalverteilung stehen ebenfalls geeignete Tests zur Verfügung (siehe Kapitel 13).

5 Lokalisationsparameter

(auch: *Lageparameter* / *Lagemaße*)

5.1 Der Mittelwert

Der Mittelwert ist der passende Lokalisationsparameter für intervallskalierte und normalverteilte Variablen. Er ist weniger geeignet für nicht normalverteilte oder nur ordinalskalierte Variablen. Trivialerweise ist er unsinnig für nominalskalierte Variablen.

Es gibt drei Varianten des Mittelwerts, welche im folgenden vorgestellt werden.

5.1.1 Arithmetisches Mittel

Das arithmetische Mittel von n Werten x_i ist die Summe dieser Werte, geteilt durch ihre Anzahl n.

Definition 5.1. *Arithmetisches Mittel*

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Hier kann man bereits einen Grund für die Popularität des arithmetischen Mittels erkennen: Man benötigt für seine Berechnung im Prinzip nicht die genauen Einzelwerte. Deren Summe zusammen mit der Anzahl reicht aus.

Als Beispiel seien Gewichtsangaben von $n = 11$ Personen angegeben.

$$65, 78, 83, 74, 105, 92, 88, 65, 95, 112, 73$$

Die Summe dieser Werte beträgt 930. Somit ergibt sich:

$$\overline{x} = \frac{930}{11} = 84,545$$

Die Personen sind also im Mittel 84,545 kg schwer. Als eine praktikabel handhabbare Antwort kann es auch ausreichend sein, den Mittelwert mit 84,5 oder auch nur mit 85 kg anzugeben.

In den Fällen, wo verschiedene Mittelwerte miteinander verglichen werden, sollte man in der Regel drei Nachkommastellen angeben.

Sonderfall Hat man keine Urliste vor sich, sondern eine Häufigkeitstabelle wie in Tabelle 4, so gibt es eine Modifizierung, was die Formel zur Berechnung des arithmetischen Mittels anbelangt.

Note (x_j)	**Häufigkeit (f_j)**	$f_j \cdot x_j$
1	3	3
2	5	10
3	9	27
4	6	24
5	5	25
Summe	28	89

Tabelle 4: Noten einer Schularbeit

In wie weit es sinnvoll ist, bei Schulnoten die Methode des arithmetischen Mittels als Kennwert zu wählen, soll an dieser Stelle kurz diskutiert werden. Grundsätzlich sei erwähnt, daß die Möglichkeit besteht, sie anzuwenden und dies auch allzu oft geschieht. Allerdings ist dabei zu beachten, daß Schulnoten nicht *metrisch* sind, weil die Abstände zwischen den einzelnen Noten sicherlich nicht als gleich anzusehen sind. Dazu kommt noch, daß ein Schüler mit *Sehr gut* dieses vielleicht knapp erreicht hat, während ein anderer mit der gleichen Note möglicherweise bereits in der Schule reif für ein Leistungsstipendium ist. Gleichermaßen gilt dies für die anderen Noten. Dies ist nun also der Grund, warum man bei Schulnoten von der Verwendung des arithmetischen Mittels absehen sollte.

Die angesprochene modifizierte Formel lautet jedenfalls:

$$\overline{x} = \frac{\sum_{j=1}^{k} f_j \cdot x_j}{\sum_{j=1}^{k} f_j}$$

Wie man sieht, wird hier nur die Multiplikation für die Zusammenfassung der Werte im Zähler benutzt, während man weiterhin durch die Gesamtanzahl (hier nun die Summe der Einzelhäufigkeiten) dividiert.

In unserem Beispiel beträgt der Mittelwert

$$\overline{x} = \frac{89}{28} = 3,179$$

[Eigenschaften des Mittelwerts]Zwei wichtige Eigenschaften des Mittelwerts

- Die Summe der Differenzen aller Werte von ihrem gemeinsamen Mittelwert ist null.

  Wir stellen folgende Gleichung auf:
  $\sum\limits_{i=1}^{n} (x_i - x) = 0$
  Nun fragen wir: Für welches x gilt diese Eigenschaft?
  $\sum\limits_{i=1}^{n} (x_i - x) = \sum\limits_{i=1}^{n} x_i - \sum\limits_{i=1}^{n} x = \sum\limits_{i=1}^{n} x_i - n \cdot x$
  Für den Mittelwert gilt: $\overline{x} = x = \frac{1}{n} \cdot \sum\limits_{i=1}^{n} x_i$
  Daraus folgt: $\sum\limits_{i=1}^{n} x_i - n \cdot \frac{\sum\limits_{i=1}^{n} x_i}{n} = 0$
  Die aufgestellte Gleichung stimmt also genau für den Wert,
  den wir eingesetzt haben, nämlich den Mittelwert.

- Die Summe der Quadrate der Differenzen aller Werte von ihrem Mittelwert ist kleiner als die Summe der Quadrate der Differenzen aller Werte zu irgendeinem anderen Wert, sie ist also minimal.

  $\sum\limits_{i=1}^{n} (x_i - x)^2 \rightarrow \text{Minimum}$
  Der Ausdruck wird einmal abgeleitet und Null gesetzt:
  $\sum\limits_{i=1}^{n} 2 \cdot (x_i - x) \cdot (-1) = 0$
  $\sum\limits_{i=1}^{n} (x_i - x) = 0$
  Daß diese Eigenschaft genau für den Mittelwert gilt,
  wurde im vorhergehenden Punkt gezeigt.

Ein Vorteil des arithmetischen Mittels gegenüber dem Median (vgl. Kapitel 5.2) liegt darin, daß die Information der Stichprobe effizienter ausgenützt wird. So kann man etwa vom arithmetischen Mittel und der Anzahl der Werte auf die Wertesumme schließen.

Abschließend sei noch einmal darauf hingewiesen, daß die Berechnung des Mittelwerts bei nominalskalierten Variablen unsinnig ist. Wollten wir ihn etwa anwenden auf Werte zu den Kategorien aus Tabelle 1 von Seite 16 (Geschlecht, Familienstand, Beruf), so kann das niemals sinnvoll sein, denn ein Wert von beispielsweise 3,4 kann ja nicht mehr sinnvoll interpretiert werden. Bereits wegen solcher Feinheiten ist die getroffene Einteilung möglicher Daten in verschiedene Skalenniveaus sinnvoll und notwendig.

5.1.2 Geometrisches Mittel

Das geometrische Mittel dient zum Beispiel der Ermittlung von Wachstumsraten aufeinanderfolgender Perioden. Seine Berechnung setzt Verhältnisniveau voraus (vgl. Seite 17). In der Tat wird diese Methode meist in Verbindung mit Wachstumsraten angewandt.

Definition 5.2. *Geometrisches Mittel*

$$\overline{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

Tabelle 5 gibt Preise für ein bestimmtes Produkt zu bestimmten Zeitpunkten (Jahren) sowie die zugehörigen Wachstumsfaktoren an.

Jahr	Preis	Wachstumsfaktor
1997	1425	
1998	1490	1,046
1999	1535	1,030
2000	1570	1,023
2001	1840	1,172
2002	1950	1,060
2003	2100	1,077

Tabelle 5: Wachstumsfaktoren

Das geometrische Mittel der Wachstumsfaktoren ist

$$\overline{x}_G = \sqrt[6]{1,046 \cdot 1,030 \cdot 1,023 \cdot 1,172 \cdot 1,060 \cdot 1,077} = 1,067$$

6,7 % ist also die konstante Rate, die zum gleichen Gesamtwachstum über alle Jahre hinweg geführt hätte, wie man leicht testen kann, da bis auf Rundungsfehler gilt:

$$1425 \cdot 1{,}067^6 \approx 2100$$

5.1.3 Harmonisches Mittel

Definition 5.3. *Harmonisches Mittel*

$$\overline{x}_H = \frac{n}{\dfrac{1}{x_1} + \dfrac{1}{x_2} + \cdots + \dfrac{1}{x_n}}$$

Die dritte Möglichkeit eines Mittelwerts sei hier anhand eines praktischen Beispiels vorgestellt.

Mittlere Geschwindigkeiten Bei diesen führt die Methode des arithmetischen Mittels nicht zum korrekten Ergebnis.

Beispiel: Ein Zug fährt von A nach B (100 km) mit einer durchschnittlichen Geschwindigkeit von 80 km/h und von B nach C (ebenfalls 100 km) mit durchschnittlich 120 km/h. Das möglicherweise augenscheinliche Ergebnis von 100 km/h für die Durchschnittsgeschwindigkeit ist falsch. Es ist hier nicht das arithmetische, sondern das harmonische Mittel zu wählen:

$$\overline{x}_H = \frac{2}{\frac{1}{80} + \frac{1}{120}} = 96$$

Die durchschnittliche Geschwindigkeit für die Gesamtstrecke beträgt also 96 km/h, wie leicht nachgerechnet werden kann:

$$t = \tfrac{200}{96} = 2{,}083 \quad t = \tfrac{100}{80} + \tfrac{100}{120} = 2{,}083$$

Insgesamt wurden nämlich 200 km zurückgelegt und dabei für das erste Teilstück $\frac{100}{80}$ Stunden verbraucht, für das zweite Teilstück $\frac{100}{120}$ Stunden. Daher berechnet sich die mittlere Geschwindigkeit für die Gesamtstrecke wie folgt:

$$\frac{200}{\frac{100}{80} + \frac{100}{120}} = \frac{2}{\frac{1}{80} + \frac{1}{120}}$$

Auf der rechten Seite erhalten wir wiederum die Anwendung der Formel zur Berechnung des harmonischen Mittels.

5.1.4 Berechnung eines gemeinsamen arithmetischen Mittelwerts verschiedener Stichproben

Bei der Berechnung des gemeinsamen Mittelwerts zweier Stichproben, von denen die Mittelwerte $\overline{x}_1$ und $\overline{x}_2$ sowie die dazugehörigen Fallzahlen n_1 und n_2 vorliegen, müssen diese Mittelwerte entsprechend gewichtet werden:

$$\overline{x} = \frac{n_1 \cdot \overline{x}_1 + n_2 \cdot \overline{x}_2}{n_1 + n_2}$$

Die Gewichtung erfolgt analog bei k Mittelwerten ($k \in \mathbb{N}, \quad k > 2$):

$$\overline{x} = \frac{n_1 \cdot \overline{x}_1 + n_2 \cdot \overline{x}_2 + \ldots + n_k \cdot \overline{x}_k}{n_1 + n_2 + \ldots + n_k}$$

Zusammenhang zwischen arithmetischem und geometrischem Mittel

Für den Fall $n = 2$ gilt:

$$\sqrt{x_1 \cdot x_2} \leq \frac{x_1 + x_2}{2}$$

$$\begin{aligned}
x_1 \cdot x_2 &\leq \frac{x_1^2 + 2 \cdot x_1 \cdot x_2 + x_2^2}{2} \\
\cancel{x_1 \cdot x_2} &\leq \frac{x_1^2}{2} + \frac{x_2^2}{2} + \cancel{x_1 \cdot x_2} \\
0 &\leq \frac{x_1^2 + x_2^2}{2}
\end{aligned}$$

$\square$

5.2 Der Median

Der Median wird bei ordinalskalierten bzw. intervallskalierten, aber nicht normalverteilten Variablen berechnet.

Definition 5.4. *Median*

$$m = \begin{cases} x_{\frac{n+1}{2}} & \textit{falls n ungerade} \\[2ex] \dfrac{x_{\frac{n}{2}+1} + x_{\frac{n}{2}}}{2} & \textit{falls n gerade} \end{cases}$$

Der Median ist derjenige Wert, unterhalb und oberhalb dessen jeweils die

Hälfte der Meßwerte liegen. Es handelt sich hiebei im wahrsten Sinne um den Wert in der Mitte der (geordneten) Meßwerte.

Liegen einzelne Messwerte vor, bei denen *Ausreißer*[2] vorkommen, so wird der Median angewandt.

Hat man beispielsweise eine Urliste (es könnten gemessene Zeiten in Sekunden verschiedener Probanden sein, die eine bestimmte Aufgabe bewältigen mußten wie in Tabelle 6) wie die folgende, so ordnet man sie zuerst der Größe nach:

Werte ungeordnet	489 113 141 120 217 109 675 218 96 225 132
Werte geordnet	96 109 113 120 132 141 217 218 225 489 675

Tabelle 6: Dauer in Sekunden verschiedener Probanden beim Lösen eines Problems

Aufgrund der beiden Ausreißerwerte (489, 675) erscheint es sinnvoll, den Median zu wählen. Da hier elf Messwerte vorliegen, ist der sechste von ihnen genau der Wert in der Mitte. Somit beträgt der Median

$$m = 141$$

Davor und danach (oder links und rechts) liegen gleich viele Werte, nämlich fünf.

Im Falle einer geraden Anzahl von Meßwerten erhält man den Median als arithmetisches Mittel der beiden Werte „in der Mitte“, unterhalb und oberhalb derer jeweils die Hälfte der Meßwerte liegen (vgl. Definition 5.4).

Beispiel: Wir fügen in Tabelle 6 noch einen Wert (690) an, um eine gerade Anzahl zu erhalten.

Werte geordnet	96 109 113 120 132 141 217 218 225 489 675 690

Somit erhalten wir den Median wie folgt:

$$m = \frac{141 + 217}{2} = 179$$

Die stärkste Eigenschaft des Medians ist offensichtlich: Er ist gänzlich *unempfindlich gegenüber der Größe der Ausreißer*, da der größte oder auch der kleinste Wert ja keinen Einfluß auf die Berechnung des Medians nehmen.

[2] Werte, die augenscheinlich viel kleiner oder größer sind als das Hauptfeld

Ein weiterer Vorteil des Medians ist die Tatsache, daß er im Vergleich zum arithmetischen Mittel einfach zu berechnen ist.

Und schließlich verläßt der Median auch nicht seine Ausgangsmenge, sieht man von der Ausnahme ab, daß bei einer geraden Fallzahl das arithmetische Mittel der beiden in der Mitte der geordneten Datenmenge liegenden Werte berechnet. Dies ist in der Praxis oftmals anschaulicher. Im Gegensatz dazu kommt es beim arithmetischen Mittel allzu oft vor, daß etwa eine Familie 1,34 Kinder besitzt oder ähnliches.

Als letzte positive Eigenschaft sei nochmals genannt, daß der Median auch bei nichtmetrischen Daten anwendbar ist, also bei ordinalskalierten Variablen.

Sonderfall Berechnung des Medians bei ordinalskalierten Variablen, die als gehäufte Daten vorliegen

Antwort	**Skalenwert**	**Häufigkeit**	**kumulierte Häufigkeit**
trifft gar nicht zu	1	12	12
trifft eher nicht zu	2	25	37
trifft manchmal zu	3	23	60
trifft ziemlich zu	4	53	113
trifft voll zu	5	47	160

Tabelle 7: Beobachtete und kumulierte Häufigkeiten

Hier wäre die bekannte Berechnung des Medians selbstverständlich möglich. Er beträgt $m = 4$. Da es 160 Werte gibt, wäre der Mittelwert aus dem 80. und dem 81. Wert zu berechnen. Wie wir aus Tabelle 7 ablesen können, betragen beide Werte 4.

Da dies allerdings etwas ungenau erscheint, benutzt man folgende verfeinerte Formel:

$$m = x_m - 0,5 + \frac{1}{f_m} \cdot \left(\frac{n}{2} - F_{m-1} \right)$$

wobei	x_m	......	Wert der m-ten Kategorie
	f_m	......	Häufigkeit der m-ten Kategorie
	F_{m-1}	......	kumulierte Häufigkeit bei der Kategorie $m - 1$
	n	......	Gesamtsumme der Häufigkeiten

Der Index m steht hier für genau diejenige Kategorie, in welcher der Median liegt.

In diesem Beispiel sind die Werte:

$$m = 4 \qquad x_m = 4 \qquad f_4 = 53 \qquad F_3 = 60 \qquad n = 160$$

Somit berechnet sich der Median wie folgt:

$$m = 4 - 0,5 + \frac{1}{53} \cdot \left(\frac{160}{2} - 60 \right) = 3,877$$

Den Mittelwert berechnen wir wie gewohnt:

$$\overline{x} = \frac{\sum\limits_{m=1}^{k} f_m \cdot x_m}{n} = \frac{578}{160} = 3,613 \quad \text{mit k} \ldots \text{Anzahl der Kategorien}$$

Anmerkung: Der Mittelwert ist kleiner als der Median, da es sich hier um eine rechtsgipflige Verteilung handelt. Hätten wir eine linksgipflige Verteilung, so wäre der Mittelwert größer als der Median.

5.3 Der Modus

Ein anderes, nicht sonderlich brauchbares Lagemaß ist der sogenannte Modus.

Definition 5.5. *Der Modus ist der am häufigsten auftretende Wert einer Stichprobe. Er muß nicht eindeutig existieren.*

6 Dispersionsparameter

(auch: *Streuungsparameter / Streuungsmaße*)

Hier wird - im Gegensatz zu den Lokalisationsparametern, wo die Lage einer Verteilung oder ihre zentrale Tendenz beschrieben werden - die *Breite einer Verteilung* betrachtet.

6.1 Spannweite

Die Spannweite errechnet sich als Differenz zwischen Maximum und Minimum.

Definition 6.1. *Spannweite = größter Wert minus kleinster Wert*

Der Vorteil dieses Streuungsmaßes liegt augenscheinlich in seiner Einfachheit. Für einen ersten Überblick wird die Spannweite herangezogen. Der Nachteil hingegen liegt ebenfalls in der Einfachheit, da nur zwei Werte (die Extrema) zur Berechnung herangezogen werden. Alle dazwischen liegenden Werte bleiben unberücksichtigt, weshalb aussagekräftigere Maße hinsichtlich der Streuung entwickelt werden.

6.2 Standardabweichung und Standardfehler

Die Standardabweichung ist das gebräuchlichste Streuungsmaß. Sie kommt bei intervallskalierten und normalverteilten Variablen zum Einsatz (analog dem Mittelwert als geeigneten Lokalisationsparameter).

Definition 6.2. *Standardabweichung*

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n - 1}}$$

Teilweise, für die Praxis aber eher unüblich, wird im Nenner auch n statt $n-1$ angegeben. Bei größeren Fallzahlen ist der daraus resultierende Unterschied der Ergebnisse vernachlässigbar.

Verschiebungssatz von Steiner Durch eine Umformung ergibt sich:

$$\begin{aligned}
\sum_{i=1}^{n}(x_i-\overline{x})^2 &= \sum_{i=1}^{n}\left(x_i^2+\overline{x}^2-2\cdot x_i\cdot\overline{x}\right)= \\
&= \sum_{i=1}^{n}x_i^2+n\cdot\overline{x}^2-2\cdot\overline{x}\cdot\sum_{i=1}^{n}x_i= \\
&= \sum_{i=1}^{n}x_i^2+\cancel{n}\cdot\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n^{\cancel{2}}}-2\cdot\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}= \\
&= \sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}
\end{aligned}$$

und somit gilt:

$$s=\sqrt{\frac{\sum_{i=1}^{n}x_i^2-\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{n}}{n-1}}$$

Für die Rechenpraxis hat sich diese Modifizierung der Formel bewährt. Sie stellt lediglich eine Umformung dar.

x_i	x_i^2
40	1600
37	1369
67	4489
23	529
45	2025
39	1521
29	841
51	2601
56	3136
24	576
42	1764
38	1444
491	21895

Tabelle 8: Altersangaben, deren Summe und die Summe deren Quadrate

So kann die Standardabweichung mit Hilfe der umgewandelten Formel berechnet werden:

$$s = \sqrt{\frac{21895 - \frac{491^2}{12}}{12 - 1}} = 12,81$$

Anmerkung: Vergleicht man zwei Standardabweichungen miteinander, so ist dies nur sinnvoll, wenn die Mittelwerte ähnlich sind. Eine Standardabweichung von 5 hat bei einem Mittelwert von 10 (Abweichung durchschnittlich um die Hälfte des Mittelwerts!) klarerweise eine völlig andere Bedeutung als bei einem Mittelwert von 100 (Abweichung durchschnittlich nur um ein Zwanzigstel des Mittelwerts).

Um die Standardabweichung zu relativieren, wird folgendermaßen der Variationskoeffizient definiert.

Definition 6.3. *Variationskoeffizient*

$$V = \frac{s}{\overline{x}}$$

Die Relativierung wird also mit der Division durch den Mittelwert erreicht.

Im Beispiel von Tabelle 8 ergibt sich:

$$V = \frac{12,81}{40,9} = 0,313$$

Dieser Variationskoeffizient wird benutzt, um Standardabweichungen zwischen Stichproben zu vergleichen, deren Mittelwerte deutlich voneinander abweichen.

Ein weiteres Maß für die Streuung ist die *Varianz*.

Definition 6.4. *Varianz*

$$V = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n - 1}$$

Berechnet wird sie als Quadrat der Standardabweichung.

6.2.1 Berechnung einer gemeinsamen Standardabweichung verschiedener Stichproben

Ähnlich wie beim Mittelwert (vgl. Seite 24) können Standardabweichungen verschiedener Stichproben zu einer gemeinsamen Standardabweichung zusammengeführt werden.

$$s = \sqrt{\frac{1}{N-1} \cdot \left[\sum_{j=1}^{k} \left(\left(n_j - 1 \right) \cdot s_j^2 \right) + \sum_{j=1}^{k} \left(n_j \cdot \left(\bar{x}_j - \bar{x} \right)^2 \right) \right]}$$

wobei	k	......	Anzahl der Stichproben
	n_j	......	Umfänge der Stichproben
	$\bar{x}_j$	......	Mittelwerte der Stichproben
	s_j	......	Standardabweichungen der Stichproben
	$\bar{x}$	......	Gesamtmittelwert der Stichproben
	N	......	Summe der Stichprobenumfänge $N = \sum_{j=1}^{k} n_j$

An dieser Stelle sei eine wichtige **Faustregel** angemerkt:

Im Intervall $[\bar{x} - s; \quad \bar{x} + s]$ liegen etwa zwei Drittel (67 %) aller Werte, im Intervall $[\bar{x} - 2 \cdot s; \quad \bar{x} + 2 \cdot s]$ etwa 95 % aller Werte.

Die Standardabweichung erlaubt über die Angabe eines so genannten Konfidenzintervalls auch eine Voraussage über den Mittelwert der betreffenden Grundgesamtheit (siehe Kapitel 12.2). Dazu wird später ein etwas abgeändertes Streuungsmaß benötigt: der *Standardfehler des Mittelwerts* oder kurz *Standardfehler*.

Definition 6.5. *Standardfehler*

$$s_m = \frac{s}{\sqrt{n}}$$

6.3 Der Quartilabstand

Erinnern wir uns an den Median, welcher jener Wert ist, ober dem und unter dem 50 % aller Werte liegen. Nach derselben Überlegung definieren wir das 1. Quartil (Q1) sowie das 3. Quartil (Q3).

Definition 6.6. *Quartile*

Unterhalb des 1. Quartils liegen 25 %, unterhalb des 3. Quartils 75 % der Werte.

Der Median ist also - in diesem Zusammenhang betrachtet - das 2. Quartil. Das 1. Quartil, der Median und das 3. Quartil teilen die Meßwertskala also in vier Teile mit gleichen Häufigkeiten ein.

Zwischen Q1 und Q3 liegen 50 % der Werte. Somit ist der Quartilenabstand per se ein Maß für die Streuung. In der Praxis ist allerdings der mittlere Quartilenabstand wichtiger.

Definition 6.7. *Mittlerer Quartilabstand*

$$QA = \frac{Q3 - Q1}{2}$$

Der mittlere Quartilenabstand wird als Streuungsmaß dort verwendet, wo anstelle des Mittelwerts der Median als Lageparameter benutzt wird. Das ist deshalb der Fall, weil bei beiden die Reihenfolge der Werte und die so definierten Kennwerte (Q1, Median, Q3) eine Rolle spielen. Analog entsprechen Mittelwert und Standardabweichung einander, wenn eine Datenmenge beschrieben wird.

Wie auch der Median bei gehäuften Daten durch eine modifizierte Formel berechnet werden kann (vgl. Seite 26), besteht diese Möglichkeit auch beim 1. Quartil sowie beim 3. Quartil.

$$Q1 = x_m - 0,5 + \frac{1}{f_m} \cdot \left(\frac{n}{4} - F_{m-1} \right)$$

$$Q3 = x_m - 0,5 + \frac{1}{f_m} \cdot \left(\frac{3 \cdot n}{4} - F_{m-1} \right)$$

Bei nicht gruppierten Werten, also intervallskalierten Variablen, werden Q1 und Q3 wie beim Median durch Auszählen bestimmt.

7 Statistische Wahrscheinlichkeit

Definition 7.1. *Statistische Wahrscheinlichkeit*

Tritt unter n Versuchen ein Ereignis k-mal auf und nähert sich mit größer werdendem n die relative Häufigkeit

$$\frac{k}{n}$$

einer festen Zahl, so wird diese Zahl als (statistische) Wahrscheinlichkeit dieses Ereignisses bezeichnet.

Die Tatsache, daß sich bei immer größer werdender Versuchszahl die relative Häufigkeit eines Ereignisses immer mehr einem festen Wert annähert, wird als *empirisches Gesetz der großen Zahlen* bezeichnet. Es sei allerdings darauf hingewiesen, daß dies für das einzelne Ereignis völlig ohne Relevanz bleibt. Hat man beim Roulette schon n mal auf Schwarz gesetzt (mit n sehr groß), heißt das in keinem Fall, daß die Wahrscheinlichkeit dafür steigt, daß im $(n + 1)$-ten Fall nicht mehr Schwarz kommen wird.

8 Zufallsvariablen

Variablen werden auch als Zufallsvariablen bezeichnet, um klar zu machen, daß deren mögliche Werte (auch: *Ausprägungen* oder *Realisationen*) Ergebnisse eines Zufallsvorgangs sind.

Zufallsvariablen werden üblicherweise mit Großbuchstaben bezeichnet, ihre Ausprägungen mit Kleinbuchstaben.

Faßt man alle möglichen Werte in einer Menge zusammen, spricht man vom *Ereignisraum*, der meist mit dem griechischen Großbuchstaben Ω bezeichnet wird. Dieser Ereignisraum kann auch eine unendliche Menge sein. Betrachten wir beispielsweise die Variable Körpergewicht oder ähnliche, so kommen mathematisch gesehen (je nach Meßgenauigkeit) theoretisch unendlich viele Zahlen in Betracht. In der Praxis werden die möglichen Werte aber eine endliche Menge bilden, da wir etwa die Körpergröße in Metern mit maximal zwei Nachkommastellen angeben. Noch einfacher sind typische Beispiele wie Geschlecht ($\Omega = \{\textit{männlich}, \textit{weiblich}\}$) oder das Alter in Jahren.

Definition 8.1. *Zufallsvariable*

Eine Variable X, deren Werte (Ausprägungen) x_i aus dem zugeordneten Ereignisraum Ω die Ergebnisse eines Zufallsvorgangs sind, bezeichnet man als Zufallsvariable.

Von diskreten Werten sprechen wir in diesem Zusammenhang dann, wenn der Ereignisraum nur endlich viele oder abzählbar unendlich viele Ausprägungen besitzt.

Definition 8.2. *Diskrete Zufallsvariable*

Eine Zufallsvariable X heißt diskret, wenn sie nur endlich viele oder abzählbar unendlich viele Werte x_i annehmen kann.

Im Gegensatz dazu stehen stetige Zufallsvariablen.

Definition 8.3. *Stetige Zufallsvariable*

Eine Zufallsvariable X heißt stetig, wenn sie zumindest in einem bestimmten Bereich jeden reellen Zahlenwert annehmen kann.

Beispiele diesbezüglich sind Längen oder Größen aller Art oder eine bestimmte Zeitspanne. Diese Werte können ja im Prinzip alle reellen Werte annehmen.

Querverbindung zu Skalenniveaus Nominal- und ordinalskalierte Variablen sind stets diskret, bei intervall- und verhältnisskalierten Variablen entscheidet es sich insofern, als es auf die Meßgenauigkeit ankommt, ob man sie dem diskreten oder dem stetigen Bereich zuordnet. Tendenziell werden solche Variablen als stetig eingeordnet.

9 Wahrscheinlichkeits- und Verteilungsfunktion

9.1 Der diskrete Fall

Definition 9.1. *Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen*

Die Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen ist die Funktion $f(x_i)$, die für jede Ausprägung der Zufallsvariablen die Wahrscheinlichkeit ihres Auftretens angibt.

Hat der Ereignisraum n Ausprägungen, so gilt für die Wahrscheinlichkeitsfunktion:

$$\sum_{i=1}^{n} f(x_i) = 1$$

Definition 9.2. *Verteilungsfunktion einer diskreten Zufallsvariablen*

Die Verteilungsfunktion einer diskreten Zufallsvariablen berechnet sich aus ihrer Wahrscheinlichkeitsfunktion $f(x_i)$ zu

$$F(x_i) = \sum_{j=1}^{i} f(x_j)$$

Der Graph einer Verteilungsfunktion einer diskreten Zufallsvariablen ist somit der einer Treppenfunktion, die an den jeweiligen Stellen x_i zum nächsten Wert springt.

9.2 Der stetige Fall

Definition 9.3. *Dichtefunktion einer stetigen Zufallsvariablen*

Die Dichtefunktion $f(x_i)$ einer stetigen Zufallsvariablen hat die Eigenschaft

$$F(a \leq x \leq b) = \int_{a}^{b} f(x)\mathrm{d}x$$

Hieraus folgt analog zur Eigenschaft der Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen:

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$$

Definition 9.4. *Verteilungsfunktion einer stetigen Zufallsvariablen*
Die Verteilungsfunktion einer stetigen Zufallsvariablen ist gegeben durch

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$$

Die Verteilungsfunktion einer stetigen Zufallsvariablen wird somit durch das Integral zwischen dem linken Ende der Verteilung und dem betreffenden Wert x bestimmt.

10 Verteilungen

10.1 Diskrete Verteilungen

10.1.1 Gleichverteilung

Bei einer gleichverteilten Zufallsvariablen ist die Wahrscheinlichkeitsfunktion für alle n Ausprägungen gleich:

$$f(x_1) = f(x_2) = \ldots = f(x_n) = \frac{1}{n}$$

Einfache Beispiele hierfür sind das Werfen eines Würfels ($f(x_i) = \frac{1}{6}$) oder das Setzen auf eine bestimmte Zahl beim Roulette ($f(x_i) = \frac{1}{37}$).

10.1.2 Binomialverteilung

Im Rahmen der Binomialverteilung interessieren wir uns für Ereignisse, die in zwei Alternativen auftreten, wobei diese Alternativen gleich oder ungleich wahrscheinlich sein können. Beispielsweise sind die beiden möglichen Ereignisse beim Münzwurf (Kopf, Zahl) in ihrem Auftreten gleichwahrscheinlich ($f(x_1) = f(x_2) = \frac{1}{2}$), wobei hingegen die beiden Alternativen $x_1 =$ „Würfle eine 6 (oder jede beliebige andere beim Würfeln mögliche Zahl)“ und $x_2 =$ „Würfle keine 6“ verschieden wahrscheinlich sind ($f(x_1) = \frac{1}{6}$, $f(x_2) = \frac{5}{6}$).

Ein Zufallsexperiment, bei dem bei wiederholter Durchführung ein bestimmtes Ereignis mit derselben konstanten Wahrscheinlichkeit p auftritt, nennt man ein *Bernoulli-Experiment*. Eine solche Abfolge von zufälligen alternativen, voneinander unabhängigen Ereignissen, heißt demnach ein *Bernoulli-Prozeß*. Wenn n Versuche durchgeführt werden, kann ein bestimmtes Alternativereignis A i-mal ($i = 1 \ldots n$) auftreten. Die Binomialverteilung ist die Wahrscheinlichkeitsfunktion f für die Zufallsvariable „Häufigkeiten des Auftretens von A bei n Bernoulli-Versuchen“.

Definition 10.1. *Binomialverteilung*

A trete mit einer Wahrscheinlichkeit von p ein. Die Wahrscheinlichkeit, k-mal A in n Versuchen zu erhalten, errechnet sich als

$$f(X = k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$$

10.1.3 Hypergeometrische Verteilung

In einer Urne mögen sich N Elemente befinden, davon M solche, die mit einem besonderen Merkmal gekennzeichnet sind. Mithilfe der hypergeometrischen Verteilung läßt sich die Frage beantworten, wie groß die Wahrscheinlichkeit dafür ist, daß sich unter n gezogenen Elementen *ohne Zurücklegen* genau x Elemente befinden, die das besondere Merkmal aufweisen.

Definition 10.2. *Hypergeometrische Verteilung*

Diese Wahrscheinlichkeit ist nach der hypergeometrischen Verteilung

$$f(x, n, M, N) = \frac{\binom{M}{x} \cdot \binom{N-M}{n-x}}{\binom{N}{n}}$$

10.1.4 Poisson-Verteilung

Die Poisson-Verteilung geht für kleine Ereigniswahrscheinlichkeiten p und große Versuchszahl n aus der Binomialverteilung hervor. Sie wird daher auch die *Verteilung seltener Ereignisse* genannt.

Definition 10.3. *Poisson-Verteilung*

Die Wahrscheinlichkeit, daß ein Ereignis, das bei einem einmaligen Versuch mit der Wahrscheinlichkeit p eintritt, bei n Versuchen genau k-mal auftritt, ist nach der Poisson-Verteilung

$$f(n, k, p) = \frac{(n \cdot p)^k}{e^{n \cdot p} \cdot k!}$$

10.2 Stetige Verteilungen

10.2.1 Normalverteilung

Eine entscheidende Rolle in der Statistik spielt bei intervallskalierten Variablen die Tatsache, ob deren Werte einer Normalverteilung folgen oder nicht. Danach richtet sich, welche statistischen Kennwerte zu ihrer Beschreibung verwendet werden können beziehungsweise welche analytischen Tests gegebenenfalls bei einer Hypothesenprüfung zur Anwendung kommen. All das wird später noch behandelt.

Bei der Normalverteilung handelt es sich um eine *eingipflige und symmetrische Verteilung*. Sie wurde von Carl Friedrich Gauß entdeckt. Beschreibt man diese Verteilung mit einer Kurve, so stellt diese idealisiert eine Glocke dar. Daher kommt der Name *Gaußsche Glockenkurve* für die graphische Darstellung der Normalverteilung. Die Form dieser Glockenkurve ist durch die folgende Dichtefunktion gegeben:

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

Dabei ist μ der Mittelwert und σ die Standardabweichung der Verteilung.

Zu jedem Paar von μ und σ gibt es also eine Normalverteilung. Die Funktionen haben ihr Maximum bei $x = \mu$ und sind umso schlanker, je kleiner die Standardabweichung σ ist.

Die Fläche unter jeder Normalverteilungskurve ist jeweils gleich 1:

$$\int_{-\infty}^{\infty} f(t)\mathrm{d}t = 1$$

Die Verteilungsfunktion ist nun

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$$

und somit unter Verwendung obiger Formel

$$F(x) = \frac{1}{\sigma \cdot \sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^{x} e^{-\frac{1}{2} \cdot \left(\frac{t-\mu}{\sigma}\right)^2} \mathrm{d}t$$

Da eine Berechnung von $F(x)$ aus der gegebenen Integralformel ohne Computer respektive entsprechendes Computerprogramm nicht möglich ist, behilft man sich teilweise noch mit tabellierten Werten. Diese Werte gehören zu genau der Normalverteilung, bei der $\mu = 0$ und $\sigma = 1$ gilt. Diese Normalverteilung wird *Standardnormalverteilung* genannt und hat die Verteilungsfunktion

$$\Phi(z) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^{z} e^{-\frac{1}{2} \cdot t^2} \mathrm{d}t$$

Die Werte $\Phi(z)$ und $\Phi(-z)$ sind für z-Werte > 0 tabelliert. Aus Symmetriegründen gilt dabei

$$\Phi(-z) = 1 - \Phi(z)$$

Vor Gebrauch der z-Tabelle werden die Variablen einer *z-Standardisierung* unterzogen:

$$z = \frac{x - \overline{x}}{s}$$

Dabei sind $\overline{x}$ der Mittelwert und s die Standardabweichung der Stichprobe.

Letztlich entscheidend für die Beantwortung der Frage, ob die gegebene Häufigkeitsverteilung der Werte einer Variablen einer Stichprobe als normalverteilt angesehen werden kann, ist der Sachverhalt, ob sich diese Verteilung *signifikant* (vgl. Kapitel 11.3) von einer Normalverteilung unterscheidet oder nicht. Passende Tests dazu bietet Kapitel 13.

10.2.2 Exponentialverteilung

Ein exponentieller Abfall ist vor allem bei Zeitdauern zu beobachten (Lebenszeiten, Wartezeiten, Bearbeitungszeiten).

Definition 10.4. *Exponentialverteilung*

Eine exponentialverteilte, stetige Zufallsvariable hat die Dichtefunktion

$$f(x, \lambda) = \lambda \cdot e^{-\lambda \cdot x} \qquad mit \quad x \geq 0 \quad und \quad \lambda > 0$$

Der Parameter λ steuert, wie schnell die Exponentialfunktion für große Werte von x gegen Null geht.

Aus der Dichtefunktion berechnet sich die Verteilungsfunktion zu

$$F(x, \lambda) = 1 - e^{-\lambda \cdot x}$$

Anmerkung: Die Exponentialverteilung eignet sich zur Analyse von Zeitdauern nur dann, wenn für jeden Zeipunkt die noch verbleibende Zeitdauer nicht von der bereits verstrichenen Zeitdauer abhängt. Man spricht daher von der Exponentialverteilung als einer *gedächtnislosen* oder *nicht alternden Verteilung.*

11 Grundlagen der analytischen Statistik

Die analytische Statistik (auch: *schließende Statistik, Inferenzstatistik* - vgl. Kapitel 3.2) befaßt sich mit dem Problem, wie aufgrund von Ergebnissen, die anhand einer vergleichsweise kleinen Zahl von Personen gewonnen wurden[3], möglichst allgemeingültige Aussagen hergeleitet werden können.

Die grundlegenden Termini, die wir verwenden, sind *Stichprobe* und *Grundgesamtheit*. Die analytische Statistik versucht, von den Verhältnissen der Stichprobe auf die Verhältnisse in der Grundgesamtheit zu schließen.

Man unterscheidet zwei grundlegende Möglichkeiten, die wir im folgenden näher betrachten werden:

- den Schluß von den Kennwerten der Stichprobe auf die entsprechenden Werte der Grundgesamtheit
- die Überprüfung von Hypothesen

11.1 Schluß von der Stichprobe auf die Grundgesamtheit

Üblicherweise wird so vorgegangen, daß ein *Konfidenzintervall* (siehe Kapitel 12.2) vorgegeben wird, innerhalb dessen der untersuchte Wert in der Grundgesamtheit mit einer vorgegebenen Wahrscheinlichkeit liegt. Solche Konfidenzintervalle gibt es für Mittelwerte, Standardabweichungen und prozentuale Häufigkeiten.

11.2 Überprüfung von Hypothesen

Liegen zwei oder mehrere Stichproben vor, so können deren Kennwerte dahingehend überprüft werden, ob sie derselben Grundgesamtheit angehören oder nicht. Man spricht in diesem Zusammenhang von *Prüfstatistik*.

Typisches Beispiel: Eine Firma testet ein Medikamt A und ein weiteres Medikament B, indem an Probanden eine Messung eines bestimmten Kennwerts nach Einnahme des jeweiligen Medikaments durchgeführt wird. Nun können bezogen auf diesen Meßwert Daten erhoben werden, klassischerweise die Mittelwerte $\overline{x}_A$, $\overline{x}_B$ und die Standardabweichungen s_A, s_B.

Möchte man nun wissen, ob die erhaltenen Wertunterschiede (zum Beispiel des Mittelwerts) zufällig sind oder nicht (man spricht in diesem Zusammen-

[3] Man spricht von Stichproben.

hang von *Signifikanz* - siehe Kapitel 11.3), werden Hypothesen aufgestellt, die anschließend mit dem passenden Verfahren überprüft werden.

Wie die Vorgangsweise beim Aufstellen von Hypothesen allgemein aussieht, wird im nächsten Kapitel besprochen.

11.3 Hypothesen und Signifikanz

Bleiben wir beim vorangegangenen Beispiel und betrachten weiterhin die Unterschiede der beiden Mittelwerte aus den beiden Stichproben[4]. Es lassen sich zwei Hypothesen aufstellen:

- Hypothese 0 (H0): Der Mittelwertsunterschied ist zufällig zustande gekommen.
- Hypothese 1 (H1): Der Mittelwertsunterschied ist nicht zufällig zustande gekommen.

Die Hypothese H0 nennt man die *Nullhypothese*, die Hypothese H1 die *Alternativhypothese* oder *Arbeitshypothese*. Die Prüfstatistik entscheidet nun, ob die formulierte Nullhypothese beibehalten wird oder zugunsten der Alternativhypothese verworfen werden muß.

Statistische Tests werden üblicherweise mit Hilfe einer *Prüfgröße* durchgeführt. Das ist eine Vorschrift, nach der aus einer beziehungsweise häufig aus zwei gegebenen Stichproben eine Zahl errechnet wird, aufgrund deren Kenntnis man sich für oder gegen die Nullhypothese entscheidet. Aus einer Tabelle kann dann zu der berechneten Prüfgröße ein p-Wert entnommen werden. Wir unterscheiden drei Signifikanzstufen:

$p \leq 0,05$	signifikant
$p \leq 0,01$	sehr signifikant
$p \leq 0,001$	höchst signifikant

Für p-Werte $\leq 0,1$[5] spricht man manchmal auch von einer sogenannten „Tendenz zur Signifikanz“. Das ist deshalb der Fall, weil es bei solchen Werten unrichtig wäre zu behaupten, es läge kein signifikantes Ergebnis vor.

[4] Anmerkung: Die Probanden stellen deshalb eine Stichprobe dar, weil klarerweise nicht alle Personen, die die jeweiligen Medikamente einnehmen, getestet werden können. Diese Gesamtgruppe wäre in diesem Fall die Grundgesamtheit.

[5] Bei solchen Fällen nähert sich p bereits der Signifikanzgrenze.

Man spricht also von *signifikanten* Unterschieden, wenn das Ergebnis p eines passenden Tests kleiner oder gleich 0,05 beträgt. Nach diesem Grundprinzip werden Schlüsse gezogen, die in der Praxis ihre Anwendung finden. Die Frage der Signifikanz ist das zentrale Thema der analytischen Statistik. Das Ergebnis eines statistischen Tests ist also eine nach einer bestimmten Vorschrift berechnete Prüfgröße. Diese wird mit einem tabellierten Schwellenwert verglichen, an dessen Stelle in der Tabelle der zugehörige p-Wert angegeben ist. Dieser entscheidet darüber, ob ein Testergebnis signifikant ist oder nicht.

11.4 Prüfverteilungen

11.4.1 Standardnormalverteilung

Eine normalverteilte Prüfgröße, stets z genannt (sieht Kapitel 10.2.1), wird bei diversen Tests berechnet (vgl. Kapitel 14).

Ihre Verteilungsfunktion ist

$$\Phi(z) = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_{-\infty}^{z} e^{-\frac{1}{2} \cdot t^2} \mathrm{d}t$$

Deren Werte sind für $z > 0$ tabelliert. Es besteht die Symmetriebeziehung

$$\Phi(-z) = 1 - \Phi(z)$$

11.4.2 χ^2-Verteilung

Gegeben sei eine standardnormalverteilte Zufallsvariable z. Werden n solcher standardnormalverteilten, voneinander unabhängigen Zufallsvariablen $z_1 \ldots z_n$ quadriert und addiert, erhält man eine χ^2_n-verteilte Zufallsvariable:

$$\chi^2_n = z_1^2 + z_2^2 + \ldots + z_n^2 = \sum_{i=1}^{n} z_i^2$$

χ^2-Verteilungen unterscheiden sich somit darin, daß unterschiedliche Anzahlen von z^2-Variablen aufsummiert werden. In Abhängigkeit von der Anzahl dieser Variablen spricht man von unterschiedlichen *Freiheitsgraden df* („degrees of freedom“) Die Anzahl der Summanden n entspricht hier der

Anzahl der Freiheitsgrade df. Die Kurven zur χ^2-Verteilung sind linksgipflig.

Die χ^2-Verteilung besitzt für $x > 0$ die Dichtefunktion:

$$f_{\chi^2_n}(x) = \frac{1}{2^{\frac{n}{2}} \cdot \Gamma\left(\frac{n}{2}\right)} \cdot x^{\frac{n-2}{2}} \cdot e^{-\frac{x}{2}}$$

Die Fraktile der zugehörigen Verteilungsfunktion sind für einige Werte von Freiheitsgraden df und einige p-Werte tabelliert.

11.4.3 t-Verteilung

Sei X standardnormalverteilt und Y χ^2-verteilt mit n Freiheitsgraden (X unabhängig von Y). Dann ist die t-Verteilung die Verteilung der Zufallsvariablen

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

Die Kurven dieser Verteilung liegen symmetrisch um Null. Für große n nähert sich die Verteilungsfunktion einer t_n-verteilten Zufallsvariablen immer mehr der Standardnormalverteilung. Sie besitzt die Dichtefunktion

$$f_{t_n}(z) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \cdot \sqrt{n \cdot \pi}} \cdot \left(1 + \frac{z^2}{n}\right)^{-\frac{n+1}{2}}$$

Die Fraktile der zugehörigen Verteilungsfunktion sind für einige Werte von Freiheitsgraden df (bzw. n) und einige p-Werte tabelliert.

11.4.4 F-Verteilung

Gegeben sei eine χ^2-Verteilung mit $df1$ und eine weitere, unabhängige χ^2-Verteilung mit $df2$. Der Quotient von zwei zufällig aus diesen beiden Verteilungen entnommenen χ^2-Werten, multipliziert mit dem Kehrwert des Quotienten ihrer Freiheitsgrade wird als F-Wert bezeichnet.

$$F_{df1,df2} = \frac{\chi^2_{df1}}{\chi^2_{df2}} \cdot \frac{df2}{df1}$$

Die Verteilung dieser Zufallsvariablen heißt F-Verteilung. Verschiedene F-Verteilungen unterscheiden sich durch die Anzahl der Freiheitsgrade $df1$ (Zählerfreiheitsgrade) und $df2$ (Nennerfreiheitsgrade). Die Kurven zur F-Verteilung sind linksgipflig. Ihre Dichtefunktion lautet für $x > 0$:

$$f_{F_{df1,df2}}(x) = \frac{\Gamma\left(\frac{df1+df2}{2}\right)}{\Gamma\left(\frac{df1}{2}\right)\cdot\Gamma\left(\frac{df2}{2}\right)} \cdot \left(\frac{df1}{df2}\right)^{\frac{df1}{2}} \cdot x^{\frac{df1-2}{2}} \cdot \left(1 + \frac{df1}{df2}\cdot x\right)^{-\frac{df1+df2}{2}}$$

Für einige Anzahlen von Zähler- sowie Nennerfreiheitsgraden sind die Werte der zugehörigen Verteilungsfunktion zu einigen p-Werten tabelliert.

11.5 Fehler erster und zweiter Art

Stellt man eine Null- und eine Alternativhypothese auf, so können zwei Arten von Fehlern vorkommen:

- Die Nullhypothese wird verworfen, obwohl sie richtig ist.
- Die Nullhypothese wird beibehalten, obwohl sie falsch ist.

An erster Stelle genannter Fall heißt *Fehler erster Art* oder *α-Fehler*. Die Wahrscheinlichkeit, einen solchen Fehler zu begehen, ist gleich der Irrtumswahrscheinlichkeit α. Der andere Fall heißt *Fehler zweiter Art* oder *β-Fehler*. Tabelle 9 zeigt nochmals die möglichen Fälle.

	H0 wahr	**H0 falsch**
H0 abgelehnt	Fehler 1. Art	richtige Entscheidung
H0 beibehalten	richtige Entscheidung	Fehler 2. Art

Tabelle 9: Fehler erster und zweiter Art

Fehler erster Art werden auch als *Konsumentenrisiko* bezeichnet, Fehler zweiter Art als *Produzentenrisiko*. Erinnern wir uns an das Beispiel der Testung der beiden Medikamente A und B: Unsere Nullhypothese besagt, daß unterschiedliche Ergebnisse zufällig zustande gekommen sind. Bei einem Fehler erster Art lehnen wir die Hypothese ab, obwohl sie richtig ist. Das bedeutet, wir sind aufgrund der Testung der Meinung, daß die Unterschiede nicht zufällig zustande gekommen sind und setzen beispielsweise ein neues Produkt am Markt ein (welches nur zufällig beim Test „gut abgeschnitten“ hat). Das entstehende Risiko trägt der *Konsument*.

Begehen wir einen Fehler zweiter Art, so sind die Unterschiede der Testung nicht zufällig entstanden, wir sind allerdings weiterhin der Meinung, daß dies der Fall ist. Hätten wir als *Produzenten* das neue Produkt am Markt einsetzen wollen und tun dies nun eben nicht, so tragen wir als Produzenten das Risiko.

Teststärke Ist β die Wahrscheinlichkeit dafür, daß ein bestehender Unterschied nicht erkannt wird, so ist $1 - \beta$ die Wahrscheinlichkeit dafür, daß ein bestehender Unterschied aufgezeigt wird. Diesen Wert bezeichnet man als *Teststärke* (auch: Power, Güte, Trennschärfe).

12 Streubereiche und Konfidenzintervalle

Der Schluß von den Kennwerten einer Stichprobe auf die Parameter der zugehörigen Grundgesamtheit erfolgt bei intervallskalierten und normalverteilten Variablen über Streubereiche und Konfidenzintervalle.

12.1 Streubereiche

Mit der Angabe eines Streubereichs wird bei intervallskalierten und normalverteilten Variablen die Frage behandelt, wie viel Prozent der Werte in einem bestimmten Intervall liegen. Darunter versteht man stets ein um den Mittelwert symmetrisches Intervall. So kann nun einerseits die Intervallbreite vorgegeben sein und darauf geschlossen werden, wie viel Prozent der Werte innerhalb dieses vorgegebenen Intervalls liegen (Kapitel 12.1.1). Oder diese Prozentzahl ist bekannt, und man berechnet die Größe des zugehörigen Intervalls, innerhalb dessen die vorgegebene Prozentzahl der Werte liegen soll (Kapitel 12.1.2).

In den folgenden beiden Kapiteln werden als Grundlage der Berechnungen der Mittelwert (vgl. Kapitel 5.1)

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

und die Standardabweichung (vgl. Kapitel 6.2) verwendet.

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}}$$

12.1.1 Vorgabe der Intervallbreite

Einige relevante Intervallbreiten, die anschaulich sind und somit in der Praxis verwendet werden, sind in Tabelle 10 angeführt. Um zu betrachten, wieviel Prozent der Werte im Intervall von $\overline{x}-s$ bis $\overline{x}+s$ liegen, werden wir die z-Tabelle heranziehen. Deshalb wird zunächst eine z-Standardisierung ($z = \frac{x-\overline{x}}{s}$) durchgeführt und zwar an den Stellen, die die Grenzen des Intervalls darstellen. Somit ergibt sich für die untere Intervallgrenze $\overline{x} - s$

$$\frac{\overline{x} - s - \overline{x}}{s} = -1$$

und für die obere Intervallgrenze $\overline{x} + s$

$$\frac{\overline{x} + s - \overline{x}}{s} = 1$$

Betrachtet man nun das Flächenstück unter der Standardnormalverteilungskurve zwischen den Werten $z = -1$ und $z = 1$, so erhalten wir

$$\Phi(1) - \Phi(-1) = 0,84134 - 0,15866 = 0,68286$$

In diesem Intervall liegen also 68,3 % der Werte.

Analog dazu ergeben sich für die Intervalle $[\overline{x}-2\cdot s; \quad \overline{x}+2\cdot s]$ beziehungsweise $[\overline{x} - 3 \cdot s; \quad \overline{x} + 3 \cdot s]$ die z-Werte (-2,2) beziehungsweise (-3,3) und man erhält die dazugehörigen Prozentzahlen durch Berechnung der Differenzen aus der z-Tabelle (vgl. Tabelle 10).

Streubereich	**Prozent der Werte**
$\overline{x} \pm s$	68,3%
$\overline{x} \pm 2 \cdot s$	95,5%
$\overline{x} \pm 3 \cdot s$	99,7%

Tabelle 10: Relevante Streubereiche bei Vorgabe der Intervallbreite

12.1.2 Vorgabe der Prozentzahl

Dieser Fall ist der gebräuchlichere. Man gibt eine Prozentzahl vor und untersucht das zugehörige Intervall. Dabei wird abermals die z-Tabelle herangezogen. Möchte man beispielsweise wissen, in welchem Intervall 95 % der Werte liegen, so sucht man jene Differenz $\Phi(z) - \Phi(-z)$, welche als Ergebnis den Wert 0,95 liefert. Dies ist bei $z = 1,96$ der Fall. Dies bedeutet nun, daß im Intervall $[\overline{x} - 1,96 \cdot s; \quad \overline{x} + 1,96 \cdot s]$ 95 % der Werte liegen. Tabelle 11 gibt relevante Prozentwerte an.

Prozent der Werte	**Intervall**
95%	$[\overline{x} - 1,96 \cdot s; \quad \overline{x} + 1,96 \cdot s]$
99%	$[\overline{x} - 2,58 \cdot s; \quad \overline{x} + 2,58 \cdot s]$

Tabelle 11: Relevante Streubereiche bei Vorgabe einer bestimmten Prozentzahl

12.2 Konfidenzintervalle

Während Streubereiche in Bezug auf einzelne Meßwerte angegeben werden, behandeln Konfidenzintervalle das Problem, wie man vom Kennwert einer Stichprobe auf den entsprechenden Parameter der Grundgesamtheit schließen kann. Die Genauigkeit für das jeweilige Konfidenzintervall wird vorgegeben.

12.2.1 Konfidenzintervall für den Mittelwert

Um zu einem gegebenen Mittelwert $\overline{x}$ einer Stichprobe mit Hilfe der zugehörigen Standardabweichung s ein Konfidenzintervall bestimmen zu können, wird zunächst die Standardabweichung in den Standardfehler umgerechnet (vgl. Definition 6.5 auf Seite 31).

$$s_m = \frac{s}{\sqrt{n}}$$

Das Konfidenzintervall für den Mittelwert μ wird dann mit Hilfe der t-Verteilung ermittelt. Will man beispielsweise ein 95%-Konfidenzintervall bestimmen, so ist aus der t-Tabelle zuerst der zu $p = 0,05$ und $df = n - 1$ Freiheitsgraden gehörige Tabellenwert $t_{p;n-1}$ zu bestimmen. Man wählt deshalb den zu $p = 0,05$ gehörigen Wert, da die Irrtumswahrscheinlichkeit, die zur *Konfidenzzahl* 95 % gehört, 100% − 95% = 5% beträgt. Schließlich verwendet man folgende Intervallformel:

$$\overline{x} - t_{p;n-1} \cdot s_m < \quad \mu \quad < \overline{x} + t_{p;n-1} \cdot s_m$$

Für große Fallzahlen nähert sich die t-Verteilung der Standardnormalverteilung. Somit nähert sich auch der in die Intervallformel einzusetzende t-Wert bei einem 95%-Konfidenzintervall dem Wert 1,96.

12.2.2 Konfidenzintervall für die Standardabweichung

Möchte man aus dem bekannten Wert s der Stichprobe auf die Standardabweichung σ der Grundgesamtheit schließen, so verwendet man in diesem Fall die F-Verteilung. Das 95%-Konfidenzintervall sieht, indem die Tabellenwerte bei $p = 0,05$ und $(n-1; \infty)$ beziehungsweise $(\infty; n-1)$ Freiheitsgraden verwendet werden, wie folgt aus:

$$\frac{s}{\sqrt{F_{p;(n-1;\infty)}}} < \quad \sigma \quad < s \cdot \sqrt{F_{p;(\infty;n-1)}}$$

Bei großen Werten von $df1$ und $df2$ werden die Werte aus der F-Tabelle interpoliert. Für $df1 = \infty$ kann der Wert $df1 = 1000$ gewählt werden.

12.2.3 Konfidenzintervall für prozentuale Häufigkeiten

Auch für prozentuale Häufigkeiten (Prozentwerte) P lassen sich Konfidenzintervalle bestimmen, wenn die zugrunde liegende Fallzahl n bekannt ist.

Der Prozentwert der Grundgesamtheit wird mit π bezeichnet. Für das gesuchte Konfidenzintervall gilt

$$P - \sigma \cdot z < \quad \pi \quad < P + \sigma \cdot z$$

mit

$$\sigma = \sqrt{\frac{P \cdot (100 - P)}{n}}$$

z ist der zur gegebenen Konfidenzzahl gehörige z-Wert.

13 Überprüfung auf Normalverteilung

Die Normalverteilung spielt in der Statistik eine entscheidende Rolle. Denn je nachdem, ob eine Normalverteilung vorliegt oder nicht, sind gegebenenfalls verschiedene analytische Tests durchzuführen. Wenn ein statistischer Test eine Normalverteilung der Werte voraussetzt, so muß zuerst überprüft werden, ob eine Normalverteilung der Werte vorliegt.

13.1 Der Chiquadrat-Test

Bei diesem Test werden die Werte der zu überprüfenden Variablen in Klassen eingeteilt. Die beobachteten Klassenhäufigkeiten werden schließlich mit jenen verglichen, die unter Normalverteilung zu erwarten sind. Man spricht auch des öfteren vom *Chiquadrat-Anpassungstest*.

Beispiel Von $n = 196$ Patienten einer Klinik soll überprüft werden, ob deren Körpergewichtsangaben normalverteilt sind. Tabelle 12 gibt die beim Chiquadrat-Test auszuführenden Schritte wider.

Klasse	**Klassenende**	f_0	z	$\Phi(z)$	FD	f_e	$\frac{(f_0-f_e)^2}{f_e}$
≤ 47	47,5	12	-1,64	0,051	0,051	10,0	0,402
48-52	52,5	12	-1,25	0,106	0,055	10,8	0,138
53-57	57,5	16	-0,87	0,192	0,086	16,9	0,043
58-62	62,5	29	-0,48	0,316	0,124	24,3	0,907
63-67	67,5	23	-0,10	0,460	0,144	28,2	0,967
68-72	72,5	28	0,28	0,610	0,150	29,4	0,067
73-77	77,5	28	0,67	0,749	0,139	27,2	0,021
78-82	82,5	19	1,05	0,853	0,104	20,4	0,094
83-87	87,5	13	1,44	0,925	0,072	14,1	0,088
88-92	92,5	6	1,82	0,966	0,041	8,0	0,516
93-97	97,5	7	2,21	0,986	0,020	3,9	2,420
> 97		3		1,000	0,014	2,7	0,033
Summe		196			1,000	196,0	5,696

Tabelle 12: Rechenschritte beim Chiquadrat-Test

Die erste Spalte gibt die Klasseneinteilung an. Hier wurde die Breite 5 gewählt. In der zweiten Spalte sind die Klassenenden angeführt. In der dritten Spalte stehen die *beobachteten Häufigkeiten* f_0. In der vierten Spalte finden wir jenen z-Wert, der zum jeweiligen Klassenende gehört. Die z-Standardisierung wurde bereits durchgeführt.

$$z = \frac{x - \overline{x}}{s}$$

Für x muß hier der zum jeweiligen Klassenende gehörige Wert eingesetzt werden.

In Spalte 5 steht der zum z-Wert gehörige $\Phi(z)$-Wert aus der z-Tabelle. *FD* in der sechsten Spalte steht für *Flächendifferenz*. Es wird die jeweilige Flächendifferenz zur vorangegangenen Zeile angegeben. In der siebenten Spalte steht die *erwartete Häufigkeit* f_e. Sie wird berechnet aus

$$f_e = FD \cdot n$$

Hierbei handelt es sich um einen theoretischen Wert. Deswegen wird er auch mit einer Nachkommastelle angegeben. In der letzten Spalte werden schließlich die *standardisierten quadrierten Residuen* berechnet. Summiert man diese auf, so erhält man den Wert rechts unten in der Tabelle, nämlich die Prüfgröße χ^2, welche χ^2-verteilt ist mit $df = k - 3$ Freiheitsgraden, wobei k die Anzahl der Klassen angibt:

$$\chi^2 = \sum \frac{(f_0 - f_e)^2}{f_e}$$

Im vorliegenden Beispiel ergibt sich

$$\chi^2 = 5,696 \quad df = 12 - 3 = 9$$

Wie die χ^2-Tabelle ausweist, ist dies bei 9 Freiheitsgraden eine nicht signifikante Prüfgröße. Dies bedeutet, daß sich die gegebene Verteilung nicht signifikant von einer Normalverteilung unterscheidet. Die Variable „Körpergewicht“ wird also vorerst als normalverteilt angesehen.

Der Nachteil des Chiquadrat-Tests ist die Tatsache, daß alle Werte zuerst in Klassen eingeteilt werden müssen. Deshalb verwendet man diesen Test vor allem für große Fallzahlen.

13.2 Der Kolmogorov-Smirnov-Test

Der Kolmogorov-Smirnov-Test kommt bei kleineren Fallzahlen zum Einsatz.

Beispiel Nehmen wir an, es liegen $n = 8$ Zeitangaben vor, die auf Normalverteilung überprüft werden sollen.

200, 198, 390, 215, 171, 160, 150, 224

Anmerkung Bei derart kleinen Fallzahlen ist eine signifikante Abweichung von der Normalverteilung recht selten. Selbst der in Tabelle 13 vorkommende „Ausreißerwert“ 390 fällt hier nicht ins Gewicht, soviel sei vorweggenommen. Unsere Nullhypothese lautet also: „Die Verteilung der Werte unterscheidet sich nicht signifikant von einer Normalverteilung.“

Die Vorgehensweise beim Kolmogorov-Smirnov-Test ist die folgende (sie ist in Tabelle 13 dargestellt): Die Werte werden in eine aufsteigende Reihenfolge gebracht. Anschließend wird die z-Standardisierung durchgeführt. Danach werden die jeweiligen Werte der Verteilungsfunktion Φ der Standardnormalverteilung eingetragen. Diese Flächenstücke unter der Kurve von Φ sollten bei idealer Normalverteilung gleiche Abstände haben, wie sie in der nächsten Spalte eingetragen werden ($f = \frac{i}{n}$).

i	x_i	z_i	$\Phi(z_i)$	$f_i = \frac{i}{n}$	$\lvert\Phi(z_i) - f_i\rvert$	$\lvert\Phi(z_i) - f_{i-1}\rvert$
1	150	-0,84	0,200	0,125	0,075	0,200
2	160	-0,70	0,242	0,250	0,008	0,117
3	171	-0,56	0,288	0,375	0,087	0,038
4	198	-0,20	0,421	0,500	0,079	0,046
5	200	-0,18	0,429	0,625	0,196	0,071
6	215	0,02	0,508	0,750	0,242	0,117
7	224	0,14	0,556	0,875	**0,319**	0,194
8	390	2,32	0,990	1,000	0,010	0,115

Tabelle 13: Rechenschritte beim Kolmogorov-Smirnov-Test

Das Maximum der Differenzen in der letzten Spalte ist die Prüfgröße bei diesem Test. Dabei ist zu beachten, daß die $\Phi(z_i)$-Werte sowohl mit den f_i- als auch mit den f_{i-1}-Werten verglichen werden müssen.

Sind allgemein Zufallsgrößen $X_1, \ldots, X_n$ unabhängig verteilt mit der Verteilungsfunktion F_0 und ist F_n die empirische Verteilungsfunktion (der n

beobachteten Werte), so hat folgende Zufallsvariable eine von F_0 unabhängige Verteilung:

$$K_n = \sup|F_0(x) - F_n(x)|$$

Die Fraktile $k_{n,p}$ dieser Verteilung sind bis $n = 40$ für bestimmte p-Werte tabelliert.

Damit in unserem Fall eine signifikante Abweichung von der Normalverteilung vorliegt, muß die besprochene Prüfgröße den bei der betreffenden Fallzahl n tabellierten Grenzwert überschreiten. In diesem Fall wäre dies der Wert 0,319, welcher kleiner ist als der tabellierte Wert 0,454. Somit kann aufgrund dieses Tests kein signifikanter Unterschied zu einer Normalverteilung nachgewiesen werden.

14 Test auf signifikante Unterschiede

Am häufigsten hat man es in der analytischen Statistik wohl mit Tests zu tun, die Beziehungen zwischen zwei Variablen untersuchen. Es kommen nun unterschiedliche Testverfahren zur Anwendung. Welches in einem bestimmten Fall das passende ist, hängt davon ab, welchem Skalenniveau die Variablen angehören beziehungsweise auch, ob Normalverteilung der Variablen vorliegt oder nicht.

Prinzipiell gibt es meist zwei verschiedene Ansätze, wie eine Beziehung zwischen zwei Variablen aufgedeckt werden kann.

- Man verwendet eine der beiden Variablen als *Gruppierungsvariable*. Die entstehenden Gruppen werden dann bezüglich des Mittelwerts oder des Medians der anderen Variablen auf signifikante Unterschiede getestet. Die Gruppierungsvariable selbst muß dabei sinnvollerweise nominalskaliert oder ordinalskaliert mit recht wenigen Kategorien sein.
- Man verwendet den *Korrelationskoeffiezient*, um die Stärke des Zusammenhangs der beiden Variablen aufzuzeigen

Möchte man nun Stichproben hinsichtlich Mittelwert oder Median, also allgemein hinsichtlich ihrer zentralen Tendenzen vergleichen, so hat man drei relevante Unterscheidungsmöglichkeiten zu beachten.

- Hat man es mit *unabhängigen* oder mit *abhängigen* Stichproben zu tun?
- Vergleicht man *zwei* Stichproben miteinander, oder vergleicht man *mehrere?*
- Sind die untersuchten Werte *intervallskaliert und normalverteilt*, oder sind sie *ordinalskaliert* oder *nicht normalverteilt intervallskaliert?*

Abhängigkeit von zwei (oder mehreren) Stichproben bedeutet, daß jeweils ein Wertepaar aus beiden Stichproben sinnvoll und eindeutig einander zugeordnet werden kann. Der hier typische Fall ist der, daß ein bestimmtes Merkmal bei einem Probanden unter zwei verschiedenen Bedingungen getestet wird. Das ist insbesondere der Fall, wenn eine Variable zu zwei oder mehreren Zeitpunkten gemessen wird. Unabhängige Stichproben liegen dann vor, wenn diese Stichproben unterschiedliche Probanden (allgemein: Fälle) enthalten. Liegt dieser Fall vor, so müssen die Fallzahlen der jeweiligen Stichproben nicht gleich sein.

Die zweite Unterscheidungsmöglichkeit bei der Anwendung eines Signifikanztests ist die Anzahl der beteiligten Stichproben. Die Schwierigkeit beim Vergleich mehrerer Stichproben besteht nämlich darin, genau sagen zu können, zu welchen beiden Stichproben ein eventuell erwiesener signifikanter Unterschied gehört. Im Fall von nur zwei beteiligten Stichproben ist klar, daß nur zwischen eben diesen beiden der Unterschied aufgezeigt wird oder nicht.

Schließlich ist es wichtig zu wissen, ob die verwendeten Werte intervallskaliert und normalverteilt sind oder nicht. Sie können auch entweder ordinalskaliert oder intervallskaliert, aber nicht normalverteilt sein. In letzterem Fall verwendet man sogenannte *parameterfreie* Tests. Die dort zur Anwendung kommenden Formeln bauen nicht auf den Originalwerten auf, sondern auf Rangplätzen, die den Werten zugeordnet werden. Hat man in einigen Fällen normalverteilte und in einigen anderen Fällen nicht normalverteilte Werte, wäre es möglich, ausschließlich parameterfreie Tests zu verwenden. Diese können teilweise eine Effizienz von bis zu 95 % der entsprechenden parametrischen Tests aufweisen. Oftmals werden bereits grundsätzlich parameterfreie Tests verwendet, da diese an keine Voraussetzungen gebunden sind und normalverteilte Werte in der Praxis eher die Ausnahme darstellen.

14.1 Der t-Test nach Student

Der t-Test nach Student dient zum Vergleich zweier unabhängiger Stichproben hinsichtlich ihrer Mittelwerte, wobei die Werte der beiden Stichproben normalverteilt sein müssen. Je nachdem, ob sich die Varianzen in beiden Stichproben signifikant unterscheiden oder nicht, gibt es zwei verschiedene Formeln für eine t-verteilte Prüfgröße t. In diese Formeln gehen die jeweiligen Mittelwerte $\overline{x}_1$ und $\overline{x}_2$, die beiden Standardabweichungen s_1 und s_2, sowie die beiden Fallzahlen n_1 und n_2 ein.

Zuerst überprüft man also, ob Varianzenhomogenität (die Varianzen unterscheiden sich nicht signifikant) gewährleistet ist oder ob Varianzeninhomogenität (die Varianzen unterscheiden sich signifikant) vorliegt. Die zu berechnende Prüfgröße hierfür lautet

$$F = \frac{s_{major}^2}{s_{minor}^2}$$

Dabei steht s_{major} für die größere der beien Standardabweichungen und s_{minor} für die kleinere. Die Prüfgröße F ist F-verteilt mit

$$df = (n_{major} - 1, \quad n_{minor} - 1)$$

Freiheitsgraden. Varianzenheterogenität wird bei einer Signifikanz auf der Stufe $p < 0,05$ angenommen.

Falls nun Varianzenhomogenität vorliegt, so gilt:

$$t = \frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \qquad df = n_1 + n_2 - 2$$

Im Falle der Varianzenheterogenität gilt:

$$t = \frac{|\overline{x}_1 - \overline{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad df = \frac{n_1 + n_2 - 2}{2}$$

14.1.1 Rechenschritte des t-Tests anhand des Beispiels „Gedächtnistest"

Folgendes Beispiel soll die Rechenschritte beim t-Test erläutern: Zwanzig Männer ($n_1 = 20$) und elf Frauen ($n_2 = 11$) nahmen an einem Gedächtnistest teil, bei dem möglichst viele vorgegebene Wörter gemerkt werden sollten. Folgende Leistungen wurden erzielt:

Männer: 22, 27, 28, 30, 23, 25, 26, 29, 32, 25, 23, 29, 29, 28, 30, 21, 26, 16, 23, 25

Frauen: 35, 26, 34, 24, 27, 25, 28, 24, 25, 30, 34

Auf die Überprüfung der Werte in diesen Stichproben auf Normalverteilung soll hier verzichtet werden (vgl. dazu Kapitel 13). Die Berechnung der Mittelwerte und der Standardabweichungen ergibt:

$$\overline{x}_1 = 25{,}9 \quad \overline{x}_2 = 28{,}4 \quad \overline{s}_1 = 3{,}80 \quad \overline{s}_2 = 4{,}23$$

Es folgt nun zunächst der F-Test, um zu überprüfen, ob Varianzenhomogenität vorliegt:

$$F = \frac{4,23^2}{3,8^2} = 1,24$$

Nach der F-Tabelle ist dies bei (19,10) Freiheitsgraden ein nicht signifikanter Wert. Somit ist Varianzenhomogenität gegeben.

Nun wird die Prüfgröße t berechnet:

$$t = \frac{|25,9 - 28,4|}{\sqrt{\frac{19 \cdot 3,80^2 + 10 \cdot 4,23^2}{29}}} \cdot \sqrt{\frac{20 \cdot 11}{10 + 11}} = 1,685$$

Nach der t-Tabelle ist dies bei $df = 20 + 11 - 2 = 29$ Freiheitsgraden ein nicht signifikanter Wert ($p > 0,05$). Somit kann kein Unterschied zwischen den beiden Geschlechtern bezüglich der in diesem Gedächtnistest erzielten Leistung nachgewiesen werden.

14.2 Der t-Test für abhängige Stichproben

Dieser Test dient zum Vergleich zweier abhängiger Stichproben hinsichtlich ihrer Mittelwerte, wobei die Differenzen zusammengehöriger Meßwertepaare aus einer normalverteilten Grundgesamtheit stammen müssen.

Anhand des folgenden Beispiels soll folgender Test vorgestellt werden: Die Befindlichkeiten von zehn Patienten einer psychiatrischen Klinik wurden auf einer Skala gemessen, welche Werte zwischen 0 und 56 annehmen kann, wobei hohe Werte für schlechte Befindlichkeiten stehen. Eine Bewegungstherapie sollte die Befindlichkeit verbessern. Die Befindlichkeitswerte vor und nach der Bewegungstherapie, deren Differenzen (d) und die Quadrate dieser Differenzen sind in Tabelle 14 enthalten.

Vp	**vor**	**nach**	d	d^2
1	30	20	10	100
2	22	24	-2	4
3	38	31	7	49
4	34	28	6	36
5	25	20	5	25
6	28	28	0	0
7	33	27	6	36
8	21	24	-3	9
9	29	21	8	64
10	31	25	6	36
Σ	291	248	43	359

Tabelle 14: Befindlichkeitswerte von Patienten

Die beiden Stichproben, die durch die Befindlichkeitswerte vor und nach der Bewegungstherapie gebildet werden, sind *voneinander abhängig*, da zu jedem Probanden genau ein Wertepaar (vor, nach) existiert. Auf die Prüfung der Differenzen auf Normalverteilung wird hier verzichtet (vgl. dazu Kapitel 13).

Für die Mittelwerte und deren Differenz[6] erhält man:

$$\overline{x}_1 = \frac{291}{10} = 29,1 \quad \overline{x}_2 = \frac{248}{10} = 24,8 \quad \overline{d} = \frac{43}{10} = 4,3$$

Im Mittel ist also der Wert der Befindlichkeit um 4,3 gesunken, was in diesem Fall einer Verbesserung der Befindlichkeit entspricht. Um diesen Unterschied der beiden Mittelwerte auf Signifikanz zu prüfen, berechnet man neben dem Mittelwert der Differenzen

$$\overline{d} = \frac{\sum\limits_{i=1}^{n} d_i}{n}$$

noch deren Standardabweichung

$$s = \sqrt{\frac{\sum\limits_{i=1}^{n} d_i^2 - \frac{\left(\sum\limits_{i=1}^{n} d_i\right)^2}{n}}{n-1}}$$

Die Prüfgröße t berechnet sich zu

$$t = \frac{\overline{d} \cdot \sqrt{n}}{s}$$

Dieser Wert ist t-verteilt mit $df = n - 1$ Freiheitsgraden. Wir erhalten für $s = 4,398$, für t also:

$$t = \frac{4,3 \cdot \sqrt{10}}{4,398} = 3,092$$

Bei $df = 10-1 = 9$ Freiheitsgraden ist dieser Wert signifikant, da er den zugehörigen Wert in der t-Tabelle übersteigt (bei einer Irrtumswahrscheinlichkeit von 5 %). Somit ist der Unterschied der mittleren Befindlichkeitswerte vor und nach der Bewegungstherapie als signifikant nachgewiesen.

[6] Bis auf Rundungsfehler ist natürlich der Mittelwert der Differenzen gleich der Differenz der Mittelwerte:

$$\overline{d} = \overline{x}_1 - \overline{x}_2$$

14.3 Einfaktorielle Varianzanalyse

Die einfaktorielle Varianzanalyse dient zum Vergleich von mehr als zwei unabhängigen Stichproben hinsichtlich ihrer Mittelwerte, wobei die Werte der einzelnen Stichproben normalverteilt sein müssen. Eine weitere Voraussetzung ist die Varianzenhomogenität über die Stichproben hinweg.

Die Bezeichnung *Varianzanalyse* wird dabei teilweise als irreführend empfunden, da man meinen könnte, daß die Varianzen auf signifikante Unterschiede getestet werden. Der Name dieses Verfahrens kommt allerdings daher, daß dessen Grundlage eine Zerlegung der Gesamtvarianz ist.

14.3.1 Überprüfung auf Varianzenhomogenität

Eine der beiden Voraussetzungen zur Durchführung der Varianzanalyse ist die Homogenität der Varianzen

$$s_1^2, \quad s_2^2, \quad \ldots \quad s_k^2$$

Zur Überprüfung auf Varianzenhomogenität sind diverse Tests bekannt. Hier werden drei Tests vorgestellt.

14.3.1.1 Der Bartlett-Test Dieser Test ist wohl der am meisten angewandte. Es werden folgende Größen berechnet:

$$\begin{aligned}
n &= \sum_{i=1}^{k} n_i \\
s^2 &= \frac{1}{n-k} \cdot \sum_{i=1}^{k} \left((n_i - 1) \cdot s_i^2 \right) \\
c &= 1 + \frac{1}{3 \cdot (k-1)} \cdot \left(\sum_{i=1}^{k} \frac{1}{n_i - 1} - \frac{1}{n-k} \right) \\
\chi^2 &= \frac{1}{c} \cdot \left((n-k) \cdot \ln\left(s^2\right) - \sum_{i=1}^{k} (n_i - 1) \cdot \ln\left(s_i^2\right) \right)
\end{aligned}$$

Die Prüfgröße χ^2 ist χ^2-verteilt mit $df = k - 1$ Freiheitsgraden.

Eine mögliche Zusammenstellung der Rechenschritte zum Bartlett-Test findet sich sich Tabelle 15.

Somit ergibt sich für die nötigen Kennwerte

i	s_i^2	n_i	$\frac{1}{n_i-1}$	$(n_i-1)\cdot s_i^2$	$\ln s_i^2$	$(n_i-1)\cdot\ln s_i^2$
1	3,940	14	0,077	51,22	1,371	17,826
2	4,787	8	0,143	33,51	1,566	10,962
3	1,476	7	0,167	8,86	0,389	2,337

Tabelle 15: Rechenschritte zum Bartlett-Test

$$\begin{aligned} s^2 &= \frac{1}{26}\cdot 93,59 = 3,6 \\ c &= 1+\frac{1}{6}\cdot\left(0,387-\frac{1}{26}\right) = 1,058 \\ \chi^2 &= \frac{1}{1,058}\cdot(26\cdot\ln(3,6)-31,125) = 3,584 \end{aligned}$$

Dies ist, wie die χ^2-Tabelle ausweist, bei $df = 2$ Freiheitsgraden ein nicht signifikanter Wert. Die gegebenen Varianzen können also als homogen betrachtet werden.

14.3.1.2 Der Levene-Test Bei diesem Test werden die ursprünglichen Werte x_{ij} ersetzt durch

$$|x_{ij}-\overline{x}_i|$$

Die so erhaltenen Werte werden dann auf die vorhin beschriebene Weise (vgl. Seite 62) einer Varianzanalyse unterzogen. Der sich ergebende F-Wert gilt als *Levene-Statistik*, die mit

$$df = (k-1, \quad n-k)$$

Freiheitsgraden F-verteilt ist.

Bei den gegebenen Werten ergibt sich mit $F = 1,211$ ein mit (2, 26) Freiheitsgraden nicht signifikanter Wert.

Als Variante dieses Tests können nicht die Mittelwerte, sondern die Mediane der jeweiligen Gruppen verwendet werden.

14.3.1.3 Der Hartley-Test Der Hartley-Test stellt bei weitem den kürzesten Test dar. Hier wird die Testgröße F aus der kleinsten und der größten Varianz gebildet:

$$F = \frac{s_{max}^2}{s_{min}^2}$$

Dieser F-Wert ist F-verteilt mit $df = (k, \quad n_{max} - 1)$ Freiheitsgraden, wobei n_{max} die größte aller vorkommenden Fallzahlen ist.

Im gegebenen Beispiel ergibt sich

$$F = \frac{2,188^2}{1,215^2} = 3,24$$

Bei $df = (3,8)$ ist dies ein nicht signifikanter Wert. Während der Vorteil des Hartley-Tests in dessen Kürze und Einfachheit liegt, findet sich dort verständlicherweise auch sein Nachteil. Da nur die beiden extremen Varianzen in die Berechnung eingehen, erscheint dieser Test etwas undifferenziert.

14.3.2 Das Prinzip der Varianzanalyse

Ist k die Anzahl der Stichproben, die in diesem Zusammenhang auch *Gruppen* genannt werden, n die Gesamtzahl der Werte und x_{ij} der j-te Wert in der i-ten Stichprobe, so beträgt die Gesamtvarianz

$$\frac{1}{n-1} \cdot \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(x_{ij} - \overline{x}\right)^2$$

Hiebei ist $\overline{x}$ der Mittelwert über alle Werte und n_i der Umfang der i-ten Stichprobe.

Das Prinzip der Varianzanalyse ist eine Zerlegung dieser Gesamtvarianz in eine Varianz *innerhalb* der Gruppen und eine Varianz *zwischen* den Gruppen. Für die Summe der Abweichungsquadrate (SAQ) gilt die folgende Beziehung:

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(x_{ij} - \overline{x}\right)^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(x_{ij} - \overline{x}_i\right)^2 + \sum_{i=1}^{k} \left(n_i \cdot (\overline{x}_i - \overline{x})^2\right)$$

Die Summe auf der linken Seite dieser Gleichung ist die Aufsummierung der Abweichungen aller Werte vom Gesamtmittel $\overline{x}$. Sie wird daher SAQ_{gesamt} genannt. Als Beweis wird folgende Aufspaltung zu Grunde gelegt:

$$x_{ij} - \overline{x} = \left(x_{ij} - \overline{x}_i\right) + (\overline{x}_i - \overline{x})$$

Das erste Glied auf der rechten Seite der obigen Gleichung steht für die Abweichungen der Werte vom jeweiligen Gruppenmittel und wird daher als $SAQ_{innerhalb}$ bezeichnet. Das zweite Glied auf der rechten Seite steht für die Variabilität, die sich aus den Abweichungen der Gruppenmittel vom Gesamtmittel ergibt, und heißt deshalb $SAQ_{zwischen}$.

In Kurzschreibweise gilt also die Beziehung

$$SAQ_{gesamt} = SAQ_{innerhalb} + SAQ_{zwischen}$$

Diese Summen der Abweichungsquadrate werden durch ihre zugehörigen Anzahlen der Freiheitsgrade geteilt, woraus sich die *mittleren Quadrate* (MQ) ergeben. Liegen keine signifikanten Unterschiede vor, werden sich $MQ_{innerhalb}$ und $MQ_{zwischen}$ nur zufällig voneinander unterscheiden. Dies führt zu einer entsprechenden Prüfgröße, die einer F-Verteilung folgt. Diese Berechnungen sind sehr rechenintensiv und werden daher heutzutage nicht mehr per Hand gerechnet.

Kann keine Signifikanz betreffend der diversen Stichproben nachgewiesen werden, so kann die Nullhypothese („Es besteht kein Unterschied zwischen den Stichproben hinsichtlich der untersuchten Variablen.“) angenommen werden. Falls man allerdings ein signifikantes Ergebnis erhält, müßte man korrekterweise formulieren, daß die Nullhypothese zurückgewiesen werden muß („Es ist nicht richtig, daß kein Unterschied zwischen den Stichproben hinsichtlich der untersuchten Variablen besteht.“). Die Frage drängt sich auf, welche von den Mittelwerten sich im Einzelnen paarweise voneinander unterscheiden. Dies wird mit einem sogenannten *Post-hoc-Test* geklärt.

14.4 Ausblick: Medianbezogene Tests

Die hier folgenden Tests entsprechen ihrer Anwendung nach den bisher vorgestellten. Der wichtige Unterschied allerdings ist der, daß bei den folgenden Tests keine Voraussetzung der Normalverteilung der Variablen angenommen wird.

14.4.1 Der U-Test von Mann und Whitney

Der von H. B. Mann und D. R. Whitney im Jahre 1947 entwickelte U-Test dient zum Vergleich zweier Stichproben hinsichtlich ihrer zentralen Tendenz (Median), wobei die Werte beliebig verteilt sein oder Ordinalniveau aufweisen können.

14.4.2 Der Wilcoxon-Test

Der von F. Wilcoxon entwickelte Test dient zum Vergleich zweier abhängiger Stichproben bezüglich ihrer zentralen Tendenzen (Mediane), wobei die Differenzen zusammengehöriger Meßwertepaare nicht wie beim t-Test für abhängige Stichproben normalverteilt sein müssen.

Die häufigste Anwendung findet der Test in der Situation, daß Meßwerte einer Person zu zwei verschiedenen Zeitpunkten vorliegen.

14.4.3 Der H-Test nach Kruskal und Wallis

Der von W. H. Kruskal und W. A. Wallis im Jahr 1952 vorgestellte H-Test dient zum Vergleich von mehr als zwei Stichproben hinsichtlich ihrer zentralen Tendenzen (Mediane), wobei die Werte beliebig verteilt sein können oder Ordinalniveau aufweisen dürfen.

14.5 Der Chiquadrat-Test auf Unabhängigkeit

Das Prinzip einer Kreuztabelle und die Rechenschritte des Chiquadrat-Tests auf Unabhängigkeit sollen anhand eines Beispiels (siehe Tabelle 16) aus meiner Untersuchung erläutert werden (vgl. Kapitel 18.3 auf Seite 87).

	< 1 x / Woche	1 x / Woche	2 - 4 x / Woche	Summe
Schulsportmodell	14	0	0	14
HBLA	16	7	1	24
Summe	30	7	1	38

Tabelle 16: Kreuztabelle: Schultyp - beobachtete Häufigkeit des Alkoholkonsums

Es werden die beiden Variablen „Häufigkeit des Alkoholkonsums“ und „Schultyp“ miteinander in Verbindung gebracht. Die Nullhypothese wird aufgestellt: „Es besteht kein signifikanter Unterschied hinsichtlich der Häufigkeit des Alkoholkonsums in Bezug auf den Schultyp.“ Wir könnten auch fragen: „Unterscheiden sich die beiden Schultypen signifikant bezüglich der Häufigkeit des Alkoholkonsums?“ Korrekter formuliert lautet diese Frage allerdings: „Unterscheiden sich die beobachteten Häufigkeiten signifikant von den erwarteten Häufigkeiten?“

Dabei sind die *erwarteten Häufigkeiten* diejenigen, die man unter Zugrundelegung der gegebenen Randsummen (Zeilen- und Spaltensummen) bei

Gleichverteilung erhalten würde.

Bezeichnet man die Anzahl der Zeilen mit k, die Anzahl der Spalten mit m und die beobachtete Häufigkeit in der i-ten Zeile und j-ten Spalte mit $f_{o_{ij}}$ (f_o ... Frequenz observiert), so berechnet sich die zugehörige erwartete Häufigkeit $f_{e_{ij}}$ (f_e ... Frequenz erwartet) als Quotient vom Produkt aus Zeilensumme und Spaltensumme im Zähler und Gesamtsumme im Nenner für $i = 1 \dots k$ und $j = 1 \dots m$.

$$f_{e_{ij}} = \frac{\sum\limits_{j=1}^{m} f_{o_{ij}} \cdot \sum\limits_{i=1}^{k} f_{o_{ij}}}{\sum\limits_{j=1}^{m} \sum\limits_{i=1}^{k} f_{o_{ij}}}$$

In Tabelle 17 sind die erwarteten Häufigkeiten angeführt. Da es sich um theoretische Werte handelt, sind Nachkommastellen zulässig.

	< 1 x / Woche	1 x / Woche	2 - 4 x / Woche	Summe
Schulsportmodell	11,05	2,58	0,37	14
HBLA	18,95	4,42	0,63	24
Summe	30	7	1	38

Tabelle 17: Kreuztabelle: Schultyp - erwartete Häufigkeit des Alkoholkonsums

Möchte man nun die Frage beantworten, ob die Unterschiede zwischen den beobachteten und den erwarteten Häufigkeiten signifikant sind, muß man zunächst in jedem Feld der Kreuztabelle folgenden Wert berechnen, den man als *standardisiertes quadriertes Residuum* bezeichnet:

$$\frac{\left(f_{o_{ij}} - f_{e_{ij}}\right)^2}{f_{e_{ij}}} \qquad i = 1 \dots k, j = 1 \dots m$$

Tabelle 18 enthält die Ergebnisse:

	< 1 x / Woche	1 x / Woche	2 - 4 x / Woche
Schulsportmodell	0,788	2,58	0,37
HBLA	0,459	1,506	0,217

Tabelle 18: Standardisierte quadrierte Residuen in der Kreuztabelle

Die Prüfgröße χ^2 bei diesem Test wird als Summe aller dieser Residuen berechnet:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{\left(f_{o_{ij}} - f_{e_{ij}}\right)^2}{f_{e_{ij}}}$$

Diese Prüfgröße ist χ^2-verteilt mit $(k-1) \cdot (m-1)$ Freiheitsgraden.

In unserem Beispiel ergibt die Aufsummierung über alle Felder:

$$\chi^2 = 5{,}92$$ [7]

Dies ist nach der χ^2-Tabelle bei

$$df = (2-1) \cdot (3-1) = 2$$

Freiheitsgraden knapp kein signifikanter Wert. Der zu $\chi^2_{2;0,05}$ gehörige Wert ist als 5,991 tabelliert. Streng genommen muß also hier an der Nullhypothese festgehalten werden. Praktisch würde man interpretieren, daß hier bereits Tendenz zur Signifikanz gegeben ist; daß also anhand dieses Tests fast signifikant gezeigt werden konnte, daß die beiden Variablen „Schultyp“ und „Häufigkeit des Alkoholkonsums“ nicht voneinander unabhängig sind.

[7] *Anmerkung:* Dieser Wert stimmt aufgrund von Rundungsfehlern nicht mit dem analogen Wert aus Kapitel 18.3 überein. Dort wurde am Ende auf drei Nachkommastellen gerundet - es ergibt sich 5,911. Für die Interpretation des Ergebnisses ist dies nicht von Belang.

15 Korrelation

Der Zusammenhang zwischen zwei Variablen kann mit dem sogenannten *Korrelationskoeffizienten* beschrieben werden. Da es sich hierbei um Inhalte handelt, die Zusammenhänge direkt oder indirekt proportionaler Art wiedergeben, müssen die beteiligten Variablen mindestens ordinalskaliert (oder dichotom) sein. Ausgeschlossen sind in jedem Fall nominalskalierte Variablen mit mehr als zwei Kategorien.

Die Maßzahl, welche die Strenge oder *Stärke* des Zusammenhangs angibt, ist also der Korrelationskoeffizient, welcher üblicherweise mit r bezeichnet wird und für den gilt:

$$-1 \leq r \leq +1$$

Eine Einstufung der möglichen Werte, die der Korrelationskoeffizient annehmen kann, findet sich in Tabelle 19.

Korrelationskoeffizient	**Einstufung**
$\lvert r \rvert \leq 0,2$	sehr geringe Korrelation
$0,2 < \lvert r \rvert \leq 0,5$	geringe Korrelation
$0,5 < \lvert r \rvert \leq 0,7$	mittlere Korrelation
$0,7 < \lvert r \rvert \leq 0,9$	hohe Korrelation
$0,9 < \lvert r \rvert \leq 1$	sehr hohe Korrelation

Tabelle 19: Einstufung des Korrelationskoeffizienten

Auf einen wichtigen Umstand sei noch hingewiesen: Es sind Fälle denkbar, in denen zwischen zwei Variablen zwar ein Zusammenhang besteht, dennoch aber der Korrelationskoeffizient keinen signifikanten Wert ausweist. In der Praxis sind derartige Fälle recht selten anzutreffen. Es kann aber vorkommen, wenn der Zusammenhang nicht linear, sondern zum Beispiel U-förmig ist.

Vor Berechnung des Korrelationskoeffizienten empfiehlt es sich bei intervallskalierten Variablen dringend, zunächst das zugehörige Streudiagramm zu betrachten.

Eine Übersicht, welches Verfahren in welchem bestimmten Fall das geeignete ist, bietet Tabelle 20.

Skalenniveau	**Verfahren**
beide Variablen intervallskaliert und normalverteilt	Produkt-Moment-Korrelation nach Pearson
mindestens eine Variable ordinalskaliert oder nicht normalverteilt	Rangkorrelation nach Spearman; Rangkorrelation nach Kendall
eine Variable dichotom, die andere intervallskaliert und normalverteilt	punktbiseriale Korrelation
beide Variablen dichotom	Vierfelderkorrelation

Tabelle 20: Skalenniveaus und Korrelationskoeffizienten

15.1 Die Produkt-Moment-Korrelation

Der Produkt-Moment-Korrelationskoeffizient, auch Maßkorrelationskoeffizient oder Korrelationskoeffizient nach Pearson genannt, ist sozusagen der Klassiker unter den Korrelationskoeffizienten. Er dient zur Beschreibung des Zusammenhangs zwischen zwei intervallskalierten und normalverteilten Variablen.

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \cdot (y_i - \overline{y})}{(n-1) \cdot s_x \cdot s_y}$$

Dabei sind $\overline{x}$ und$\overline{y}$ die Mittelwerte sowie s_x und s_y die Standardabweichungen der beiden korrelierenden Variablen. Für den praktischen Gebrauch ist allerdings die folgende Formel zu empfehlen:

$$r = \frac{n \cdot \sum_{i=1}^{n} x_i \cdot y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{\left(n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right) \cdot \left(n \cdot \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)}}$$

$$B = r^2$$

Das Bestimmtheitsmaß (auch: Determinationskoeffizient) ist das Quadrat des Korrelationskoeffizienten. Es gibt den Anteil der gemeinsamen Varianz der beiden Variablen wieder.

15.2 Die Rangkorrelation nach Spearman

Der Rangkorrelationskoeffizient nach Spearman wird zwischen zwei Variablen berechnet, die mindestens ordinalskaliert sind. Er ist auch bei intervallskalierten Variablen zu benutzen, wenn keine Normalverteilung gegeben ist.

Wie bei vielen nichtparametrischen Verfahren gehen nicht die eigentlichen Meßwerte in die Formel ein, sondern die diesen Werten zugeordneten Rangplätze. So ist die Rangkorrelation nach Spearman vor allem auch dann der Korrelation nach Pearson vorzuziehen, wenn Ausreißer vorhanden sind, da diese ja bei einer Rangordnung nicht besonders ins Gewicht fallen.

Zuerst wird für jede der beiden Variablen jeweils eine Rangreihe der Werte erstellt, wobei dem höchsten Wert der Rangplatz 1 zugeteilt wird und bei gleichen Werten entsprechend gemittelte Rangplätze vergeben werden. Anschließend bestimmt man die Differenz d der zugehörigen Rangplatzpaare und hieraus das Quadrat d^2.

Der Rangkorrelationskoeffizient nach Spearman berechnet sich nach der Formel:

$$r = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot (n^2 - 1)}$$

15.3 Ausblick: Weitere Möglichkeiten der Korrelationsberechnung

15.3.1 Die Rangkorrelation nach Kendall

Der Korrelationskoeffizent nach Kendall (oftmals *Kedalls Tau* genannt) wird, ebenso wie der Rangkorrelationskoeffizient nach Spearman, zwischen zwei mindestens ordinalskalierten Variablen berechnet.

15.3.2 Die Vierfelderkorrelation

Die Vierfelderkorrelation wird zwischen zwei dichotomen Variablen berechnet. Mit den Buchstaben a, b, c und d werden üblicherweise die Häufigkeiten einer Vierfeldertafel bezeichnet. Der Vierfelder-Korrelationskoeffizient wird dann berechnet als:

$$r = \frac{a \cdot b - b \cdot c}{\sqrt{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}}$$

Verwendet man für den Zusammenhang zwischen zwei dichotomen Variablen anstelle der Vierfelderkorrelation die Korrelation nach Pearson oder die Rangkorrelation nach Spearman, erhält man für alle drei Korrelationskoeffizienten denselben Wert. Der Vorteil der Vierfelderkorrelation liegt aber in der Einfachheit der Formel, verglichen mit den anderen beiden Verfahren.

15.3.3 Die punktbiseriale Korrelation

Die punktbiseriale Korrelation wird zwischen einer dichotomen und einer intervallskalierten, normalverteilten Variablen berechnet.

Der punktbiseriale Korrelationskoeffizient berechnet sich nach der Formel

$$r = \frac{\overline{x}_1 - \overline{x}_2}{(n_1 + n_2) \cdot s} \cdot \sqrt{n_1 \cdot n_2}$$

Beispielsweise könnte man die intervallskalierte, normalverteilte Variable Körpergewicht hinsichtlich ihres Auftretens bei Frauen und Männern untersuchen. Die dichotome Variable wäre dann das Geschlecht.

16 Regression

Die Korrelationsrechnung bestimmt die Stärke des Zusammenhangs zwischen zwei Variablen. Mit Hilfe der Regressionsrechnung soll der Zusammenhang formelmäßig erfaßt werden. Man versucht, Formeln zu finden, nach denen man aus der Kenntnis des Wertes der einen Variablen den zu erwartenden Wert der anderen (abhängigen) Variablen bestimmen kann. Dabei unterscheidet man zwischen linearen und nichtlinearen Zusammenhängen.

16.1 Lineare Regression

Der Fall der linearen Regression ist der am häufigsten vorkommende. Es sind die beiden Parameter b und a der folgenden Geradengleichung zu ermitteln:

$$y = b \cdot x + a$$

Diese sogenannte *Regressionsgerade* ist diejenige Gerade, für welche die Summe der Quadrate der Abweichungen aller Punkte von dieser Geraden ein Minimum wird. Dabei sind die Abstände parallel zur Ordinate gemeint, so daß es bei der Regressionsrechnung von wesentlicher Bedeutung ist, welche der beiden gegebenen Variablen die abhängige Variable ist, deren Werte in y-Richtung aufgetragen werden.

Den Parameter b nennt man den *Regressionskoeffizient*. Das Vorzeichen von b richtet sich offensichtlich nach dem des zugehörigen Korrelationskoeffizienten: Bei positiver Korrelation ist auch b positiv, bei negativer Korrelation auch b negativ.

Werden üblicherweise mit x die Werte der unabhängigen und mit y die Werte der abhängigen Variablen bezeichnet, so ist der Regressionskoeffizient b definiert durch

$$b = \frac{\sum\limits_{i=1}^{n} (x_i - \overline{x}) \cdot (y_i - \overline{y})}{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}$$

Der Ordinatenabschnitt a berechnet sich nach Kenntnis von b zu

$$a = \frac{\sum_{i=1}^{n} y_i - b \cdot \sum_{i=1}^{n} x_i}{n}$$

In die Formeln für b und a gehen ausschließlich Größen ein, die bereits bei der Berechnung des Produkt-Moment-Korrelationskoeffizienten anfallen (vgl. Kapitel 15.1).

Eine solche Regressionsrechnung bei linearem Zusammenhang sollte nur vorgenommen werden, wenn die Punktwolke nicht allzu sehr um die Regressionsgerade streut, da dann die Vorhersage allzu unsicher wird. Auf alle Fälle muß der Korrelationskoeffizient überhaupt signifikant sein.

16.2 Nichtlineare Regression

Bei nichtlinearen Zusammenhängen, die eher selten auftreten, gibt es praktisch unbegrenzt viele Möglichkeiten der formelmäßigen Gestaltung.

Einige nichtlineare Zusammenhänge lassen sich allerdings durch Logarithmieren in lineare Zusammenhänge überführen. Es sind exponentielle Zusammenhänge der Form

$$y = a \cdot e^{b \cdot x}$$
$$y = a \cdot b^x$$
$$y = a \cdot x^b$$

16.3 Multiple lineare Regression

Von multipler linearer Regression spricht man, wenn die Abhängigkeit einer abhängigen Variablen von *mehreren* unabhängigen Variablen analysiert wird. Die in eine solche Regressionsanalyse eingehenden Variablen müssen intervallskaliert sein.

Mathematisch formuliert geht es darum, bei n unabhängigen Variablen x_i mit $i = 1 \dots n$ (auch Einflußvariablen oder Vorhersagevariablen genannt) und der abhängigen Variablen y die Regressionskoeffizienten b_i mit $i = 1 \dots n$ und die Konstante a folgender Gleichung zu bestimmen:

$$y = b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n + a$$

Derartige Berechnungen werden üblicherweise nur noch unter Zuhilfenahme eines entsprechenden Computerprogrammes durchgeführt.

17 Gütekriterien für Tests

Ein „guter" Test (beziehungsweise eine Untersuchung) muß sogenannte *Gütekriterien* erfüllen. Klassischerweise werden drei solcher Kriterien genannt, die jeder Test zu einem Mindestausmaß zu erfüllen hat.

17.1 Objektivität

Mit *Objektivität* ist gemeint, inwieweit das Testergebnis unabhängig ist von jeglichen Einflüssen außerhalb der Testumgebung, also vom Versuchsleiter, der Art der Auswertung, den Situationsbedingungen, der Zufallsauswahl etc. Teilt man beispielsweise Fragebögen aus, so darf das Ergebnis nicht davon abhängig sein, wer diese Bögen austeilt und auswertet. Sollten in Schulklassen zum Beispiel Fragen betreffend diverser Unklarheiten im Fragebogen auftauchen, so müßten diese streng genommen, tauchten sie in anderen Klassen ebenfalls auf, exakt gleich beantwortet werden, um die Beantwortung der Fragen in keinster Weise zu beeinflussen. Es wird ersichtlich, daß es hinsichtlich der Objektivität mehrere verschiedene Arten gibt, die unterschieden werden müssen. Die Wahl des Versuchsleiters stellt nur eine mögliche Fehlerquelle dar. Im einzelnen ist bei der Testentwicklung anzustreben, daß das Ergebnis unabhängig davon ist,

- wer den Test vorgibt (Durchführungsobjektivität)
- wer den Test auswertet (Auswertungsobjektivität)
- wer den Test interpretiert (Interpretationsobjektivität)

Zusätzlich läßt sich noch feststellen, daß Objektivität eine logische Voraussetzung für Reliabilität ist. Ein Test, der nicht objektiv ist, kann sicherlich ebensowenig hinreichend genau sein.

17.2 Reliabilität

Reliabilität meint das Ausmaß, wie genau ein Test das mißt, was er mißt (unabhängig davon, was er überhaupt mißt). Es ist hier lediglich die Meßgenauigkeit gemeint. Dabei geht es weniger um die Zahl der Dezimalstellen einer Messung oder ähnliches, sondern mehr um die Zuverlässigkeit, mit der bei einer wiederholten Messung unter gleichen Bedingungen dasselbe Meßergebnis zustande kommt.

Außerdem stellt Reliabilität eine logische Voraussetzung für Validität dar. Denn ein Test, der nicht ausreichend genau konstruiert ist, kann schließlich auch nicht ausreichend aussagekräftig sein.

17.3 Validität

Mit *Validität* ist gemeint, inwieweit der Test das mißt, was er messen soll. Es geht also um den Grad der Gültigkeit der Messung und in weiterer Folge um die Aussagefähigkeit des Testergebnisses bezüglich der eigentlichen Absicht der Untersuchung.

18 Untersuchung „Leben sportliche Kinder gesünder?“

Die Motivation betreffend den Inhalt dieser Untersuchung ist teilweise eine persönliche. Aufgrund meiner Tätigkeit als Kunstturntrainer bin ich, man könnte fast sagen ideologischerweise, zumindest aber aus eigenem Idealismus davon überzeugt, daß Sport einen wesentlichen Beitrag zu einer gesunden Lebensführung leistet. Durch diese Untersuchung soll dieser Sachverhalt kritisch beleuchtet werden. Der im Anhang angefügte Fragebogen wurde in zwei Schulklassen ausgeteilt.

Untersuchungsdatum: 13. 4. 2007

Untersuchungsorte:

- Sport- und Musikrealgymnasium / Oberstufenrealgymnasium Salzburg

  Akademiestraße 21

  5020 Salzburg

  Klasse 5 s

- Höhere Bundeslehranstalt für wirtschaftliche Berufe Neumarkt am Wallersee

  Siedlungsstraße 11

  5202 Neumarkt am Wallersee

  Klasse 1 a

Hierbei handelt es sich in beiden Klassen um Schülerinnen und Schüler der gleichen Altersstufe. In die Auswertung gehen 19 Bögen des Schulsportmodells und 24 der HBLA ein.

Die Auswertung des Fragebogens wurde mit Hilfe des Statistikprogramms SPSS (Version 14.0 für Windows) durchgeführt. Alle im folgenden abgebildeten Tabellen wurden vom Programm in dieser Form ausgegeben. Teilweise kommen von mir definierte Variablennamen vor. Eventuell auftretende Unstimmigkeiten betreffend deren Bedeutungen werden im Kontext geklärt.

In den folgenden Kapiteln soll versucht werden, interessante Zusammenhänge aufzudecken und exemplarisch anhand einiger Beispiele zu zeigen, wie statistische Methoden im Bereich der Psychologie zur Anwendung kommen.

18.1 Häufigkeitsbetrachtungen

Um einen ersten Überblick zu bekommen, bedient man sich gerne der Methode der Häufigkeitsbetrachtung. Exemplarisch greife ich hier einige, wie mir scheint, interessante Fragen heraus.

18.1.1 Bereich: Alkohol

alleine

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	nein	43	100,0	100,0	100,0

Gewissermaßen beruhigend erscheint mir die Tatsache, daß niemand angibt, alleine Alkohol zu konsumieren.

ich weiß es nicht

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	ja	8	18,6	18,6	18,6
	nein	35	81,4	81,4	100,0
	Gesamt	43	100,0	100,0	

Im Gegenzug allerdings gibt fast ein Fünftel der Schülerinnen und Schüler an, nicht zu wissen, warum sie Alkohol trinken.

Alkohol Frage 6

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	ja	15	34,9	34,9	34,9
	nein	28	65,1	65,1	100,0
	Gesamt	43	100,0	100,0	

Mehr als ein Drittel der Befragten gibt an, bereits einen Vollrausch gehabt zu haben. Die Interpretation dieser Zahl sei dahingestellt. Auch verweise ich in diesem Zusammenhang auf die Schwierigkeit, wie gewisse Worte verstanden werden, die in einer schriftlichen Befragung gebraucht werden. In einer Schulklasse lassen sich bestimmt mehrere verschiedene „Definitionen“ von Vollrausch finden. Diese Art von Problem stellt überhaupt eine große Schwierigkeit bei derartigen Untersuchungen dar (vgl. Kapitel 18.4).

18.1.2 Bereich: Ernährung

Ernährung Frage 3f

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	trifft nicht zu	7	16,3	16,3	16,3
	trifft wenig zu	24	55,8	55,8	72,1
	trifft zu	9	20,9	20,9	93,0
	trifft voll zu	3	7,0	7,0	100,0
	Gesamt	43	100,0	100,0	

Die Aussage „Ich esse zu fixen Zeiten“ trifft für mehr als zwei Drittel nicht oder nur wenig zu. Bedenkt man allerdings, wie ausgefüllt in unserer Zeit bereits der Alltag von Schülerinnen und Schülern meist ist, so gibt diese Häufigkeit wohl keinen besonderen Anlaß zur Besorgnis.

Ernährung Frage 3j

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	trifft wenig zu	10	23,3	23,3	23,3
	trifft zu	21	48,8	48,8	72,1
	trifft voll zu	12	27,9	27,9	100,0
	Gesamt	43	100,0	100,0	

Zumindest geben über drei Viertel der Befragten an, auf ihre Ernährung zu achten.

Ernährung Frage 3n

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	trifft nicht zu	17	39,5	39,5	39,5
	trifft wenig zu	22	51,2	51,2	90,7
	trifft zu	4	9,3	9,3	100,0
	Gesamt	43	100,0	100,0	

Über 90 % geben an, nicht in erster Linie Fertigprodukte zu essen.

Ernährung Frage 3s

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	trifft nicht zu	29	67,4	67,4	67,4
	trifft wenig zu	9	20,9	20,9	88,4
	trifft zu	3	7,0	7,0	95,3
	trifft voll zu	2	4,7	4,7	100,0
	Gesamt	43	100,0	100,0	

Auch verneinen über zwei Drittel die Aussage „Ich trinke häufig Kaffee“.

Ernährung Frage 3e

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	trifft nicht zu	29	67,4	67,4	67,4
	trifft wenig zu	8	18,6	18,6	86,0
	trifft zu	5	11,6	11,6	97,7
	trifft voll zu	1	2,3	2,3	100,0
	Gesamt	43	100,0	100,0	

Von 43 Personen geben 6 an, auf Diät zu sein. Wenn man bedenkt, daß es sich bei dieser Untersuchung um durchwegs 15- beziehungsweise 16-jährige Jugendliche handelt, erscheint mir diese Zahl schon interessant.

Fragen wir an dieser Stelle einmal nach zusätzlichen Angaben, nämlich denen des Geschlechts und der besuchten Schule:

besuchte Schule * Geschlecht * Ernährung Frage 3e Kreuztabelle

Anzahl

			Geschlecht		
Ernährung Frage 3e			männlich	weiblich	Gesamt
trifft nicht zu	besuchte Schule	Schulsportmodell	12	6	18
		HBLA	1	10	11
	Gesamt		13	16	29
trifft wenig zu	besuchte Schule	Schulsportmodell		1	1
		HBLA		7	7
	Gesamt			8	8
trifft zu	besuchte Schule	HBLA		5	5
	Gesamt			5	5
trifft voll zu	besuchte Schule	HBLA		1	1
	Gesamt			1	1

Wir sehen, daß alle 6 Personen, die angeben, auf Diät zu sein, zum einen weiblich sind und zum anderen die HBLA besuchen.

18.1.3 Bereich: Rauchen

Rauchen Frage 1

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	ja	2	4,7	4,7	4,7
	nein	38	88,4	88,4	93,0
	habe geraucht	3	7,0	7,0	100,0
	Gesamt	43	100,0	100,0	

Hier finde ich persönlich recht interessant, daß drei Leute angeben, geraucht zu haben.

Rauchen Frage 6

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	nein	42	97,7	97,7	97,7
	ja, einmal	1	2,3	2,3	100,0
	Gesamt	43	100,0	100,0	

Nur eine Person gibt an, einmal Marihuana ausprobiert zu haben.

Rauchen Frage 7

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	nein	43	100,0	100,0	100,0

Niemand gibt an, andere Drogen als Marihuana ausprobiert zu haben.

ich weiß es nicht

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	ja	1	2,3	2,3	2,3
	nein	42	97,7	97,7	100,0
	Gesamt	43	100,0	100,0	

Nur eine Person gibt an, nicht zu wissen, warum sie raucht. Im Vergleich zu dem relativ hohen Wert der analogen Variable im Bereich Alkohol scheint dies ein vergleichsweise niedriger Wert. Allerdings gibt es hier auch nur zwei Raucher und drei „Ex-Raucher“. Somit bleibt auch hier interessant, daß zumindest einer von fünf dieser Gruppe nicht weiß, warum er eigentlich raucht.

18.1.4 Bereich: Körperliche Aktivität

ka3 * Geschlecht * besuchte Schule Kreuztabelle

Anzahl

besuchte Schule			Geschlecht: männlich	Geschlecht: weiblich	Gesamt
Schulsportmodell	ka3	3 bis 5 x	2	4	6
		täglich	10	3	13
	Gesamt		12	7	19
HBLA	ka3	höchstens 1 x		2	2
		1 bis 3 x		8	8
		3 bis 5 x		4	4
		täglich		1	1
	Gesamt			15	15

Dieses Ergebnis überrascht nicht wirklich. Die Schülerinnen und Schüler des Schulsportmodells betreiben häufiger Sport als Gleichaltrige aus der HBLA. Zumindest für diese Stichprobe ist dies der Fall. Ob der Unterschied bereits als signifikant eingestuft werden kann, beantworten wir wie folgt:

besuchte Schule * ka3 Kreuztabelle

Anzahl

		ka3: höchstens 1 x	ka3: 1 bis 3 x	ka3: 3 bis 5 x	ka3: täglich	Gesamt
besuchte Schule	Schulsportmodell	0	0	6	13	19
	HBLA	2	8	4	1	15
Gesamt		2	8	10	14	34

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	20,499[a]	3	,000
Likelihood-Quotient	25,997	3	,000
Anzahl der gültigen Fälle	34		

a. 5 Zellen (62,5%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist ,88.

Mit dem Pearson-Chi-Quadrat wird die Hypothese überprüft, daß keine Abhängigkeit zwischen den Zeilen- und Spaltenvariablen besteht. Der eigentliche Wert der Statistik ist nicht sehr aussagekräftig. Der Signifikanzwert (Asymptotische Signifikanz) enthält die gesuchten Informationen. Je kleiner der Signifikanzwert ist, um so unwahrscheinlicher ist es, daß keine Abhängigkeit (Beziehung) zwischen den beiden Variablen besteht. In diesem Fall ist der Signifikanzwert so niedrig, daß er als *,000* angezeigt wird. Dies bedeutet, daß tatsächlich eine Beziehung zwischen den beiden Variablen zu bestehen scheint. Die Aussage „Schüler des Schulsportmodells betreiben häufiger Sport als Schüler der HBLA“ kann demnach als gültig angesehen werden.

ka5g

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	trifft nicht zu	2	4,7	4,7	4,7
	trifft wenig zu	6	14,0	14,0	18,6
	trifft zu	14	32,6	32,6	51,2
	trifft voll zu	21	48,8	48,8	100,0
	Gesamt	43	100,0	100,0	

Über 80 % geben an, daß Sport ihnen hilft, sich gut zu fühlen. Daraus läßt sich vermuten, daß auch diejenigen, die wenig Sport treiben, bereits dieser Ansicht sind.

ka5o

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	trifft nicht zu	1	2,3	2,4	2,4
	trifft wenig zu	2	4,7	4,8	7,1
	trifft zu	13	30,2	31,0	38,1
	trifft voll zu	26	60,5	61,9	100,0
	Gesamt	42	97,7	100,0	
Fehlend	0	1	2,3		
Gesamt		43	100,0		

Daß Sport wichtig für eine gesunde Lebensführung ist, wird wiederum von über 90 % mit „trifft zu“ beziehungsweise „trifft voll zu“ beantwortet.

18.2 Einzelkennwerte

Die Berechnung relevanter Einzelkennwerte wird von SPSS wie folgt ausgegeben:

Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabweichung
Körpergröße (in cm)	43	152	180	166,53	6,464
Körpergewicht (in kg)	36	40	68	53,81	6,568
Body Mass Index (kg/m²)	36	16,65	23,14	19,3734	1,59338
Gültige Werte (Listenweise)	36				

Daß bei Gewicht und in weiterer Folge auch beim Body-Mass-Index (BMI) sieben Werte weniger vorhanden sind, liegt daran, daß sieben Befragte ihr Körpergewicht nicht angegeben haben. Die Variable BMI wurde nicht aus dem Fragebogen direkt erhalten, sondern als zusätzliche Variable im Nachhinein hinzugefügt. Der Body-Mass-Index wird wie folgt berechnet:

$$\text{Körpermassenzahl} = \frac{\text{Masse}}{\text{Größe}^2}$$

Die Masse wird in Kilogramm und die Größe in Metern angegeben. Tabelle 21 gibt Normwerte für diesen Index an.

Kategorie	**BMI** $\left(\frac{kg}{m^2}\right)$
Untergewicht	< 18,5
Normalgewicht	18,5 - 25
Präadipositas	25 - 30
Adipositas Grad I	30 - 35
Adipositas Grad II	35 - 40
Adipositas Grad III	> 40

Tabelle 21: Normwerte für den Body-Mass-Index

Ab einem Wert ≥ 25,0 spricht man im allgemeinen von Übergewicht.

In Bezug auf das Alter gibt es ebenfalls Normwerte betreffend das jeweilige Normalgewicht (siehe Tabelle 22).

Betrachten wir den Body-Mass-Index etwas genauer:

Alter in Jahren	**BMI-Normalwert**
19-24	19 - 24
25-34	20 - 25
35-44	21 - 26
45-54	22 - 27
55-64	23 - 28
> 64	24 - 29

Tabelle 22: BMI-Normwerte betreffend Normalgewicht

Bericht

Body Mass Index (kg/m²)

besuchte Schule	Mittelwert	N	Standardabweichung
Schulsportmodell	19,0980	19	1,37611
HBLA	19,6811	17	1,79797
Insgesamt	19,3734	36	1,59338

Daß die Mittelwerte sich hier wohl nicht signifikant voneinander unterscheiden, liegt vermutlich daran, daß Angaben fehlen. Alle fehlenden Angaben stammen von Schülerinnen der HBLA.

Wir können noch untersuchen, ob die Werte der Körpergröße, des Gewichts sowie BMI-Werte über alle Personen hinweg normalverteilt sind:

Kolmogorov-Smirnov-Anpassungstest

		Body Mass Index (kg/m²)	Körpergewicht (in kg)	Körpergröße (in cm)
N		36	36	43
Parameter der Normalverteilung[a,b]	Mittelwert	19,3734	53,81	166,53
	Standardabweichung	1,59338	6,568	6,464
Extremste Differenzen	Absolut	,110	,108	,064
	Positiv	,110	,108	,053
	Negativ	-,081	-,076	-,064
Kolmogorov-Smirnov-Z		,657	,646	,417
Asymptotische Signifikanz (2-seitig)		,781	,797	,995

a. Die zu testende Verteilung ist eine Normalverteilung.

b. Aus den Daten berechnet.

Daß die Werte des Körpergewichts und somit auch die BMI-Werte nicht sehr signifikant einer Normalverteilung folgen, liegt ebenfalls daran, daß einige Werte fehlen. Man sieht allerdings sehr schön, daß die Variable Körpergröße signifikant normalverteilt ist.

18.3 Signifikante Unterschiede zwischen den Schultypen

Bisher konnten wir feststellen, daß Schüler des Schulsportmodells häufiger Sport treiben als Schüler der HBLA, was nicht sonderlich verwunderte (siehe Seite 83).

Betrachten wir nun die Häufigkeit des Alkoholkonsums:

besuchte Schule * Alkohol Frage 2 Kreuztabelle

Anzahl

		Alkohol Frage 2			
		seltener als 1 x in der Woche	1 x in der Woche	an 2 bis 4 Tagen in der Woche	Gesamt
besuchte Schule	Schulsportmodell	14	0	0	14
	HBLA	16	7	1	24
Gesamt		30	7	1	38

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	5,911[a]	2	,052
Likelihood-Quotient	8,561	2	,014
Anzahl der gültigen Fälle	38		

a. 4 Zellen (66,7%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist ,37.

Hier kann ebenfalls ein signifikanter Unterschied (wenn auch nicht so stark wie bei der Häufigkeit der körperlichen Aktivität) nachgewiesen werden (vgl. Kapitel 14.5).

Betrachten wir weiters die Stärke des Alkoholkonsums in Bezug auf das Geschlecht:

Kreuztabelle

Anzahl

		Alkohol Frage 2			
		seltener als 1 x in der Woche	1 x in der Woche	an 2 bis 4 Tagen in der Woche	Gesamt
Geschlecht	männlich	8	0	0	8
	weiblich	22	7	1	30
Gesamt		30	7	1	38

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	2,702[a]	2	,259
Likelihood-Quotient	4,319	2	,115
Anzahl der gültigen Fälle	38		

a. 3 Zellen (50,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist ,21.

Hier kann kein signifikanter Unterschied mehr festgestellt werden, wenn wohl auch Tendenz zur Signifikanz gegeben sein mag, nämlich dahingehend, daß Schülerinnen mehr Alkohol konsumieren als Schüler dieser Altersstufe.

Betrachten wir weiters das Rauchverhalten. Kann hier ein signifikanter Unterschied zwischen den Schultypen nachgewiesen werden?

besuchte Schule * Rauchen Frage 1 Kreuztabelle

Anzahl

		Rauchen Frage 1			
		ja	nein	habe geraucht	Gesamt
besuchte Schule	Schulsportmodell	0	19	0	19
	HBLA	2	19	3	24
Gesamt		2	38	3	43

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	4,479[a]	2	,107
Likelihood-Quotient	6,349	2	,042
Anzahl der gültigen Fälle	43		

a. 4 Zellen (66,7 %) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist ,88.

Offensichtlich ist dies nicht ganz der Fall. Der Wert 0,107 weist allerdings ebenfalls auf eine Tendenz zur Signifikanz hin, nämlich dahingehend, daß Schüler der HBLA mehr rauchen als Schüler des Schulsportmodells. Wie man ja auch sieht, gab niemand der 19 Schüler des Schulsportmodells an, zu rauchen beziehungsweise geraucht zu haben. Alle beantworteten die Frage mit nein.

18.4 Fazit

Hinsichtlich der drei Hauptfragen, nämlich nach der Häufigkeit des Alkoholkonsums sowie des Nikotinkonsums und der körperlichen Aktivität konnte nun festgestellt werden:

- Bei der Häufigkeit des Alkoholkonsums liegt Tendenz zur Signifikanz dahingehend vor, daß Schüler der HBLA häufiger Alkohol konsumieren als Schüler des Schulsportmodells.
- Hinsichtlich der Häufigkeit des Rauchens von Zigaretten liegt ebenfalls Tendenz zur Signifikanz dahingehend vor, daß Schüler der HBLA häufiger rauchen respektive geraucht haben als Schüler des Schulsportmodells.
- Schließlich liegt höchste Signifikanz vor in Bezug auf die Häufigkeit körperlicher Aktivität pro Woche. Schüler des Schulsportmodells treiben höchst signifikant häufiger Sport als Schüler der HBLA.

Daß damit die plakativ formulierte Frage, ob sportliche Kinder gesünder leben, nicht hinreichend beantwortet werden kann, liegt nahe. Dennoch kann das Ergebnis so interpretiert werden, daß Schüler, die ein Schulsportmodell besuchen (und somit regelmäßig Leistungssport betreiben) weniger dazu neigen, Alkohol und Zigaretten zu konsumieren als Schüler einer HBLA.

Interpretationsmöglichkeit Das erreichte Ergebnis könnte nun in der Praxis dazu beitragen, genauer zu untersuchen, in welcher Form, wie oft etc. von den Schülern des Schulsportmodells Sport betrieben wird. In weiterer Folge könnte versucht werden, auch in anderen Schulformen den Unterricht dahingehend zu verändern, daß der Sport einen höheren Stellenwert bekommt, als er im Moment hat. Selbstverständlich befinden wir uns dann bereits auf dem Niveau bildungspolitischer Entscheidungen. Mehr Sport in anderen Schultypen als ohnehin sportspezifischen verlangt nach Investitionen, die sich kurzfristig gesehen vermutlich nicht rentieren. Dennoch können in solchen Fällen wiederholte Studien Beweise erbringen, daß längerfristig derartige, wenn auch momentan sicherlich kostenintensive Investitionen Früchte tragen.

Schwierigkeiten bei psychologischen Untersuchungen

Eine große Herausforderung bei psychologischen Untersuchungen vor allem schriftlicher Art stellt die Wortwahl dar. So einfach dies klingen mag, so differenziert muß diese allerdings betrachtet werden. Der Untersuchende meint mit einem bestimmten Wort möglicherweise etwas gänzlich Verschiedenes als die untersuchten Personen. Aus diesem Grund ist man bestrebt, auch bei Untersuchungen wie Fragebögen möglichst klar zu *messen*. Das bedeutet, man versucht, Antworten numerischer Art zu erhalten. Diese sind in jedem Fall objektiv. Möchte man also eine genaue Untersuchung bezüglich der Häufigkeit eines bestimmten Verhaltens pro Woche durchführen, ist es besser ohne Umschweife zu fragen: „Wie oft gehen Sie dieser oder jener Beschäftigung in der Woche nach?“ Im Nachhinein lassen sich so gewonnene Variablen immer noch in einige (wenn nötig wenige) Klasse einteilen. Grundsätzlich sollte nach Pytagoras’ Leitmotiv „Alles ist Zahl“ vorgegangen werden, denn dann kann man zum einen besser damit rechnen. Auf der anderen Seite bleibt, wie gesagt, die Objektivität erhalten.

In meinem Fragebogen sind mir aufgrund meiner Unerfahrenheit auf diesem Gebiet solche Fehler unterlaufen, was mir im Zuge der Auswertung klarer geworden ist. Im Punkt „Ernährung 3s“ beispielsweise soll Stellung genommen werden zu der Aussage „Ich trinke häufig Kaffee“. Wieviel ist aber denn nun *häufig*? Unter diesem Begriff wird sich nicht jeder Befragte dieselbe Zahl vorstellen. Es wäre hier besser gewesen, gleich nach der getrunkenen Anzahl von Tassen Kaffee pro Woche zu fragen.

In der Psychologie kommen nun noch zusätzlich nicht selten Variablen ins Spiel, bei denen überhaupt erst eigens Einteilungen getroffen werden müssen, um eine bestimmte Eigenschaft messen zu können. In einem Gespräch mit einer Kollegin am Institut für Psychologie Wien wurde mir beispielsweise erzählt, daß es eine noch junge Studie darüber gebe, daß extrovertierte Menschen glücklicher seien. Ganz abgesehen davon, welche Absicht diese Studie verfolgt, so kann ihr Inhalt sicherlich nicht zweifelsfrei sinnvoll sein. Gibt es auf der einen Seite genormte Tests, die den Grad an Extrovertiertheit feststellen, so kann es allerdings andererseits niemals einen derartigen Test in Bezug auf den Faktor Glück geben. Wie möchte man entscheiden, ob gewisse Handlungen oder Gefühle (die wohl abgefragt werden können) tatsächlich beim Menschen generell bedeuten, daß er sich glücklich fühlt. Mit Problemen dieser Art befaßt sich die *Methodenlehre* in der Psychologie, ein Teilgebiet derer auch die Statistik darstellt.

Zu guter Letzt stellt sich bei psychologischen Untersuchungen immer die Frage, inwieweit die Ergebnisse tatsächlich aussagekräftig sind und zwar ob der Tatsache, daß die erhaltenen Informationen nicht unbedingt der Wirklichkeit entsprechen müssen. Man muß damit rechnen, daß einige Befragte bei einigen Antworten möglicherweise bewußt (oder auch unbewußt) falsche Angaben machen (beispielsweise bei der Angabe des Körpergewichts). Teilweise ist es dann möglich, an verschiedenen Stellen Kontrollfragen einzubauen. Dann können alle Fragen zu einem bestimmten Item zusammengefaßt und miteinander verglichen werden. Ergeben sich Widersprüche, wird der betroffene Fall in diesem Punkt aus der Auswertung ausgeschlossen. Daß dies allerdings nicht immer möglich ist, zeigt bereits die Tatsache, daß es zu verschiedenen Fragestellungen aus der Psychologie immer wieder ähnliche Untersuchungen mit konträren Ergebnissen gibt.

Verbesserungsmöglichkeiten der Untersuchung

Im Nachhinein weiß man meist einiges besser. Diese Weisheit trifft auch auf meine kleine Untersuchung zu. Deshalb möchte ich hier noch anführen, welche Dinge adhoc verändert werden könnten, um Verbesserungen zu erzielen.

Für eine exaktere Auswertung könnte in Frage 3 im Bereich Körperliche Aktivität nach der genauen Stundenanzahl gefragt werden. Also etwa: „Wieviele Stunden pro Woche betreibst du Sport?“ anstelle der vorgegebenen Auswahlmöglichkeiten. Man hätte anstatt einer ordinalskalierten eine intervallskalierte Variable gewonnen. Zusätzlich stellte der Wert Null kein Problem dar.

Bei den Bereichen Alkohol und Rauchen fehlt bei mehreren Auswahlfragen die Option *Trinke nicht* beziehungsweise *Rauche nicht*. Beim Punkt Alkohol wirkte sich das in der Auswertung nicht so stark aus wie beim Punkt Rauchen. Beinahe jeder der Befragten trinkt ab und zu Alkohol. Allerdings waren fast alle Befragten ihren Angaben nach Nichtraucher. Deshalb blieben die Fragen 2 bis 8 bei fast allen Antwortbögen leer, da die Option für die Angabe, daß man Nichtraucher ist, bei diesen Fragen fehlt.

Bei Frage 9 im Bereich Rauchen sollte entschieden werden, wieviel „beste/r Freund/in“ und wieviel „Geschwister“ rauchen. Nun kam allerdings auch der Fall vor, daß hier Mehrfachantworten gemacht wurden, da man ja von vorn herein nicht ausschließen kann, daß jemand eine beste Freundin *und* einen besten Freund hat, die beide nicht dieselben Rauchgewohnheiten haben. Das gleiche gilt auch für Geschwister. Hier müßte man die Angaben als Untersuchender präziser machen.

Abschließend möchte ich noch auf die relativ geringen Fallzahlen hinweisen. Bei einer derartigen Untersuchung reicht es natürlich in der Praxis nicht aus, allein zwei Schulklassen unterschiedlicher Schultypen zur Betrachtung heranzuziehen. Die Stichproben sollten deutlich größer sein.

Anmerkung: Die Befragung der Klasse des Schulsportmodells wurde in einer Supplierstunde durchgeführt. Die Klasse hätte sonst Religion gehabt. Mir wurde im Nachhinein mitgeteilt, daß nicht alle Schüler für den Religionsunterricht angemeldet sind. Nun könnte es beispielsweise sein, daß die Variable „Besuch des Religionsunterrichts“ mit von mir untersuchten Variablen positiv korreliert ist. Das könnte bedeuten, daß im Schulsportmodell fünf Schüler nicht anwesend waren und daß genau diese fünf Schüler, die Religion nicht besuchen, die fünf Raucher der Klasse sind. In der Praxis müssen auch solche Feinheiten beachtet werden.

Alles in allem gestaltet sich eine derartige Untersuchung meist recht aufwendig. Wird sie jedoch gewissenhaft ausgewertet, können interessante, für die Praxis bedeutende Zusammenhänge aufgedeckt werden.

Literatur

[1] Jürgen BORTZ: *Statistik für Human- und Sozialwissenschaftler*, 6., vollständig überarbeitete und aktualisierte Auflage, Heidelberg: Springer, 2005

[2] Jürgen BORTZ, Nicola DÖRING: *Forschungsmethoden und Evaluation*, 4., überarbeitete Auflage, Heidelberg: Springer, 2006

[3] Joerg M. DIEHL, Heinz Ulrich KOHR: *Deskriptive Statistik*, 4., erweiterte Auflage, Frankfurt am Main: Eschborn - Fachbuchhandlung für Psychologie, Verlagsabteilung, 1982

[4] Edith HARTL: *Statistische Testverfahren und ihre Anwendung in der Psychologie*, Technische Universität Wien, 1982

[5] Joachim HARTUNG: *Statistik. Lehr- und Handbuch der angewandten Statistik*, 14., unwesentlich veränderte Auflage, München: Oldenbourg, 2005

[6] Walter KRÄMER: *Statistik verstehen*, 5. Auflage, München: Piper, 2006

[7] Jürgen JANSSEN, Wilfried LAATZ: *Statistische Datenanalyse mit SPSS für Windows*, 5., neu bearbeitete und erweiterte Auflage, Heidelberg: Springer, 2005

[8] Jürgen ROST: *Lehrbuch Testtheorie, Testkonstruktion*, Göttingen: Huber, 1996

[9] Lothar SACHS: *Statistische Methoden*, 3., neu bearbeitete Auflage, Heidelberg: Springer, 1976

[10] Ulrich TRÄNKLE: *Statistische Methoden in der Psychologie*, Darmstadt: Wissenschaftliche Buchgesellschaft, 1985

[11] Ulrich TRÄNKLE: *Mathematische und statistische Methoden für Studierende der Psychologie, Biologie, Medizin, Pädagogik und Soziologie*, Münsteraner Skripten zur Psychologie, Heft 1, 2., neubearbeitete und erweiterte Auflage, Münster: Aschendorff, 1991

[12] Peter ZÖFEL: *Statistik für Psychologen*, München: Pearson Studium, 2003

A Fragebogen

Geschlecht: ○ weiblich ○ männlich
Geburtsjahr: 19_____
Größe (in cm): ____________ Gewicht (in kg): ____________

Alkohol

1. **Mit welchem Alter hast du das erste mal Alkohol getrunken?**

   mit _____ Jahren

2. **Wie oft trinkst du aktuell Alkohol?** (kreuze eine Möglichkeit an)

   ○ seltener als 1 x in der Woche ○ 1 x in der Woche
   ○ an 2 bis 4 Tagen in der Woche ○ öfters

3. **Welche Alkoholika trinkst du?** (Mehrfachantworten möglich)

   ○ Mixgetränke, Alkopops ○ Bier
   ○ Wein, Spritzer ○ Prosecco, Frizzante, Sekt
   ○ Liköre ○ Schnäpse pur (Vodka, Tequila, . . .)

4. **In welchen Situationen trinkst du Alkohol?** (Mehrfachantworten möglich)

   ○ bei Feiern (Weihnachten, Geburtstag, . . .) ○ bei Veranstaltungen
   ○ mit Freundinnen und Freunden ○ in Discos und Discoparties
   ○ alleine

5. **Warum trinkst du Alkohol?** (Mehrfachantworten möglich)

   ○ weil andere auch trinken ○ zur Entspannung
   ○ aus Langeweile ○ bei Problemen in der Schule
   ○ aus Protest ○ um Regeln zu brechen
   ○ um erwachsen zu sein ○ um zu einer Clique dazuzugehören
   ○ um betrunken zu sein ○ ich weiß es nicht
   ○ andere Gründe

6. **Hast du schon einmal einen Vollrauch gehabt?**

   ○ ja ○ nein

7. **Fühlst du dich betrunken, nachdem du Alkohol konsumiert hast?**

   ○ eher ja ○ eher nein

Ernährung

1. **Wie ernährst du dich?**

   ○ normal ○ vegetarisch ○ vegan ○ mache Diät

2. **Wieviel trinkst du pro Tag?**

   ○ weniger als 1 Liter ○ 1 bis 2 Liter ○ 2 bis 3 Liter ○ mehr als 3 Liter

3. **Beurteile folgende Aussagen:** (kreuze in jeder Zeile eine Möglichkeit an)

		trifft voll zu	trifft zu	trifft wenig zu	trifft nicht zu
a)	Ich koche manchmal selbst	○	○	○	○
b)	Ich esse täglich Obst oder Gemüse	○	○	○	○
c)	Einmal pro Woche esse ich bei McDonalds	○	○	○	○
d)	Ich esse gerne Süßigkeiten	○	○	○	○
e)	Ich bin auf Diät	○	○	○	○
f)	Ich esse zu fixen Zeiten	○	○	○	○
g)	Bei uns zu Hause wird immer gekocht	○	○	○	○
h)	Einmal täglich esse ich eine Wurstsemmel	○	○	○	○
i)	Wenn ich in der Nacht Hunger bekomme, plündere ich den Kühlschrank	○	○	○	○
j)	Ich achte auf meine Ernährung	○	○	○	○
k)	Ich kaufe ausschließlich Bioprodukte	○	○	○	○
l)	Einmal pro Woche esse ich eine Wurstsemmel	○	○	○	○
m)	Ich nehme Nahrungsergänzungsmittel wie Vitamine, Eiweiß etc. . .	○	○	○	○
n)	Ich esse hauptsächlich Fertigprodukte	○	○	○	○
o)	Nach dem Essen bin ich oft so richtig voll	○	○	○	○
p)	Abends esse ich nichts mehr	○	○	○	○
q)	Ich trinke fast nur Mineralwasser oder Wasser	○	○	○	○
r)	Ich trinke regelmäßig Cola, Fanta, Sprite, Keli, Gröbi, Radelberger oder ähnliche Getränke	○	○	○	○
s)	Ich trinke häufig Kaffee	○	○	○	○

Rauchen

1. **Rauchst du Zigaretten?**

   ○ ja ○ nein ○ habe geraucht

2. **Wie alt warst du, als du mit dem Rauchen begonnen hast?**

   _____ Jahre

3. **Wieviele Zigaretten rauchst du durchschnittlich pro Tag?**

   _____ Stück / Tag

4. **Rauchst du am Wochenende (beim Fortgehen) mehr als unter der Woche?**

ja, viel mehr	ja, etwas mehr	in etwa gleich	nein, rauche weniger	rauche am Wochenende gar nicht
○	○	○	○	○

5. **Falls du rauchst, willst du mit dem Rauchen aufhören?**

   ○ ja ○ nein

6. **Hast du schon einmal gekifft?**

   ○ ja, öfters ○ ja, selten ○ ja, einmal ○ nein

7. **Hast du schon einmal andere Drogen außer Marihuana probiert?**

   ○ ja, öfters ○ ja, selten ○ ja, einmal ○ nein

8. **Warum rauchst du?** (Mehrfachantworten möglich)

   ○ weil andere auch rauchen
   ○ aus Langeweile
   ○ es schmeckt mir
   ○ ich mag das Schwindelgefühl
   ○ ich rauche zum Kaffee
   ○ zur Entspannung
   ○ bei Problemen in der Schule
   ○ um zu einer Clique dazuzugehören
   ○ ich weiß es nicht
   ○ andere Gründe

9. **Beurteile folgende Aussagen:** (kreuze in jeder Zeile eine Möglichkeit an)

		raucht täglich	raucht manchmal	raucht micht	ich weiß es nicht	habe oder sehe diese Person nicht
a)	Mutter	○	○	○	○	○
b)	Vater	○	○	○	○	○
c)	beste/r Freund/in	○	○	○	○	○
d)	Geschwister	○	○	○	○	○

Körperliche Aktivität

1. **Wie hältst du dich körperlich fit?** (Mehrfachantworten möglich)

   ○ ich halte mich nicht bewußt fit
   ○ ich betreibe folgende Sportart: ____________
   ○ ich gehe laufen
   ○ ich besuche ein Fitness-Studio
   ○ anderes, nämlich: ____________

2. **Falls du eine Sportart betreibst, wie lange tust du das schon?**

   _____ Jahre

3. **Falls du eine Sportart betreibst, wie oft pro Woche tust du das?**

   ○ höchstens 1 x ○ 1 bis 3 x ○ 3 bis 5 x ○ täglich

4. **Bist du Mitglied in einem Sportverein?**

   ○ ja ○ nein

5. **Beurteile folgende Aussagen:** (kreuze in jeder Zeile eine Möglichkeit an)

		trifft voll zu	trifft zu	trifft we-nig zu	trifft nicht zu
a)	Ich bin sportlich	○	○	○	○
b)	Ich bin Leistungssportler/in	○	○	○	○
c)	Ich gehe laufen	○	○	○	○
d)	Ich habe einen Trainer	○	○	○	○
e)	Im Sommer treibe ich eher Sport als im Winter	○	○	○	○
f)	Ich schwitze gerne beim Sport	○	○	○	○
g)	Sport hilft mir, mich gut zu fühlen	○	○	○	○
h)	Ich trainiere nach einem Plan	○	○	○	○
i)	Sport hilft mir, gut auszusehen	○	○	○	○
j)	Ich treffe Freunde beim Sport	○	○	○	○
k)	Österreich ist ein sportliches Land	○	○	○	○
l)	Ich trainiere zu festgelegten Zeiten	○	○	○	○
m)	Ich halte mich fit	○	○	○	○
n)	Sport ist wichtig für eine gesunde Lebensführung	○	○	○	○

Alkohol und Zigaretten

1. **Welche Gefahr für körperliche Schäden oder Schäden anderer Art gehen deiner Meinung nach Leute ein, wenn sie . . .** (kreuze in jeder Zeile eine Möglichkeit an)

		kein Risiko	geringes Risiko	großes Risiko	ich weiß es nicht
a)	gelegentlich Zigaretten rauchen	○	○	○	○
b)	eine oder mehrere Schachteln Zigaretten pro Tag rauchen	○	○	○	○
c)	ein oder zwei Gläser Alkohol fast jede Woche trinken	○	○	○	○
d)	ein- oder zweimal Substanzen (wie Klebstoff, Benzin) schnüffeln	○	○	○	○
e)	jeden Tag Alkohol trinken	○	○	○	○
f)	ein- oder zweimal Marihuana probieren	○	○	○	○
g)	regelmäßig Marihuana rauchen	○	○	○	○
h)	gelegentlich andere Drogen konsumieren	○	○	○	○

2. **Wenn du die folgenden Fragen beantwortest, denke an deine Freunde, mit denen du die meiste Zeit in der Freizeit verbringst.** (kreuze in jeder Zeile eine Möglichkeit an)

		niemand	weniger als die Hälfte	ungefähr die Hälfte	mehr als die Hälfte	alle	ich weiß es nicht
a)	Wie viele deiner Freunde rauchen Zigaretten?	○	○	○	○	○	○
b)	Wie viele deiner Freunde trinken Alkohol?	○	○	○	○	○	○
c)	Wie viele deiner Freunde nehmen Marihuana oder andere Drogen?	○	○	○	○	○	○

3. Wenn du Alkohol trinken wolltest (oder bereits tust), denkst du, daß deine Eltern es dir erlauben würden?

würden es mir erlauben / erlauben es	würden es mir nicht zu Hause erlauben / erlauben es nicht zu Hause	würden es mir überhaupt nicht erlauben / erlauben es überhaupt nicht	ich weiß es nicht
○	○	○	○

4. Macht eines deiner Geschwister folgendes? (kreuze in jeder Zeile eine Möglichkeit an)

		eher ja	eher nein	ich weiß es nicht	ich habe keine Geschwister
a)	Alkohol trinken	○	○	○	○
b)	sich betrinken	○	○	○	○
c)	Marihuana rauchen	○	○	○	○
d)	Substanzen schnüffeln	○	○	○	○
e)	andere Drogen nehmen	○	○	○	○

5. Wenn du rauchen wolltest (oder bereits tust), denkst du, daß deine Eltern es dir erlauben würden?

würden es mir erlauben / erlauben es	würden es mir nicht zu Hause erlauben / erlauben es nicht zu Hause	würden es mir überhaupt nicht erlauben / erlauben es überhaupt nicht	ich weiß es nicht
○	○	○	○