

EMPIRISCHE SOZIALFORSCHUNG

V 2.0 (01.06.2003)

HERBERT NAGEL

Inhaltsverzeichnis

1	Grundlagen und Übersicht	4
1.1	Wissen und Wissenschaft	4
1.2	Aufgaben von Studien	5
1.3	Phasen einer Studie	6
1.4	Fragen	6
2	Formulierung und Präzisierung des Untersuchungsproblems	7
2.1	Hypothesen	7
2.2	Variablen und Beobachtungseinheiten	8
2.3	Fragen	9
3	Planung und Vorbereitung der Erhebung	11
3.1	Operationalisierung	11
3.2	Messung	12
3.2.1	Gütekriterien der Messung	12
3.2.2	Skalenniveaus	12
3.3	Zeitliche Abgrenzung von Studien	13
3.4	Experimente	15
3.4.1	Abhängige, unabhängige und Störvariable	15
3.4.2	Pretest und Posttest	16
3.4.3	Versuchsgruppe und Kontrollgruppe	16
3.4.4	Quasi – Experimente	17
3.5	Population und Stichprobe	17
3.5.1	Zufallsauswahl	18
3.5.2	Quotenauswahl	19
3.5.3	Willkürliche Auswahl	19
3.6	Erhebungsinstrument Fragebogen	20
3.6.1	Fragestellungen	20
3.6.2	Fragearten	20
3.6.3	Richtlinien für die Fragenformulierung	22
3.6.4	Pretests des Fragebogens	23
3.7	Fragen	23
4	Datenerhebung	26
4.1	Befragung	26
4.1.1	Schriftliche Befragung	26
4.1.2	Face-to-Face Interviews	27
4.1.3	Telefoninterviews	28
4.2	Beobachtung (Feldforschung)	28
4.2.1	Teilnehmende versus nichtteilnehmende Beobachtung	29
4.2.2	Offene versus verdeckte Beobachtung	29
4.2.3	Feldbeobachtung versus Beobachtung im Labor	30
4.2.4	Unstrukturierte versus strukturierte Beobachtung	30
4.3	Nichtreaktive Verfahren	31
4.3.1	Inhaltsanalyse	31
4.3.2	Verwendung bestehender Daten	32
4.4	Fragen	32
5	Datenauswertung	34
5.1	Datenmatrix, Fehlerkontrolle und Fehlerbereinigung	34
5.2	Einfache deskriptive Statistiken	35
5.3	Inferenzstatistische Methoden	36

5.3.1	Ziele der Inferenzstatistik	36
5.3.2	Signifikanzniveau und p-Wert	36
5.3.3	Fehlermöglichkeiten bei Signifikanztests	37
5.4	Fragen.....	38
6	Berichterstattung	40
6.1	Projektbericht	40
6.2	Ethik und Politik in der Forschung.....	40
6.3	Fragen.....	41
7	Eine kategoriale Variable	42
7.1	Numerische und grafische Beschreibung	42
7.2	Test für den Anteilswert	44
7.3	Fragen.....	45
8	Zwei kategoriale Variablen.....	48
8.1	Numerische und grafische Beschreibung	48
8.1.1	Gesamtprozent und gruppierte Balkendiagramme	48
8.1.2	Zeilen- und Spaltenprozent, gestapelte Balkendiagramme	49
8.2	Homogenitätstest	51
8.3	Fragen.....	52
9	Eine metrische Variable	54
9.1	Numerische und grafische Beschreibung	54
9.1.1	Histogramm.....	54
9.1.2	Lagemaße.....	55
9.1.3	Streuungsmaße.....	56
9.1.4	Boxplot	56
9.1.5	Vergleich mehrerer Verteilungen	57
9.2	Test für den Mittelwert.....	59
9.3	Fragen.....	60
10	Zwei metrische Variablen.....	65
10.1	Numerische und grafische Beschreibung	65
10.1.1	Streudiagramm.....	65
10.1.2	Korrelationskoeffizient.....	66
10.1.3	Weitere Beispiele	66
10.2	Test für den Korrelationskoeffizienten.....	67
10.3	Fragen.....	68
11	Eine metrische und eine kategoriale Variable.....	71
11.1	Grafische Beschreibung.....	71
11.2	Vergleich der Mittelwerte von zwei Gruppen.....	72
11.3	Fragen.....	73
12	Weitere statistische Verfahren	75
13	Bibliographie	76

1 Grundlagen und Übersicht

1.1 Wissen und Wissenschaft

Unser herkömmliches Wissen basiert zum Großteil auf Übereinkunft und Glauben. Im Rahmen des Aufwachsens, der Erziehung und des Erwachsenwerdens wird der Grundstock individuellen Wissens aufgebaut, in erster Linie durch Akzeptieren von bereits bekanntem Wissen anderer. **Tradition** oder „überliefertes Wissen“ ist ein kumulativer Prozess, der uns von bereits vorhandenem Wissen profitieren lässt. **Autorität**, basierend auf Kompetenz, Glaubwürdigkeit aber auch Status, führt dazu, dass aus der Mitteilung des Wissens ein Akzeptieren wird. Unsere Anstrengung kann auf die Entdeckung und Entwicklung neuen Wissens konzentriert werden. Tradition und Autorität führen uns dadurch zu einem Ausgangspunkt für unser persönliches Streben nach neuem Wissen; wir müssen nicht mehr bei Null beginnen. Es muss uns aber auch bewusst sein, dass wir möglicherweise von einem falschen Standpunkt ausgehen oder in eine falsche Richtung suchen.

Für dieses Skriptum ist **Wissenschaft** eine Forschungsmethode, die Wege weist, die Welt um uns herum besser kennen zu lernen. Um bereits vorhandenes Wissen anderer zu akzeptieren, soll dieses entweder logisch ableitbar oder empirisch (anhand der Wirklichkeit) nachweisbar sein. Wissenschaft stellt aber auch Werkzeuge für neuen Erkenntnisgewinn zur Verfügung, für uns ist es das Tripel: **Theorie + Datengewinnung + Datenanalyse**. Aus der Theorie werden Aussagen abgeleitet, die anhand von gewonnenen Daten (also empirisch) überprüft werden. Passen die Daten zu den Aussagen, so unterstützen die Daten diese Aussagen. Ist dies nicht der Fall, widersprechen sich also Aussagen und Daten, so muss die Theorie verworfen oder zumindest revidiert werden. Nach Karl Popper kann eine Theorie nie bewiesen (verifiziert) sondern nur widerlegt (falsifiziert) werden. Im Wesentlichen folgt empirische Sozialforschung einem Konzept, wie es auch in den Naturwissenschaften zur Anwendung kommt.

Die Anwendung dieser Instrumente schließt aber nicht aus, dass nicht auch Fehler auftreten können. Mögliche Ursachen gibt es viele, die häufigsten sind:

- **Eingeschränkte Sichtweise:** Wir sind geprägt von unserer jeweiligen Vergangenheit (Erziehung, Ausbildung,...) und Gegenwart (Lebensphase, Beruf,...). Dadurch hat jeder Mensch eine andere Sichtweise der Dinge und es besteht die Gefahr, dass der Forscher die beobachteten Tatsachen auf seine individuelle Art interpretiert.
- **Unterdrückte Tatsachen:** In Studien müssen sog. „nicht relevante“ Informationen vernachlässigt werden. Was relevant, was nicht relevant ist, unterliegt persönlicher Beurteilung und ist unter Umständen fehlerhaft.
- **Selektive Wahrnehmung:** Persönliche Erwartungen können dazu führen, dass Beobachtungen verfälscht oder ignoriert werden (man sieht nur das, was einem „ins Konzept passt“). Es werden Regelmäßigkeiten dort „entdeckt“, wo es keine gibt. Das geschieht (meist) nicht willentlich, sondern passiert einem.
- **Unzulässige Verallgemeinerungen und Schlüsse:** Von wenigen Beobachtungen wird auf die Allgemeinheit geschlossen. Aus dem Zusammenhang von Ereignissen werden Ursache-Wirkung-Beziehungen abgeleitet (Speiseeiskonsum und Tote bei Badeunfällen).
- **Unlogische Begründungen:** der oft zitierte Satz, dass „Ausnahmen die Regel bestätigen“, macht wissenschaftlich keinen Sinn. In keinem logischen System kann eine Ausnahme als Beweis dienen. Auch die landläufige Annahme von einigen Casino-Besuchern, dass beim

Roulette nach mehrmaligem Rot die Wahrscheinlichkeit, und damit die Gewinnchancen, für Schwarz größer werden, ist falsch und steht im Widerspruch zur Wahrscheinlichkeitsrechnung.

Kritik an der wissenschaftlichen Forschung

Das Auffinden von Regelmäßigkeiten ist ein großes Ziel in der Wissenschaft. In den Sozial- und Wirtschaftswissenschaften ist es im Konkreten das Auffinden von Regelmäßigkeiten im sozialen und wirtschaftlichen Leben. Werden solche Regelmäßigkeiten entdeckt, wird oft einer der folgenden Kritikpunkte geäußert:

- Die Regelmäßigkeiten sind trivial (Umweltbewusste wählen „Die Grünen“, Selbstständige wählen ÖVP,...).
- Ausnahmen sind möglich und widersprechen den Regelmäßigkeiten (einzelne Frauen verdienen mehr als bestimmte Männer, es gibt auch schwarze, blaue und grüne Gewerkschafter,...).
- Personen haben einen freien Willen und können sich bewusst diesen Regelmäßigkeiten widersetzen.

Diese Kritikpunkte können entkräftet werden:

- Die Dokumentation und Bestätigung von Offensichtlichem kann durchaus wertvoll im Sinn wissenschaftlicher Forschung sein. Allzu oft schon haben sich „offensichtliche Erkenntnisse“ als falsch erwiesen.
- Soziale und wirtschaftliche Regelmäßigkeiten sind probabalistische (Wahrscheinlichkeits-) Aussagen, sie müssen nicht für 100 Prozent der Betroffenen gültig sein.
- Soziale und wirtschaftliche Regelmäßigkeiten sind keine Naturgesetze, nach denen sich Individuen zu verhalten haben; sie dienen zur Beschreibung des Verhaltens des Großteils der Individuen.

1.2 Aufgaben von Studien

Die wichtigsten Gründe für Studien sind:

- **Forschung:** Studien zu Forschungszwecken werden üblicherweise durchgeführt, um (1) die Machbarkeit einer größeren Studie schon im Vorfeld zu klären, um (2) Methoden zu entwickeln, die in anschließenden Studien zum Einsatz kommen sollen und um (3) die Neugier von Forschern zu stillen und deren Wunsch nach einem besseren Verstehen der sozialen Welt nachzukommen.
- **Beschreibung:** Aufgabe vieler sozialwissenschaftlicher Studien ist es, das Verhalten von Personen zu beschreiben. Eine Umfrage während des Wahlkampfs beschreibt das mögliche Abstimmungsverhalten der Bevölkerung. Eine Studie unter Interessenten für ein bestimmtes Produkt versucht das Konsumentenverhalten zu beschreiben.
- **Erklärung:** Eine weitere wichtige Aufgabe von Studien ist es, Vorgänge zu erklären. Mit der sog. Sonntagsfrage kann man das Abstimmungsverhalten beschreiben. Mit einer sog.

Wählerstromanalyse beschreibt man die Verschiebung von Wählerstimmen und kann (teilweise) Wahlergebnisse erklären.

- **Prognose:** Das Hauptziel vieler wirtschaftlicher Studien ist die Prognose von Ereignissen. Wenn ein Freizeitzentrum errichtet wird, mit wie viel Besuchern kann man pro Tag rechnen? Wie gehen die Wahlen aus? Meist bilden gute Erklärungsmodelle die Basis für Prognosen.

1.3 Phasen einer Studie

Wir haben jetzt ein theoretisches Fundament für die Art und Weise wie üblicherweise empirische Sozialforschung betrieben wird. Darauf aufbauend können empirische Untersuchungen durchgeführt werden. Diekmann (2002) gliedert den Ablauf einer empirischen Untersuchung grob in fünf Hauptphasen:

- Formulierung und Präzisierung des Forschungsproblems
- Planung und Vorbereitung der Erhebung
- Datenerhebung
- Datenauswertung
- Berichterstattung

Nicht jede Untersuchung muss genau diesem Schema entsprechen. Es ist aber ein Gerüst, das erste empirische Arbeiten aufzubauen helfen kann. Die nächsten fünf Kapitel des Skriptums widmen sich diesen fünf Phasen, ihren Aufgaben aber auch ihren Gefahren.

In weiteren fünf Kapiteln wird für bestimmte Datensituationen gezeigt, wie diese Daten numerisch und/oder grafisch beschrieben werden können und wie eine Schlussfolgerung aus diesen Daten gezogen werden kann.

1.4 Fragen

Für dieses Kapitel ist es ausreichend, wenn Sie Antworten auf folgende Fragen geben können:

- 1) Wodurch erwerben wir den Großteil unseres Wissens?
- 2) Was sind mögliche Fehlerquellen bei sozialwissenschaftlichen Studien?
- 3) Warum werden sozialwissenschaftliche Studien durchgeführt?.

2 Formulierung und Präzisierung des Untersuchungsproblems

Eine empirische Arbeit sollte mit einem klar definierten Forschungsproblem beginnen, etwas moderater formuliert, mit klar formulierten Fragestellungen. Nicht wenige Arbeiten kranken daran, dass „irgend etwas“ im sozialen Bereich untersucht werden soll, ohne dass ein Forschungsziel auch nur annähernd klar umrissen worden wäre. Die Hoffnung wird in die Zukunft gesetzt, wenn dann die Daten erhoben, werden sich die „Hypothesen von selbst anbieten“.

Viele empirische Studien dienen der Überprüfung von Hypothesen. Was sind Hypothesen?

2.1 Hypothesen

Im allgemeinen Sinn ist eine Hypothese eine Vermutung über einen bestimmten Sachverhalt. In der Regel formulieren Hypothesen Beziehungen zwischen Variablen. So könnte eine Hypothese lauten, dass Wellnessurlaubsangebote Frauen eher ansprechen als Männer. Die beiden Variablen, die in Beziehung stehen, sind das Interesse an Wellnessurlauben und das Geschlecht.

Hypothesen können im Rahmen von Untersuchungen anhand von Daten überprüft werden. Diese Überprüfung geschieht in der Phase der Datenauswertung. Die ursprüngliche Hypothese wird in ein Hypothesenpaar gepackt, in eine **Nullhypothese** und eine **Alternativhypothese**. Die Nullhypothese unterstellt immer, dass keine Unterschiede zwischen zwei oder mehreren Gruppen bestehen, dass kein Zusammenhang zwischen Variablen besteht, dass eine Maßnahme keinen Effekt hat, etc. In die Alternativhypothese wird die eigentliche Hypothese gestellt; also es besteht ein Unterschied zwischen verschiedenen Gruppen, es besteht ein Zusammenhang zwischen Variablen, eine Maßnahme hat einen Effekt, etc. Die Alternativhypothese ist also die interessantere der beiden Hypothesen. Im Beispiel des Wellnessurlaubs würde die Nullhypothese lauten: Frauen und Männer haben gleiches Interesse an Wellnessurlauben, die Alternativhypothese wäre: Frauen haben ein größeres Interesse an Wellnessurlauben als Männer.

Wie gelangt man zu Hypothesen?

Mehrere Wege können zu interessanten Fragestellungen führen, einige sind eher Imitation früherer Studien, andere geben der eigenen Kreativität großen Raum.

- **Wissenschaftliche Literatur:** beim Studium wissenschaftlicher Zeitschriften entdeckt man einen interessanten Artikel. Eine kleine Modifikation der dort vorgestellten Studie kann eine neue Studie sein. Wenn etwa für die Schweiz gezeigt worden ist, dass Frauen mehr Interesse an Wellnessurlauben zeigen, so kann eine Untersuchung für Österreich nach demselben Design durchgeführt werden.
- **Replikation einer Studie:** verdienstvoll aber nicht sehr beliebt sind Replikationen (Wiederholungen) von Studien. Eine einmal unter speziellen geprüfte Hypothese ist vielleicht nur unter bestimmten Bedingungen gültig, findet aber ohne Replikation als

generelle Hypothese Eingang in die Lehrbücher. Bei einer Replikationsstudie ist der Aufwand weit geringer als bei der Ursprungsstudie.

- **Theorie:** aus wissenschaftlichen Theorien können überprüfbare Hypothesen abgeleitet werden. Die empirische Prüfung der Hypothesen (und damit der Theorie) ist der Normalfall in theoretisch-empirischen Wissenschaften und entspricht dem deduktiv-empirischen Wissenschaftsmodell Poppers.
- **Deskriptive Studien:** es gibt empirische Studien, deren Zweck nicht die Überprüfung von Hypothesen ist, sondern die Gewinnung von Datenmaterial über meist neue Phänomene. Aus den Erkenntnissen dieser Studie können Hypothesen entwickelt und formuliert werden, die in einer weiteren Studie überprüft werden können.
- **Eigenes Interesse:** man interessiert sich für bestimmte Themen, hat auch Beobachtungen dazu gemacht (nicht im Sinn von Stichproben) und hat damit den Ausgangspunkt für eine Studie.
- **Auftragsforschung:** nicht wenige Studien dienen nicht primär dem wissenschaftlichen Erkenntnisgewinn, sondern sind Auftragsstudien von Unternehmen, Ministerien, Interessensverbänden etc. Bei solchen Studien bestimmt der Auftraggeber das Forschungsziel und legt damit auch weitgehend die Hypothesen fest, die überprüft werden sollen.

Es kommt vor, dass sich interessante Fragestellungen im Verlauf einer Arbeit ergeben. Eine allgemeine Empfehlung, ob diese Fragestellungen in einer Anschlussstudie untersucht oder durch eine Revision des Forschungsplans in die aktuelle Studie aufgenommen werden sollen, ist nicht möglich.

2.2 Variablen und Beobachtungseinheiten

Die Datenerhebung dient üblicherweise dazu, Informationen über eine bestimmte wohldefinierte Menge von **Beobachtungseinheiten** zu gewinnen. Oft sind diese Beobachtungseinheiten Personen, es können aber auch Transaktionen (Buchungen, Käufe,...), Ereignisse (Unfälle, Erkrankungen,...), Organisationen (Vereine, Gewerkschaften, ..) untersucht werden. Die Gesamtmenge aller Beobachtungseinheiten nennt man **Population**. Das können je nach Untersuchung alle wahlberechtigten Österreicher, alle Gäste einer Wintersportregion, alle potentiellen Kunden eines Internetversands, alle Unfälle auf einer Autobahn,... sein.

An diesen Beobachtungseinheiten werden bestimmte Charakteristika beobachtet, man nennt ein solches Charakteristikum **Variable** (Merkmal).

Sind die Forschungshypothesen formuliert, ist implizit auch festgelegt, welches die abhängigen und welches die unabhängigen Variablen in der Studie sind. Als **abhängige Variablen** (erklärte Variablen, Responsevariablen) werden jene Variable bezeichnet, die in der Studie erklärt oder genauer untersucht werden. **Unabhängige Variablen** (erklärende Variablen) sind jene Variablen, die zur Erklärung oder zur Definition von Gruppen dienen.

Im Beispiel mit dem Wellnessurlaub wären die Beobachtungseinheiten Personen (das Gebiet, das untersucht wird legt die Population fest, also Großraum Wien, Westösterreich, EU, ...). Das

Interesse am Wellnessurlaub (wie es gemessen werden kann, sehen wir erst im nächsten Kapitel) ist die abhängige Variable, das Geschlecht der Person die unabhängige Variable.

In die erste Phase gehört auch das Studium von schon vorhandenen Studien zu ähnlich gelagerten Fragestellungen. Wie wurde dort vorgegangen, welche Methoden wurden verwendet, welche Variablen wurden erfasst? Es kann vorkommen, dass in einer solchen Studie Daten erhoben wurden, mit denen auch die eigenen Fragestellungen bearbeitet werden können. Im akademischen Bereich ist es leicht, solche Daten zu erhalten; im kommerziellen Bereich werden Daten nur in Sonderfällen weiter gegeben. Angenommen man kommt zu solchen Daten, ist es dann noch sinnvoll, sich selbst den üblicherweise sehr großen Aufwand der Planung und Durchführung der Datenerhebung anzutun? Ein Grund, wohl der einzig argumentierbare, ist die Aktualität der Daten.

2.3 Fragen

Frau Maier hat mit ihrer Abteilung trotz mancher Schwierigkeiten ein Projekt erfolgreich abgeschlossen. Zum Teil als Belohnung zum Teil als Fortbildung kann sie an einem einwöchigen Seminar über Mitarbeiterführung teilnehmen. Sie interessiert sich, ob sie aus dem Seminar etwas für ihre alltägliche Arbeit mitnehmen kann.

- 1) Wie könnte eine Hypothese lauten, die zu Frau Maiers Frage passt?
- 2) Wie würde die diese Hypothese in einem Hypothesenpaar, bestehend aus Null- und Alternativhypothese formuliert lauten?
- 3) Was ist bei dieser Hypothese die abhängige Variable? Gibt es eine unabhängige Variable?
- 4) Wenn Frau Maier ihre Hypothese überprüfen wollte, was wäre die Population ihrer Untersuchung, was wären die Untersuchungseinheiten?

Mögliche Antworten:

- 1) Wie könnte eine Hypothese lauten, die zu Frau Maiers Frage passt?

Eine (man kann auch andere finden) Hypothese könnte sein: „Die Teilnahme am Seminar fördert das Arbeitsklima in meiner Abteilung“.

- 2) Wie würde die diese Hypothese in einem Hypothesenpaar, bestehend aus Null- und Alternativhypothese formuliert lauten?

Obige Hypothese kann in die Nullhypothese „Die Teilnahme am Seminar führt zu keiner Verbesserung des Arbeitsklimas in meiner Abteilung“ und die Alternativhypothese „Die Teilnahme am Seminar führt zu einer Verbesserung des Arbeitsklimas in meiner Abteilung“ aufgesplittet werden.

- 3)** Was ist bei dieser Hypothese die abhängige Variable? Gibt es eine unabhängige Variable?

Die abhängige Variable ist das Arbeitsklima. In diesem Beispiel gibt es keine unabhängigen Variablen.

- 4)** Wenn Frau Maier ihre Hypothese überprüfen wollte, was wäre die Population ihrer Untersuchung, was wären die Untersuchungseinheiten?

Die Überprüfung sollte wohl an der Abteilung von Frau Maier stattfinden. Die Beobachtungseinheiten wären die Mitarbeiter ihrer Abteilung (Streitfall, ob Frau Maier auch mitzuzählen ist), die Population wäre also die ganze Abteilung.

3 Planung und Vorbereitung der Erhebung

Nach der konkreten Formulierung des Forschungsproblems geht es in dieser Phase darum, die in den Hypothesen auftretenden Begriffe zu definieren und operationalisieren, d.h. einer Messung zugänglich zu machen. Das kann bei mehrdimensionalen Begriffen (wie etwa „Umweltbewusstsein“) eine aufwändige Arbeit sein, bei einfacheren Begriffen (etwa „Interesse an Wellnessurlaube“) mit einigen sorgfältigen Überlegungen abgedeckt sein.

In diesen Bereich gehört auch die Auswahl der Mess- und Skalierungsmethoden.

Eine wichtige Entscheidung bezieht sich auf den zeitlichen Aspekt der Datenerhebung, auf die Wahl zwischen Querschnitt- und Längsschnitterhebung. Ebenfalls ist zu bedenken, ob eine Vergleichs- oder Kontrollgruppe explizit berücksichtigt werden soll.

Sind diese Entscheidungen getroffen, ist die Bestimmung von Typ und Größe der Stichprobe offen. Spätestens hier ist auch das Stichwort Studienbudget zu erwähnen.

In jedem Fall sollte das gewählte Erhebungsinstrument (Fragebogen, Beobachtungsschema, ..) einem Pretest unterzogen werden, bei großen Projekten können es auch mehrere Pretests sein.

3.1 Operationalisierung

Unter Operationalisierung eines Begriffes versteht man die „Übersetzung“ des Begriffes in Operationen zur Messung jener beobachtbaren Variablen, die mit dem Begriff zusammenhängen. Jede empirische Untersuchung setzt Operationalisierungen jener Begriffe voraus, zu denen die Untersuchung durchgeführt werden soll. Das bedeutet, dass man diese Begriffe präzise fasst und angibt, durch welche Indikatoren sie in der Realität wahrgenommen werden können.

Für viele Begriffe ist unmittelbar ersichtlich, was damit gemeint ist und wie sie zu messen sind. Beispiele hierfür ist etwa das Geschlecht einer Person oder die Absatzmenge eines Produktes (die operationale Eigenschaft dafür erfordert zumindest, wann und wo zu zählen ist). Der Begriff Monatseinkommen ist schon etwas schwieriger; zwar ist wohl den meisten klar, was gemeint ist (nach einer Einigung auf Brutto- oder Nettomonatseinkommen). Wie misst man das aber den in immer häufiger auftretenden Fällen einer nicht dauernden Anstellung, bei Selbstständigen?

Bei Begriffen wie „Werbewirkung“, „Einstellung“, „soziale Schicht“ ist aber nicht so klar, was gemeint ist. Nicht nur eine Begriffsdefinition ist gefordert sondern die Angabe von Indikatoren, über die man den Begriff messbar machen kann.

Noch schwieriger sind Begriffe wie „Umweltbewusstsein“ zu fassen, die komplex und mehrdimensional sind. In einer sog. Konzeptspezifikation müssen die einzelnen Dimensionen des Begriffs herausgearbeitet werden und die einzelnen Dimensionen gesondert analysiert werden.

3.2 Messung

3.2.1 Gütekriterien der Messung

Messungen sollen möglichst objektiv, zuverlässig und gültig sein.

Objektivität

Objektivität einer Messung bedeutet, dass das Messergebnis unabhängig von der Person ist, die das Messinstrument anwendet. Das klassische Beispiel ist ein Schulaufsatz, der von mehreren Deutschlehrern unabhängig voneinander benotet wird. Wenn die Noten stark variieren, kann man nur schwer von einer objektiven Beurteilung sprechen. Allgemein ist Objektivität immer dann gefährdet, wenn es um Beurteilungen geht. Genaue Kriterien für die Beurteilung sind daher notwendig. Auch das unterschiedliche Auftreten von Interviewern gegenüber Befragten kann zu nicht objektiven Messergebnissen führen. Auch hier sind einheitliche Richtlinien für die Interviewer bei einer Befragung notwendig.

Reliabilität (Zuverlässigkeit)

Reliabilität eines Messinstruments liegt vor, wenn bei mehrfacher Messung das gleiche Messergebnis erzielt wird. Wird die Körpergröße eines Menschen bestimmt, so sollte (zumindest approximativ) annähernd derselbe Wert ermittelt werden. Diese Forderung, die man an physikalische Messinstrumente stellt, erwartet man auch von sozialwissenschaftlichen Messinstrumenten.

Validität (Gültigkeit)

Validität bedeutet, dass das Messinstrument auch tatsächlich das misst, was es messen soll. Messen Intelligenztests tatsächlich Intelligenz oder, wie manchmal kritisiert wird, nur „das, was Intelligenztests messen“. Validität ist stark von einer guten Operationalisierung und Konzeptspezifikation abhängig.

Diese Gütekriterien werden in der Literatur zum Teil noch weiter untergliedert: Durchführungsobjektivität, Auswertungsobjektivität, Inhaltsvalidität, Kriteriumsvalidität, Konstruktvalidität, etc. Für kleine Untersuchungen sollte es genügen, sich dieser Anforderungen bewusst zu sein. Für große Untersuchungen ist der Einsatz von Tests zu überlegen, mit denen überprüft wird, ob diese Kriterien eingehalten werden.

3.2.2 Skalenniveaus

Die üblicherweise vorgenommene Einteilung der Skalenniveaus führt zu einer Hierarchie von Skalentypen. Beginnend mit dem niedrigsten Skalenniveau erhält man:

- **Nominalskala:** im einfachsten Fall des Messens sind die möglichen Ausprägungen der Variablen lediglich Bezeichnungen von Klassen. So sind die Ausprägungen der Variable

„Geschlecht“ „männlich“, bzw. „männlich“. Weitere Beispiele für nominalskalierte Variablen sind Beruf, Religionsbekenntnis, Wohnort etc.

- **Ordinalskala:** Ausprägungen ordinalskalierteter Variablen können in eine sinnvolle Rangreihe gebracht werden. Beispiele sind Güteklassen bei Lebensmitteln, Schulnoten, Präferenzen bei Speisen, etc.
- **Intervallskala:** bei intervallskalierten Variablen können Differenzen zwischen den Ausprägungen sinnvoll interpretiert werden, nicht aber Verhältnisse. Die Zeitrechnung im westlichen Sinn, mit dem aus nichtreligiöser Sicht gesehen, willkürlichen Nullpunkt Christi Geburt ist ein Beispiel dafür. Die Differenz von 2000 und 1997 hat, als Zeitspanne gesehen, die gleiche Bedeutung wie die Differenz von 2003 und 2000. Allerdings macht es keinen Sinn davon zu sprechen, dass ein Ereignis im Jahr 2000 doppelt so spät erfolgt ist wie ein vergleichbares Ereignis im Jahr 1000.
- **Verhältnisskala (Ratioskala, Rationalskala):** in Ergänzung zur Intervallskala sind bei dieser Skala auch Verhältnisse sinnvoll interpretierbar. Beispiele dafür sind etwa Einkommen (es macht Sinn, davon zu sprechen, dass Person X doppelt so viel wie Person Y verdient), Dauer eines Urlaubsaufenthaltes, etc.
- **Absolutskala:** Beispiele hierfür sind Häufigkeiten (etwa Tischreservierungen für einen Ball, Beschäftigtenzahl eines Betriebes, etc.) und Wahrscheinlichkeitswerte (etwa für die „0“ bei Roulette, etc.).

Achtung: Meist werden nominal- oder ordinalskalierte Variable für die Datenanalyse mit Zahlen codiert, etwa die Variable „Geschlecht“ wird mit „1“ für „männlich“ und „2“ für „weiblich“ codiert. Wenn man nur mehr die Zahlen sieht und vergisst, dass es sich dabei um die Variable „Geschlecht“ handelt, werden leicht Operationen durchgeführt, die für nominalskalierte Variablen eigentlich keinen Sinn machen, etwa das arithmetische Mittel oder die Standardabweichung berechnet.

Für viele statistische Anwendungen ist eine gröbere Einteilung ausreichend, nämlich die Einteilung in kategoriale und metrische Variable.

- **Kategoriale Variable** sind nominal- oder ordinalskalierte Variable.
- **Metrische Variable** sind mindestens intervallskaliert. Bei metrischen Variablen ist nicht nur die Zuordnung von Zahlen zu den Variablenwerten sinnvoll, es sind auch so gut wie alle statistischen Verfahren (im einfachsten Fall die Berechnung des arithmetischen Mittels) anwendbar.

Diese Einteilung wird auch für die Besprechung statistischer Methoden in den letzten Kapiteln des Skriptums verwendet.

3.3 Zeitliche Abgrenzung von Studien

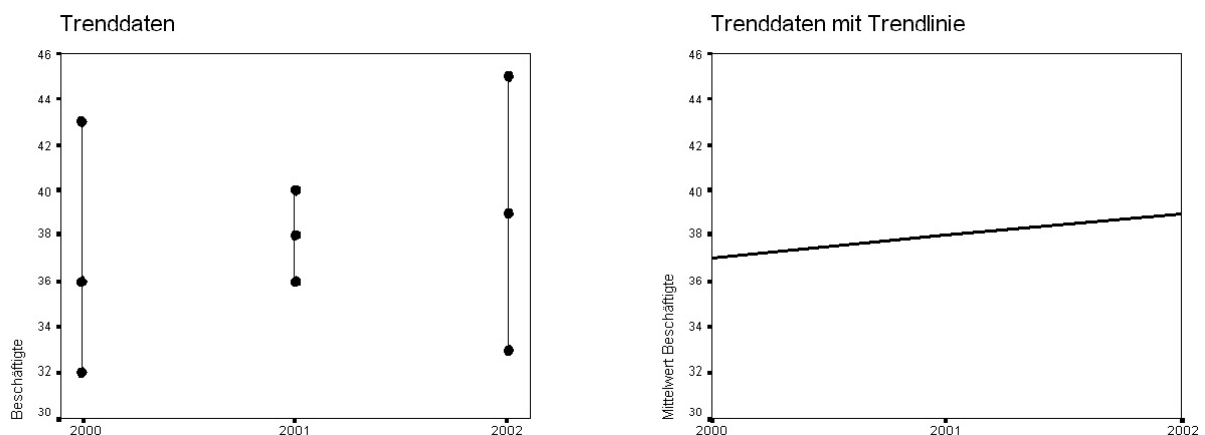
Mit dem Erhebungsdesign wird der zeitliche Modus der Datenerhebung festgelegt. Wir unterscheiden drei Arten von Erhebungsdesigns:

- Querschnittsdesign
- Trenddesign
- Paneldesign

Die Datenerhebung wird entsprechend als Querschnitt-, Trend- oder Panelerhebung bezeichnet.

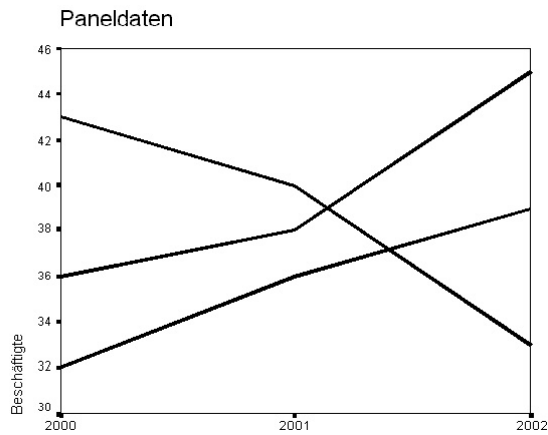
Die Datenerhebung beim **Querschnittsdesign** bezieht sich auf einen Zeitpunkt oder eine kurze Zeitspanne, in der eine einmalige Erhebung der Eigenschaften (Variablen) bei N Untersuchungseinheiten vorgenommen wird. Die meisten sozialwissenschaftlichen Studien folgen einem Querschnittsdesign.

Bei einem **Trenddesign** werden die Werte der gleichen Variablen zu mehreren Zeitpunkten an jeweils unterschiedlichen Stichproben erhoben. Man kann sich eine Trenderhebung einfach als Abfolge mehrerer Querschnitterhebungen zum gleichen Thema vorstellen. Vergleichbar über die Zeit sind dann Kennziffern über die Stichprobe, wie Mittelwerte, Prozentwerte. Man kann also aggregierte Trends ableiten.



In den Jahren 2000, 2001 und 2002 wurde bei jeweils drei Betrieben die Anzahl an Beschäftigten erhoben. Es liegen Querschnittsdaten für drei Zeitpunkte vor. Da es sich nicht um dieselben Betriebe handeln muss, kann keine Beschäftigungsentwicklung der einzelner Betriebe aus den Daten abgeleitet werden. Als Zusammenfassung kommt die Berechnung von Mittelwerten für die drei Zeitpunkte in Frage und deren Darstellung als Trendlinie.

Mit dem **Paneldesign** werden die Werte der gleichen Stichprobe zu mehreren Zeitpunkten an ein und derselben Stichprobe erhoben. Die einzelnen Erhebungen eines Panels werden als Panelwellen bezeichnet. Man kann damit Veränderungen auf der individuellen Ebene nachvollziehen. Allerdings bringen Panelerhebungen einen hohen organisatorischen Aufwand mit sich, müssen doch die Elemente der Stichprobe zu mehreren Zeitpunkten befragt werden. Da in vielen Fällen die Elemente der Stichprobe Personen sein werden, muss mit folgenden Schwierigkeiten gerechnet werden: Tod, Umzug an eine andere Adresse (Aktualisierung einer Adressdatei), Umzug in ein anderes Land (bedeutet meist ein Ausscheiden aus der Stichprobe), etc. Panelerhebungen mit vielen Wellen (mehr als drei) oder über einen langen Zeitraum sind daher selten.



Im Unterschied zu Trenddaten erlauben Paneldaten die Beobachtung individueller Veränderungen. Basierend auf den identen Zahlenwerten wie im Trenddatenbeispiel kann man hier erkennen, dass in einem Betrieb die Beschäftigtenzahl zweimal gesunken ist, im Unterschied zu den beiden anderen Betrieben, wo sowohl 2001 und 2002 mehr Beschäftigte tätig waren.

Zwischen den drei Designtypen existiert eine Informationshierarchie. Panelerhebungen sind informativer als Trenderhebungen, diese informativer als Querschnitterhebungen.

Sowohl Trend- als auch mit Paneldaten können in ein sog. **Kohortendesign** eingehen. Als Kohorte wird eine Bevölkerungsgruppe bezeichnet, die durch ein zeitlich gemeinsames, längerfristig prägendes Startereignis definiert ist. Je nach Startereignis kann es sich um Alters- oder Geburtenkohorten, Eheschließungskohorten oder Berufseintrittskohorten handeln, um die häufigsten Kohortendefinitionen zu erwähnen.

3.4 Experimente

Ebenfalls in die Planung und Vorbereitung einer Untersuchung gehört die Entscheidung, ob ein Experiment oder eine Umfragestudie durchgeführt wird.

3.4.1 Abhängige, unabhängige und Störvariable

Bei einem Experiment wird die Auswirkung der **unabhängigen Variablen**, die in Form eines Stimulus auf die **abhängige Variable** einwirkt, gemessen. Die unabhängige Variable ist der verursachende Reiz, die abhängige (zu untersuchende) Variable ist die Auswirkung des Reizeinflusses. Es geht also darum, Ursache-Wirkungsverhältnisse aufzudecken.

Auf die abhängige Variable haben meist viele Variable einen Einfluss, im Experiment wird aber üblicherweise nur der Einfluss einiger weniger (oft nur einer einzigen) Variablen untersucht. Die anderen Einflüsse werden für das konkrete Experiment als **Störvariable** betrachtet. Störvariable können im Experiment berücksichtigt werden, indem diese konstant gehalten oder kontrolliert werden, indem durch Randomisieren der Versuchspersonen die Störeinflüsse in der Versuchs- und Kontrollgruppe etwa gleich sind oder dass die Störvariablen als unabhängige Variablen in das Experimentaldesign aufgenommen werden.

3.4.2 Pretest und Posttest

Beim einfachsten Fall des Experiments wird die abhängige Variable vor (**Pretest**) und nach (**Posttest**) dem Einwirken des Reizes gemessen. Die Abweichungen zwischen den beiden Messungen werden auf den Einfluss der unabhängigen Variablen, also auf den Stimulus, zurückgeführt.

Allerdings besteht die Gefahr von sog. **Pretest-Effekten**, dass etwa Lerneffekte aus dem Pretest die Ergebnisse des Posttests beeinflussen.

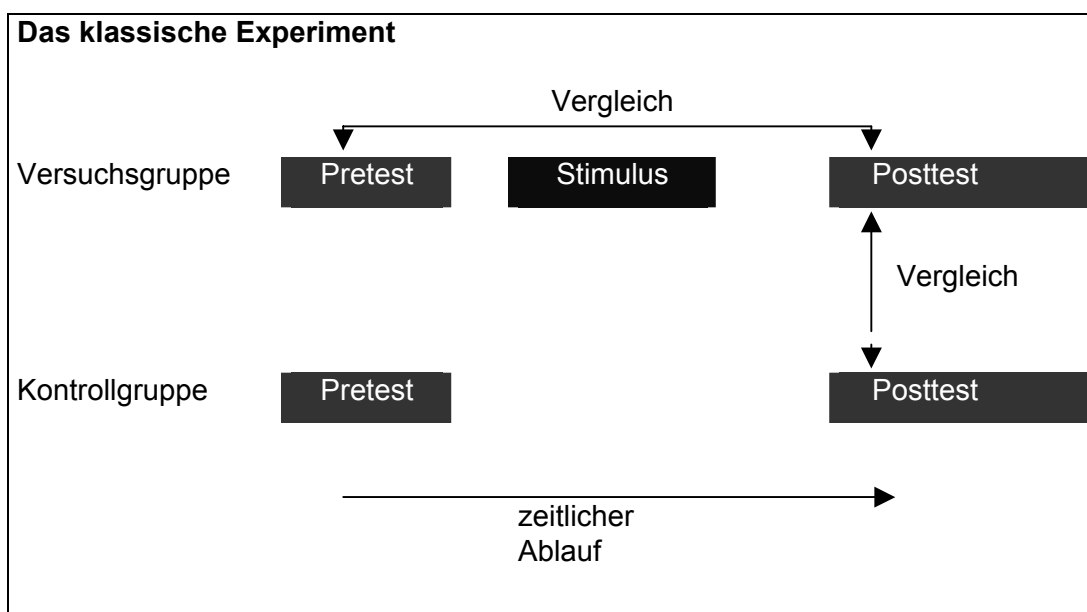
3.4.3 Versuchsgruppe und Kontrollgruppe

Meist enthält das Design von Experimenten neben der **Versuchsgruppe** (Experimentalgruppe), die dem Stimulus ausgesetzt ist, eine **Kontrollgruppe**, auf die der Reiz nicht einwirkt. Dadurch kann der Einfluss der unabhängigen Variablen kontrolliert werden. Da die Kontrollgruppe den Reiz nicht erfährt, kann überprüft werden, ob der Stimulus das Ergebnis (der Versuchsgruppe) bewirkt hat oder ob andere Effekte die Messung beeinflusst haben.

Randomisierung

Die Zuteilung der Untersuchungsobjekte auf die Versuchs- und Kontrollgruppe ist zentrales Thema des Experiments. Unterschiede zwischen den Gruppen werden in experimentellen Untersuchungen durch Randomisierung, also die Zuteilung der Personen auf die Gruppen nach dem Zufallsprinzip, minimiert.

Mit diesen Begriffen ist der Aufbau des klassischen Experiments abgeschlossen. Die folgende Abbildung soll noch einmal deren Beziehung unter einander verdeutlichen.



Versuchsleitereffekt

Die unbeabsichtigte oder unbewusste Beeinflussung der Versuchspersonen durch den Versuchsleiter nennt man Versuchsleitereffekt. Das kann etwa dadurch geschehen, dass der Versuchsleiter unbewusst etwas freundlicher blickt, wenn die zu prüfenden Hypothesen bestätigt werden. Ein Ausweg ist, einen neutralen Versuchsleiter zu installieren oder nicht bekannt zu geben, ob die Versuchsperson der Versuchs- oder Kontrollgruppe zugehört.

3.4.4 Quasi – Experimente

Die zufällige Aufteilung der Versuchspersonen auf Experimental- und Kontrollgruppe ist manchmal nicht möglich, sie ist vorgegeben. Wenn diese Prämisse experimenteller Designs nicht erfüllt ist, spricht man von quasi-experimentellen Designs oder Quasi-Experimenten.

Wenn man die Leistungen in Mathematik zwischen Hauptschülern und Gymnasiasten vergleichen will, ist die Aufteilung der Stichprobe vorgegeben. So bestehen zwischen den Schultypen nicht nur Unterschiede hinsichtlich der unabhängigen Variablen (z.B. Art des Unterrichtsstils), sondern zusätzlich hinsichtlich vieler anderer Variablen (z.B. Motivation, Intelligenz, sozialer Status), die das Messergebnis ebenfalls beeinflussen können (Störvariable).

3.5 Population und Stichprobe

Bei Umfragen ist es im Allgemeinen nicht möglich, sämtliche nur denkbaren Beobachtungseinheiten zu befragen. Solche als **Vollerhebungen** bezeichneten Datenerhebungen sind zeit- und kostenintensiv. Beispiele dafür sind periodisch wiederkehrende Volkszählungen, Befragungen von Interessensvertretungen (Wirtschaftskammer, Ärztekammer, etc.) ihrer Mitglieder, Abstimmung in einem Verein. Allerdings wird auch bei diesen Beispielen ein Ausschöpfungsgrad von 100% so gut wie nie erreicht.

In den meisten Fällen wird man nur einen Teil der Population befragen können, wird also eine **Stichprobenerhebung** durchführen. Nach der Art wie die Stichprobe ermittelt wird, unterscheidet man nach:

- Zufallsauswahl
- Quotenauswahl
- Willkürliche Auswahl

Repräsentativität

Markt- und Meinungsforschungsinstitute sprechen gerne von einem „repräsentativen Querschnitt“, von einer Stichprobe als „verkleinertem Abbild der Bevölkerung“ oder von „repräsentativen Stichproben“. Gemeint sind damit meist Quotenstichproben, bei denen für einige wenige Variable (Geschlecht, Alter, Bundesland,..) die Quoten so fixiert wurden, dass sie mit der Merkmalsverteilung in der Population übereinstimmen.

Eine Stichprobe „repräsentiert“ aber nie sämtliche Merkmalsverteilungen der Population. In der Statistik ist der Begriff der repräsentativen Stichprobe kein Fachbegriff.

Stichprobenumfang

In manchen Statistikbüchern gibt es Abhandlungen darüber, wie groß eine Stichprobe sein soll, damit die Genauigkeit einer Schätzung gewährleistet ist. Diese Abhandlungen gehen immer von sehr einfachen Annahmen aus; meist wird nur eine Variable untersucht, von der gewisse Kenntnisse (speziell über deren Varianz) vorausgesetzt werden. In der Praxis hat man es mit vielen Variablen zu tun, über die man kaum Angaben machen kann.

Natürlich sind Studien mit mehreren tausend Untersuchten vertrauenswürdiger als Studien mit einem viel kleineren Stichprobenumfang. Nur setzen Finanz-, Zeit- und andere Ressourcen gewisse Grenzen, über die die Größe der Stichprobe nicht hinausgehen kann. Daher meine Empfehlung: Sorgfalt beim Ziehen der Stichprobe und Erheben der Daten ist mehr wert als mit Ach und Weh den Stichprobenumfang um 10 Prozent zu erhöhen.

3.5.1 Zufallsauswahl

Bei diesem Auswahlverfahren werden die Probanden durch einen Zufallsmechanismus bestimmt. Auch hier unterscheidet man je nach Technik wieder:

- **Einfache Zufallsstichproben:** die Auswahlwahrscheinlichkeit jedes Elements der Population ist gleich und die Auswahl erfolgt direkt in einem einstufigen Auswahlvorgang. Man benötigt dazu allerdings Listen sämtlicher Elemente der Population (bei Telefoninterviews etwa Telefonbücher auf CD-ROM), die nicht immer verfügbar sind. Ein weiterer Nachteil aus statistischer Sicht ist, dass kleine Gruppen der Grundgesamtheit in der so ermittelten Stichprobe stark über- oder stark unterrepräsentiert sein können und damit die Stichprobenvarianz sehr hoch ist. Bei Handelsbetrieben schwankt der Umsatz stark. Einigen wenigen Großhandelsbetrieben stehen viele Klein Händler gegenüber. Es kann leicht passieren, dass in einer kleinen Stichprobe gar kein Großhandelsbetrieb oder überverhältnismäßig viele Großhandelsbetriebe aufscheinen.
- **Geschichtete Zufallsstichproben:** Die Grundgesamtheit wird in sich ausschließende Untergruppen (Schichten) aufgeteilt, in jeder Schicht wird eine einfache Zufallsstichprobe gezogen.
Häufig entsprechen die Stichprobengrößen in den einzelnen Schichten den Anteilen der Schichten in der Grundgesamtheit. Wenn im Beispiel mit den Handelsbetrieben 2% als Großhandelsbetriebe sind, sollte eine geschichtete Stichprobe von 400 Handelsbetrieben 8 zufällig ausgewählte Großhandelsbetriebe und 392 zufällig ausgewählte andere Handelsbetriebe enthalten.
- **Klumpenstichproben:** die Grundgesamtheit wird in eine Anzahl sich gegenseitig ausschließender Gruppen (in diesem Zusammenhang Klumpen oder Cluster) genannt eingeteilt. Per Zufallsauswahl wird eine Anzahl von Klumpen gezogen, in denen im einfachsten Fall alle Elemente in die Stichprobe aufgenommen werden (einstufige Klumpenauswahl), oder in jedem Klumpen wieder eine Zufallsauswahl durchgeführt wird (zweistufige Klumpenauswahl).
In einer Stadt gibt es kaum eine Liste mit den Benützern öffentlicher Verkehrsmittel, aber sicher mit der Liste aller Haushalte. Die Haushalte bilden in diesem Beispiel die Klumpen. Nach einer Zufallsauswahl von Haushalten werden in jedem Haushalt alle Benutzer öffentlicher Verkehrsmittel in die Stichprobe aufgenommen. Dieses Beispiel zeigt auch die

Vorteile der Klumpenmethode; Zeitaufwand und bei persönlichen Interviews auch Wegkosten lassen sich wesentlich verringern.

3.5.2 Quotenauswahl

In der kommerziellen Markt- und Meinungsforschung ist die Quotenauswahl sehr beliebt. Einige Merkmale (z.B. Geschlecht, Alter, etc.), deren Verteilung in der Grundgesamtheit bekannt ist, werden herangezogen, um eine Stichprobe so aufzubauen, dass die (univariate) Verteilung dieser Merkmale in der Stichprobe der in der Grundgesamtheit entspricht. Jeder einzelne Interviewer erhält dazu einen Quotenplan, auf dem die Anzahl der Interviews, die Quotenmerkmale und die Quoten pro Merkmal angegeben sind.

Die üblichsten Quotenmerkmale sind Geschlecht, Alter, Berufstätigkeit, Bundesland, Größe des Wohnortes. So könnte ein Quotenplan für einen Interviewer im Land Salzburg so ausschauen:

Interviews insgesamt		10
Geschlecht	Männlich	6
	Weiblich	4
Alter	16-29 Jahre	2
	30-44 Jahre	3
	45-59 Jahre	4
	60 und älter	1
Gemeindegröße	Unter 2000	4
	2000 – 5000	3
	5000 – 10000	3
Beruf	Landwirte	2
	Arbeiter	1
	Angestellte	2
	Beamte	1
	Pensionisten	2
	Schüler, Studenten	2

3.5.3 Willkürliche Auswahl

Durch Befragungen von Studenten in der Mensa, von Passanten in einer Fußgängerzone, durch TED-Umfragen oder Internet-Umfragen kommt man schnell zu umfangreichen Stichproben. Allerdings sind daraus abgeleitete Aussagen kaum vertrauenswürdig.

3.6 Erhebungsinstrument Fragebogen

Schriftliche Befragungen und mündliche Interviews bilden die hauptsächliche Datenquelle empirischer Untersuchungen. Ihnen liegt das Konzept des Fragebogens zugrunde, dessen Aufbau in diesem Abschnitt besprochen wird.

3.6.1 Fragestellungen

In einem Fragebogen werden Fragen verschiedener Art gestellt. Diese dienen zur Erhebung von:

- **Einstellungen:** Häufig werden Aussagen (Statements, Items) vorgelegt, die vom Befragten auf einer Ratingskala oder einer Likert-Skala (siehe weiter unten) eingestuft werden soll. Meist wird nicht eine Frage allein, sondern eine ganze Fragenbatterie zur Messung einer Einstellung verwendet.
- **Überzeugungen:** Mit einer Überzeugung sind subjektive Aussagen zu Fakten gemeint. „Was glauben Sie, wie viel Verkehrstote gibt es jährlich auf Österreichs Straßen?“
- **Verhalten:** Diese sind in der Regel Retrospektivfragen. Erfragt wird die Häufigkeit, Dauer und Art von Handlungen in der Vergangenheit. „Haben Sie in den letzten drei Monaten eine Städtereise in eine Stadt außerhalb Österreichs unternommen?“
- **Sozialstatistische Merkmale:** Erhoben werden die „sozialdemografischen“ Merkmale (Geschlecht, Alter, Einkommen, ..) der Befragten (Selbstauskünfte) aber auch von anderen Personen (Fremdauskünfte), etwa Ehepartner.

3.6.2 Fragetypen

Offene und geschlossene Fragen

Grundsätzlich werden offene und geschlossene Fragen unterschieden. Bei den offenen Fragen sind keine Antwortalternativen vorgegeben. Den Auskunftspersonen ist die Beantwortung vollkommen selbst überlassen.

Bei geschlossenen Fragen muss sich der Proband zwischen vorgegebenen Alternativen entscheiden. Dieser Fragetypus wird weit häufiger angewendet, weil durch die Kalkulierbarkeit der Antworten eine schnellere Datenauswertung möglich ist. Die Vollständigkeit der Antwortalternativen ist wünschenswert, aber nicht für jeden Fall möglich. Daher wird meist die Möglichkeit eingeräumt, die bestehenden Antworten durch den Probanden ergänzen zu lassen („weitere Nennung“). Diese Ergänzungen zu bearbeiten kann sehr zeitaufwändig sein.

Likert-Skala

Die Likert-Skala ist eines der gebräuchlichsten Formate zur Präsentation von Einstellungsfragen. Wenn von einer Likert-Skala die Rede ist, wird das zu beurteilende Item in Form eines Statements präsentiert, gefolgt von Antwortkategorien, die eine unterschiedlich starke Zustimmung zu dem Item ermöglichen. Die Anzahl der Skalenstufen kann variieren (nie mehr als 10), ebenso, ob ein neutraler Mittelpunkt aufgenommen wird (bei einer ungeraden Anzahl von Skalenstufen). Die Likert-Skala ist eine Ordinalskala.

Als Beispiel für eine Likert-Skala seien zwei Fragen zu den Stadtwerken einer Stadt angeführt:

Die Stadtwerke		trifft voll zu	trifft eher zu	weder noch	trifft eher nicht zu	trifft nicht zu
bieten alles aus einer Hand		1	2	3	4	5
sind ein sehr kundenfreundliches Unternehmen		1	2	3	4	5

Semantisches Differenzial

Bei dieser Technik werden die Befragten darum gebeten, ein Einstellungsobjekt mit einer Reihe Adjektiven zu beschreiben. Hier wird eine Liste von Adjektiven verwendet, die jeweils ein Gegensatzpaar bilden, etwa gut/schlecht, schön/hässlich. Jeweils ein solches Gegensatzpaar kann als eine einfache Skala aufgefasst werden.

Als Beispiel seien drei Adjektivpaare angeführt, die zur Beschreibung von Stadtwerken dienen können:

modern						traditionell
ehrllich						unehrllich
sympathisch						unsympathisch

Matrixfragen

Sehr oft kommt es vor, dass Fragestellungen eine größere Anzahl gleicher Antwortkategorien ermöglichen. Der Einfachheit und Übersichtlichkeit des Fragebogens halber können solche Fragen zu sog. Matrixfragen kombiniert werden. Diese Fragen können sowohl als Likert-Skala als auch als semantisches Differenzial vorliegen. Das obige Beispiel zur Likert-Skala könnte ebenso als Minibeispiel einer Matrixfrage aufgefasst werden (nur zwei Fragen) wie das Beispiel für das semantische Differenzial (nur drei Fragen). Meist sind aber weit mehr Fragen in einer solchen Fragebatterie zusammen gefasst.

Die einfache Auswertung von Matrixfragen über Mittelwerte kann zu sog. Polaritätsprofilen führen.

Filterfragen

In Fragebögen kommt es vor, dass manche Fragen nur von bestimmten Personen beantwortet werden dürfen, oder dass Fragen von der Antwort einer oder mehrerer davor gestellter Fragen,

den sog. Filterfragen, abhängig sind. Bei schriftlichen Befragungen sollten daher abhängige Fragen deutlich von den Filterfragen unterschieden werden. Einen Fragenblock „Ehezufriedenheit“ passieren nur verheiratete Personen. Die Filterfrage wäre der Familienstand.

Gabelfragen

Eine Erweiterung der Filterfrage ist die Gabel. Im Anschluss an die Frage nach dem Familienstand werden ledige Personen mit Fragenblock A, verheiratete Personen mit Fragenblock B, eventuell geschiedene oder verwitwete Personen mit Frageblöcken C oder D konfrontiert.

3.6.3 Richtlinien für die Fragenformulierung

Grobe Fehler in einem Fragebogen kann man durch Einhalten einiger Richtlinien vermeiden:

- **Kurz, verständlich und hinreichend präzise:** Fragen sollten kurz, verständlich, mit einfachen Worten und hinreichend präzise formuliert sein. Sie sollten nicht gestelzt klingen, und es sollten Fremdworte vermieden werden, die in der Zielgruppe nicht allgemein üblich sind.
- **Keine platten Anbieterungen:** Fragen sollten in einfachem Hochdeutsch ohne bürokratische Verrenkungen gestellt werden. Kumpelhafte Anbieterungen oder Subkultur-Formulierungen sind meist lächerlich und sollten vermieden werden.
- **Keine doppelte Verneinung:** Im Ablauf eines persönlichen oder telefonischen Interviews werden ziemlich viele Fragen in kurzer Zeit gestellt. Längeres Nachdenken über eine gestellte Frage ist meist nicht möglich. Daher sollte man Fragen nicht unnötigerweise verkomplizieren. „Sind Sie gegen ein Verbot der Gentechnik in der Landwirtschaft?“
- **Antwortkategorien:** Die Antwortkategorien von geschlossenen Fragen sollen disjunkt (nicht überlappend), erschöpfend und präzise sein. Je nach Zielsetzung der Untersuchung sollten sie hinreichend genau zwischen verschiedenen Sachverhalten diskriminieren können. Eine Frage nach dem Einkommen mit nur drei Kategorien wird kaum eine gute Schätzung des Einkommens erlauben. Mit der Wahl der Kategorien wird auch eine Vorentscheidung über das Messniveau der Variablen getroffen und damit über die anwendbaren statistischen Analyseverfahren getroffen.
- **Vorsicht bei stark wertbesetzten Begriffen:** Begriffe wie „Gerechtigkeit“, „Freiheit“, „Verbrechen“ haben stark positiven oder negativen Beigeschmack. Unabhängig von der Fragestellung kann allein die Verwendung solcher Begriffe die Antwortreaktion in die eine oder andere Richtung lenken. Kaum jemand wird sich gegen Gerechtigkeit oder für Verbrechen äußern.
- **Keine mehrdimensionalen Fragen:** Antworten auf mehrdimensionale Fragen sind nicht eindeutig einer Zieldimension zurechenbar. Die Frage „Kernkraftwerke verringern die Kosten der Stromerzeugung, stellen aber ein Sicherheitsrisiko dar“ wird besser durch zwei Einzelfragen „Kernkraftwerke verringern die Kosten der Stromerzeugung“ und „Kernkraftwerke sind ein Sicherheitsrisiko“ ersetzt.
- **Keine Suggestivfragen:** Die Frage soll die Befragten auffordern, das zu sagen, was sie für richtig halten. Die Fragen sollten so formuliert sein, dass sie der Auskunftsperson keine bestimmten Antworten besonders nahe legen („Sie sind sicher auch der Meinung, dass ..“)
- **Keine Überforderung der Befragten:** Die Frage „Wieviel Prozent Ihres Einkommens geben Sie für Versicherungen aus?“ klingt harmlos, verlangt vom Befragten aber eine gewaltige Rechenleistung.

- **Vorsicht bei Matrixfragen:** Man weiß aus Experimenten, dass Personen Fragen eher bejahen als verneinen. Sind in einer Fragenbatterie (Matrixfragen) alle Fragen gleich gepolt, erhält man leicht ein falsches Bild.

3.6.4 Pretests des Fragebogens

Ein neuer Fragebogen sollte einem oder mehreren Pretests unterzogen werden. Im Pretest sollten die Befragten ermuntert werden, weniger verständliche Fragen zu kritisieren. Die Interviewer sollen ihre Einschätzungen von Fragebogen und Interviewverlauf protokollieren.

Zweck von Pretests ist:

- die Ermittlung der durchschnittlichen Befragungszeit,
- die Prüfung der Verständlichkeit von Fragen,
- die Prüfung der Vollständigkeit von Antwortkategorien,
- die Prüfung, ob die Reihenfolge der Fragen gut gewählt wurde,
- eine erste Schulung der Interviewer.

Als Folge des Pretests kann der Fragebogen leicht umformuliert, gekürzt aber auch völlig neu entworfen werden..

3.7 Fragen

Frau Maier hat mit ihrer Abteilung trotz mancher Schwierigkeiten ein Projekt erfolgreich abgeschlossen. Zum Teil als Belohnung zum Teil als Fortbildung kann sie an einem einwöchigen Seminar über Mitarbeiterführung teilnehmen. Sie interessiert sich, ob sie aus dem Seminar etwas für ihre alltägliche Arbeit mitnehmen kann. Speziell interessiert es sie, ob das Betriebsklima in ihrer besser geworden ist.

- 1) Was könnte alles zur Operationalisierung des Begriffs „Betriebsklima“ verwendet werden?
- 2) Wenn zur Operationalisierung des Begriffs „Betriebsklima“ die Einschätzung der Vorgesetzten von Frau Maier („sehr gut“, „gut“, „schlecht“, „sehr schlecht“), die Einschätzung der Mitarbeiter selbst und die Anzahl der Krankenstandstage der einzelnen Mitarbeiter verwendet werden, wie steht es bei diesen Messungen um die Gütekriterien Objektivität, Reliabilität, Validität?
- 3) Welches Skalenniveau haben obige Messungen?
- 4) Wenn nach dem Seminar 19 von 20 Mitarbeitern der Abteilung von Frau Maier von einem guten bis sehr guten Betriebsklima sprechen, kann man daraus folgern, dass das Seminar ein Erfolg war?
- 5) Wenn nach dem Seminar 19 von 20 Mitarbeitern der Abteilung von Frau Maier von einem guten bis sehr guten Betriebsklima sprechen, während es vor dem Seminar nur 16 von 20 waren, kann man daraus folgern, dass das Seminar ein Erfolg war?
- 6) Wenn vor dem Seminar fünf Mitarbeiter über das Betriebsklima befragt werden, fünf andere Mitarbeiter nach dem Seminar, handelt es sich um eine Querschnitts-, Trend- oder

Paneluntersuchung? Was wäre ein Paneldesign mit Stichprobenerhebung, was eine Vollerhebung?

- 7) Wenn vor dem Seminar fünf Mitarbeiter über das Betriebsklima befragt werden sollen, wie könnte eine Zufallsstichprobe dazu gezogen werden?
- 8) Wenn zur Messung des Betriebsklimas die Einschätzung der Vorgesetzten von Frau Maier („sehr gut“, „gut“, „schlecht“, „sehr schlecht“) erhoben wird, ist das mit einer Likert-Skala oder einem semantischen Differenzial möglich?
- 9) Wiederholen Sie: Randomisierung, Quasi-Experiment, offene und geschlossene Fragen, Matrixfragen, Richtlinien für die Fragenformulierung.

Mögliche Antworten:

- 1) Was könnte alles zur Operationalisierung des Begriffs „Betriebsklima“ verwendet werden?

„Betriebsklima“ ist natürlich ein mehrdimensionaler Begriff. Die Geschäftsleitung versteht unter Betriebsklima vermutlich nicht genau das, was einfache Angestellte darunter verstehen. Fragen über den Umgang untereinander, über die Art, wie Aufgaben verteilt und bearbeitet werden, wie Konflikte (viele Ursachen) gelöst oder auch nicht gelöst werden, fallen darunter. Bei all diesen Fragen ist auch zu entscheiden, ob eine Einschätzung von außen, besser als eine von innen ist.

- 2) Wenn zur Operationalisierung des Begriffs „Betriebsklima“ die Einschätzung der Vorgesetzten von Frau Maier („sehr gut“, „gut“, „schlecht“, „sehr schlecht“), die Einschätzung der Mitarbeiter selbst und die Anzahl der Krankenstandstage der einzelnen Mitarbeiter verwendet werden, wie steht es bei diesen Messungen um die Gütekriterien Objektivität, Reliabilität, Validität?

Einschätzung der Vorgesetzten: es besteht die Gefahr (wie bei fast allen Einschätzungen), dass die Objektivität nicht gegeben ist. Reliabilität ist gegeben (einen Tag später wird die Einschätzung von außen vermutlich gleich ausfallen). Validität ist fraglich, da mit einer einzelnen Frage ein sehr komplexes Gebiet schwer erfasst werden kann.

Einschätzung der Mitarbeiter: es besteht noch mehr die Gefahr, dass die Objektivität nicht gegeben ist. Reliabilität ist weniger klar als bei der Einschätzung von außen. Validität ist fraglich, da mit einer einzelnen Frage ein sehr komplexes Gebiet schwer erfasst werden kann. Anzahl der Krankenstandstage: ist objektiv und reliabel, valide ist diese Messung kaum.

- 3) Welches Skalenniveau haben obige Messungen?

Einschätzung der Vorgesetzten: Ordinalskala

Einschätzung der Mitarbeiter: Ordinalskala

Anzahl der Krankenstandstage: Absolutskala

- 4) Wenn nach dem Seminar 19 von 20 Mitarbeitern der Abteilung von Frau Maier von einem guten bis sehr guten Betriebsklima sprechen, kann man daraus folgern, dass das Seminar ein Erfolg war?

Nein. Vielleicht war die Einschätzung vorher schon sehr gut. Wenn man eine Veränderung messen will, benötigt man eine Messung vor (Pretest) und eine Messung nach (Posttest) einem Ereignis.

- 5) Wenn nach dem Seminar 19 von 20 Mitarbeitern der Abteilung von Frau Maier von einem guten bis sehr guten Betriebsklima sprechen, während es vor dem Seminar nur 16 von 20 waren, kann man daraus folgern, dass das Seminar ein Erfolg war?

Kaum. Es ist zwar zu einer Veränderung gekommen, man kann diese Veränderung aber nicht mit Sicherheit als Auswirkung des Seminars bezeichnen. Vor dem Seminar musste ein Projekt abgeschlossen werden, Stress und Gereiztheit sind keine Grundlagen eines guten Betriebsklimas. Bei einem Experiment gibt es daher immer eine Versuchsgruppe und eine Kontrollgruppe (fehlt hier).

- 6) Wenn vor dem Seminar fünf Mitarbeiter über das Betriebsklima befragt werden, fünf andere Mitarbeiter nach dem Seminar, handelt es sich um eine Querschnitts-, Trend- oder Paneluntersuchung? Was wäre ein Paneldesign mit Stichprobenerhebung, was eine mit Vollerhebung?

Das ist eine Trenduntersuchung, bestehend aus zwei Querschnittserhebungen (vor und nach dem Seminar). Bei einem Paneldesign werden dieselben Mitarbeiter, die vor dem Seminar befragt wurden, auch nach dem Seminar befragt. Bei einer Vollerhebung werden alle Mitarbeiter, bei einer Stichprobenerhebung nur ein Teil davon befragt.

- 7) Wenn vor dem Seminar fünf Mitarbeiter über das Betriebsklima befragt werden sollen, wie könnte eine Zufallsstichprobe dazu gezogen werden?

Am einfachsten durch Durchnummerieren der Mitarbeiter, dann mittels Zufallszahlen (Zufallszahlengenerator, Zufallszahlentabellen) Auswahl derjenigen, die in die Stichprobe fallen.

- 8) Wenn zur Messung des Betriebsklimas die Einschätzung der Vorgesetzten von Frau Maier („sehr gut“, „gut“, „schlecht“, „sehr schlecht“) erhoben wird, ist das mit einer Likert-Skala oder einem semantischen Differenzial möglich?

So wie hier formuliert, ist es ein klassisches Beispiel für eine Likert-Skala.

4 Datenerhebung

In diesem Kapitel werden mehrere Formen der Datenerhebung, deren Vor- und Nachteile vorgestellt. Befragung und Beobachtung sind reaktive Verfahren, es besteht die Möglichkeit, dass die Datenerhebung die Untersuchungsobjekte beeinflusst. Bei der Inhaltsanalyse und der Verwendung bestehender Daten gibt es diese Gefahr nicht, man bezeichnet sie daher auch als nichtreaktive Verfahren.

4.1 Befragung

Die in den Sozialwissenschaften am häufigsten angewandte Methode, Daten zu erheben, ist die Befragung. Je nach Art, wie diese Befragung erfolgt, unterscheiden wir:

- Schriftliche Befragung
- Mündliche Befragung durch Face-to-Face Interviews
- Telefoninterviews

4.1.1 Schriftliche Befragung

Bei einer schriftlichen Befragung führt die Auskunftsperson schriftlich vorgelegte Fragen selbstständig aus. Dies kann sowohl mit als auch ohne persönlichen Interviewerkontakt erfolgen. Wichtig ist die Bereitstellung einer Anleitung (nicht zu umfangreich), wie bestimmte Fragen auszufüllen sind oder wie einige Begriffe aufzufassen sind.

Bei einer **postalischen Befragung (Mailing)** wird ein Fragebogen mit einem Begleitbrief versandt. Der Brief muss eine Erklärung des Fragebogens und ein frankiertes Rückantwortkuvert mit Adresse beinhalten.

Immer öfter wird auch die Möglichkeit eines **elektronischen Fragebogens** genutzt. Die Auskunftsperson erhält den Fragebogen per Mail, Diskette oder CD-ROM, füllt den Fragebogen aus und schickt ihn retour. Zusätzlich gibt es die Möglichkeit, dass Fragebögen auf Homepages eingerichtet sind.

Vorteile schriftlicher Befragungen, die ohne Beisein eines Interviewers erfolgen:

- die einfachste Möglichkeit, Auskunftspersonen zu erreichen
- einfache Stichprobenziehung, sofern vollständige Adresslisten verfügbar sind
- Merkmale und Verhalten von Interviewern haben keinen Einfluss
- die Auskunftsperson füllt den Fragebogen in ihrer gewohnten Umgebung aus
- die Anonymität der Auskunftsperson ist gewährleistet, dadurch sind auch persönliche und heikle Fragen möglich
- die Kosten sind verhältnismäßig gering

Als **Probleme und Nachteile** der schriftlichen Befragung gelten:

- bei Verständnisproblemen erfolgt keine Hilfe durch den Interviewer, komplexe oder mehrdeutige Fragestellungen sind daher kaum möglich
- die Erhebungssituation ist unkontrolliert, es ist nicht klar, ob die angeschriebene Person, den Fragebogen ausfüllt
- ohne Begleitmaßnahmen (Erinnerungsschreiben, Teilnahme an einer Preisverlosung) ist die Rücklaufquote gering

4.1.2 Face-to-Face Interviews

Hier werden der Auskunftsperson die Fragen mündlich durch den Interviewer gestellt, man spricht von **Face-to-Face Interviews** oder auch von **persönlichen Interviews**. An diesen müssen besondere Anforderungen gestellt werden:

- Er muss ein entsprechendes Auftreten und äußeres Erscheinungsbild aufweisen.
- Er muss über eine hohe Anpassungsfähigkeit verfügen, um eine entspannte Gesprächsatmosphäre herstellen und aufrecht erhalten zu können.
- Er muss das Verhalten anderer aufmerksam beobachten und verstehen können.
- Er muss über das Befragungsthema ausreichend informiert und mit dem Fragebogen vertraut sein.
- Er darf die Antworten des Befragten nicht durch eigene Urteile und Bewertungen beeinflussen.
- Er muss offene Fragen exakt und unverfälscht notieren.

Ein besonderes Kriterium ist daher die Schulung der Interviewer.

Vorteile von Face-to-Face Interviews:

- geeignetste Methode, um ganz bestimmte Personen bzw. Zielgruppen zu erreichen
- ermöglicht den Einsatz von komplexen Fragestellungen, offenen Fragen, ..
- geringe Anzahl von Antwortverweigerungen
- mehrdeutige Fragen können erklärt werden
- visuelle Hilfsmittel (Skalen, Bilder,..) können der Auskunftsperson vorgelegt werden

Nachteile von Face-to-Face Interviews:

- Erreichbarkeit der Auskunftsperson
- Aufwand, nicht erreichte Personen erneut zu kontaktieren („Not-at-Homes“)
- regionale Streuung der Interviews schwierig und teuer
- Gefahr sozial erwünschter Antworten
- Interviewereinfluss
- ordnungsgemäße Durchführung der Interviews erfordert eine Kontrolle der Interviews

Bei kleineren Studien ist auch zu überlegen, ob nicht das Forschungsteam besser selbst die Umfrage durchführt und die Organisation der sog. Feldarbeit selbst in der Hand behält.

4.1.3 Telefoninterviews

Zunehmend beliebter wird die Befragung über **Telefoninterviews**. Besonders geeignet sind einfache, kurze Befragungen, die keine besondere Motivation der Befragten voraussetzen und keinerlei visuelle Unterstützung benötigen. Methodische Forschung zum Vergleich persönlicher und telefonischer Interviews bezüglich Datenqualität und Ausschöpfungsquoten hat ältere Vorurteile zum Anwendungsbereich und der Qualität telefonischer Befragungsdaten eindeutig widerlegt.

Vorteile von Telefoninterviews:

- verhältnismäßig geringe Kosten
- große Streuung der Interviews möglich
- hohe Antwortrate (besonders im Vergleich zu Mailings)
- schnelle Art der Datenerhebung
- einfacheres Handling im Falle von „Not-at-Homes“
- einfachere Interviewerkontrolle

Nachteile von Telefoninterviews:

- visuelle Hilfsmittel können nicht vorgelegt werden
- begrenzte Interviewdauer
- Vertrauensbasis ist schwieriger herstellbar als bei Face-to-Face Interviews
- Interviewereinfluss (allerdings geringer als bei Face-to-Face Interviews)
- schwer kontrollierbar, wer die Antworten am Telefon erteilt

4.2 Beobachtung (Feldforschung)

In einem allgemeinen Sinn sind sämtliche empirische Methoden Beobachtungsverfahren. Durch Beobachtung ermittelt man die Position eines Zeigers an einem Messgerät oder die Stelle, an der die Ratingskala in einem schriftlichen Interview angekreuzt wurde. Ist jedoch von der Erhebungsmethode der **Beobachtung** in der Sozialforschung die Rede, so wird dadurch die direkte Beobachtung menschlicher Handlungen, sprachlicher Äußerungen, nonverbaler Reaktionen (Mimik, Gestik, Körpersprache) und anderer sozialer Merkmale (Kleidung, Symbole, Gebräuche, etc.) verstanden.

Aus der Ethnologie kommend, wird diese Methode auch **Feldforschung** genannt. Wohl die bekannteste soziologische Studie, die in Österreich durchgeführt wurde, „Die Arbeitslosen von Marienthal“ von Marie Jahoda, Paul Lazarsfeld und Hans Zeisel (1975, Erstausgabe 1933), beruht zu einem Großteil auf Beobachtung. In dem kleinen niederösterreichischen Ort Marienthal wurde durch das Schließen einer Textilfabrik 1930 ein Großteil der dort ansässigen Bevölkerung arbeitslos. Das Forschungsteam um Jahoda, Lazarsfeld und Zeisel setzte eine Vielzahl von unterschiedlichen Methoden ein, nur ein kleiner Teil war eine klassische Fragebogenstudie, weit mehr waren unterschiedliche Beobachtungsstudien.

Zur Anwendung kommt Feldforschung vor allem in Fällen, wo Befragung nicht möglich ist, weil der Zugang zu den Untersuchungspersonen fehlt (Sekten, Rechtsradikale ..), wo soziale Prozesse, die über einen bestimmten Zeitrahmen hinweg stattfinden, erklärt werden sollen oder dort, wo zu wenig Wissen über die zu untersuchende Gruppe bzw. den Untersuchungsgegenstand besteht, um einen brauchbaren Fragebogen zu entwerfen.

Feldforschung darf man nicht mit Datenerheben in der konventionellen Form mit Fragebogen verwechseln, die manchmal als „Feldarbeit“ bezeichnet wird. Ebenso falsch ist es aber auch, Feldforschung als eine „nur qualitative“ Methode abzuwerten, bei der keine quantitative Analyse von Daten möglich ist. Eine, meist nachträgliche, Protokollierung der Beobachtungsergebnisse macht auch diese Methode einer statistischen Datenanalyse zugänglich.

Von den Untersuchungszielen und dem Untersuchungsgegenstand hängen die Methoden ab, die eingesetzt werden. Eine kurze Beschreibung dieser Methoden und mit ihnen einhergehender Probleme folgt jetzt.

4.2.1 Teilnehmende versus nichtteilnehmende Beobachtung

Die **teilnehmende Beobachtung** bietet sich an, wenn der Beobachter eine klare Rolle im sozialen Feld übernehmen kann. Sollen Einkaufsfahrten untersucht werden, so ist die Teilnahme an einer oder mehreren solcher Einkaufsfahrten sicher wertvoll. Sind die Arbeitsbedingungen von Taxilenkern Untersuchungsgegenstand, so ist es der Untersuchung dienlich, wenn der Beobachter selbst eine Taxilizenz erwirbt, um auch persönliche Erfahrungen sammeln zu können. Eine Gefahr der Teilnahme des Beobachters ist die Beeinflussung des sozialen Geschehens in Richtung auf eine Bestätigung der zu prüfenden Hypothesen.

Die **nichtteilnehmende Beobachtung** weist den Vorzug auf, dass der Beobachter nicht gleichzeitig zwei Dinge tun muss: im Feld interagieren und sich gleichzeitig auf die Beobachtung des sozialen Geschehens konzentrieren. Er kann die Beobachtungen direkt protokollieren, entweder als Notiz oder mittels eines strukturierten Beobachtungsschemas.

4.2.2 Offene versus verdeckte Beobachtung

Die teilnehmende wie auch die nichtteilnehmende Beobachtung kann verdeckt oder offen erfolgen. Bei der **verdeckt teilnehmenden** Beobachtung gibt sich der Beobachter seinen Interaktionspartnern nicht als solcher zu erkennen (Undercover-Methode). Bei der **verdeckt nichtteilnehmenden** Beobachtung wird der Beobachter bestrebt sein, von den untersuchten Personen unbemerkt zu bleiben (Schlüssellochmethode). Die verdeckte Beobachtung hat den Vorteil, dass sie nicht reaktiv ist, d.h. die untersuchten Personen ändern ihr Verhalten nicht deshalb, um (z.B.) in einem günstigeren Licht zu erscheinen. Einige Sozialforscher lehnen die verdeckte Beobachtung als prinzipiell unethisch ab.

Bei der **offenen Beobachtung** kennen die beobachteten Personen die Rolle des Beobachters, es besteht die Gefahr, dass seine Anwesenheit das zu beobachtende Geschehen beeinflusst. So ist im Rahmen einer Untersuchung zu den Arbeitsbedingungen von Saisonarbeitern in der

Tourismuswirtschaft eine verdeckt teilnehmende Beobachtung durch ein Mitglied des Untersuchungsteams besser als eine offene, da sonst sowohl Arbeitskollegen als auch die Leitung des Tourismusbetriebs ihr Verhalten ändern könnten.

4.2.3 Feldbeobachtung versus Beobachtung im Labor

Ein Vorteil der Feldforschung gegenüber Experimenten ist die Tatsache, dass die Beobachtung im gewohnten Umfeld stattfindet und nicht in einem künstlichen Umfeld (Labor). Dem gegenüber können in einem Experiment experimentelle Stimuli gezielt gesetzt und Störfaktoren kontrolliert werden.

Ein Mittelweg zwischen Experiment und Feldbeobachtung ist das **Feldexperiment**. Anstatt in einem Labor werden in natürlicher Umgebung experimentelle Untersuchungen durchgeführt. Randomisierung, also die zufällige Zuteilung zu Versuchs- oder Kontrollgruppe, kann nicht immer gewährleistet werden, ebenso wie die Kontrolle der Störvariablen. Die Datenerhebung erfolgt meist verdeckt nichtteilnehmend, also nicht-reaktiv.

Beispiel Hilfeleistung: Die Hypothese lautet: Personen in guter Stimmung werden anderen Menschen eher Hilfe leisten als Personen in schlechter Stimmung. Zur Beeinflussung der Stimmung erhält die Versuchsperson unerwartet eine kleine Belohnung. Der Aufbau des Feldexperiments ist der folgende: Im Rückgabeschacht eines öffentlichen Telefons wird in der Versuchsgruppe eine 10-Cent-Münze deponiert. Nach einem Telefongespräch greifen fast alle Menschen sicherheitshalber in den Rückgabeschacht. Die Wahrscheinlichkeit ist somit recht hoch, dass eine Versuchsperson die deponierte Münze findet (und damit kurzfristig in guter Stimmung ist). Sobald eine Person die Telefonzelle verlässt, lässt eine Person vor der Telefonzelle einen Stapel mit Akten und Papieren fallen. Beobachtet wird, ob die Versuchsperson zur Hilfeleistung bereit war. Bei einer Kontrollgruppe werden keine Münzen im Telefonapparat deponiert; die Stimmung der Versuchspersonen wird also nicht experimentell gehoben.

4.2.4 Unstrukturierte versus strukturierte Beobachtung

Mit einem **strukturierten Beobachtungsschema** wird versucht, die Objektivität und Zuverlässigkeit der Beobachtung zu erhöhen. Man begegnet damit der Gefahr, durch selektive Wahrnehmung eine Verzerrung der Beobachtung zu erhalten; gerade in der Feldforschung ist diese Gefahr nicht zu unterschätzen. Es kann ein einheitliches Schema für das Beobachtungsprotokoll ausgearbeitet werden, das ähnlich einem Fragebogen bei der Befragung, alle wichtigen Punkte der Beobachtung enthält. Für eine spätere Datenauswertung ist dies ein wichtiges Hilfsmittel.

Andererseits ist ein Vorteil der Feldforschung ihre Flexibilität. Das Studiendesign kann – ohne große Kosten – immer wieder neu gestaltet werden, die Beobachtung neu strukturiert werden. Die Gefahr besteht, dass am Ende nur ein Datensalat vorhanden ist, sehr viele Variablen, die im Laufe der Beobachtungsphase zusätzlich aufgenommen wurden, aber kaum Beobachtungen, bei denen allen diesen Variablen auch Werte zugewiesen sind. Änderungen des Studiendesigns sollen daher sorgfältig bedacht und nur in Absprache mit dem gesamten Forschungsteam unternommen werden.

4.3 Nichtreaktive Verfahren

4.3.1 Inhaltsanalyse

Bei der Inhaltsanalyse werden Bücher, Zeitungen, Magazine, TV- und Rundfunksendungen, Gesetzestexte, Plakate, etc. auf bestimmte Inhalte hin untersucht. Im einfachsten Fall, der sog. **Frequenzanalyse**, wird nur gezählt, wie oft ein bestimmter Begriff auftritt. In einer erweiterten Form, der sog. **Kontingenzanalyse**, wird gezählt, wie oft der Begriff in Zusammenhang mit bestimmten anderen Begriffen auftaucht. In beiden Fällen ist die Kodierung nicht sehr schwierig, sie geschieht über Häufigkeitstabellen (Strichlisten). Weit schwieriger ist das Entdecken **latenter Inhalte**, die aus dem Sprach- oder Schreibstil abgeleitet werden. Diese sog. **qualitative Analyse** ist von der Idee her sehr interessant, ihre Messungen genügen aber nicht immer den Kriterien der Objektivität, Reliabilität und Validität.

In der schon erwähnten Studie „Die Arbeitslosen von Marienthal“ von Jahoda, Lazarsfeld und Zeisel wurden in einer Inhaltsanalyse Schulaufsätze zum Thema „Weihnachtswünsche“ einer formalen Inhaltsanalyse unterzogen. Es zeigte sich, dass die Kinder der arbeitslosen Marienthaler häufiger den Konjunktiv wählten als Schulkinder aus Nachbarorten.

Beispiel: Eine Untersuchung zur Akzeptanz von Festspielen soll durchgeführt werden. Dabei soll auch der Begriff „Umwegrentabilität“ ausgeleuchtet werden.

Eine Frequenzanalyse könnte darin bestehen, für einen bestimmten Zeitraum die Tageszeitungen nach dem Begriff Umwegrentabilität zu durchforsten. Man zählt wie oft dieses Wort in Artikeln zu den Festspielen auftaucht.

Will man hingegen Umwegrentabilität und gleichzeitig Finanzierung von Festspielen untersuchen, bietet sich eine Kontingenzanalyse an. Der Begriff „Finanzierung“ ist allerdings noch genauer zu konkretisieren.

Eine qualitative Analyse könnte im Auswerten der Artikel auf die Frage hin bestehen, wo die verantwortlichen Festspielleiter darüber klagen, dass der finanzielle Druck immer härter wird und die Umwegrentabilität als Argument für Subventionsgeber verwendet wird.

Vorteile der Inhaltsanalyse:

- relativ geringer zeitlicher und finanzieller Aufwand
- es können längere Zeitspannen (Jahre bis Jahrzehnte) analysiert werden
- im Gegensatz zu misslungenen Experimenten, Befragungen oder Beobachtungen, die meist gänzlich wiederholt werden müssen, muss bei der Inhaltsanalyse in der Regel nur ein Teil der Studie neu durchgeführt werden
- der Forscher übt keinen Einfluss auf die Untersuchungsobjekte aus

Nachteile der Inhaltsanalyse:

- es können nur aufgezeichnete Inhalte analysiert werden
- Gütekriterien der Messung bei der qualitativen Analyse nicht immer erfüllt

4.3.2 Verwendung bestehender Daten

Dabei ist nicht an die Sekundäranalyse eines Datensatzes eines anderen Projektes gedacht, sondern als zusätzliche Datenquelle, um sich etwa im Vorfeld einer Studie Informationen und Input für die Fragebogenentwicklung zu besorgen. Allerdings liegen Daten selten in der gewünschten Form vor.

Ebenso in diese Kategorie fällt die Auswertung von Datenspuren, die Benutzer des Internets hinterlassen, wenn etwa der Kauf eines Produktes oder die Buchung einer Reise elektronisch durchgeführt wird.

4.4 Fragen

- 1) Wenn eine Studie zum Verhalten von Fußballfans bei Auswärtsspielen ihrer Mannschaft geplant ist:
Warum sollte eine Beobachtung neben oder statt einer Befragung erfolgen?
Was ist eine teilnehmende Beobachtung?
Was ist eine verdeckte, was eine offene Beobachtung?
- 2) Welche Schwierigkeiten hätte man bei obiger Studie (Fußballfans) mit schriftlichen Befragungen, welche bei Telefoninterviews?
- 3) Welche Schwierigkeiten hätte man bei obiger Studie (Fußballfans) mit persönlichen Interviews?
- 4) Wiederholen Sie Vor- und Nachteile der einzelnen Datenerhebungsformen.

Mögliche Antworten:

- 1) Wenn eine Studie zum Verhalten von Fußballfans bei Auswärtsspielen ihrer Mannschaft geplant ist:
Warum sollte eine Beobachtung neben oder statt einer Befragung erfolgen?
Was ist eine teilnehmende Beobachtung?
Was ist eine verdeckte, was eine offene Beobachtung?

Es ist anzunehmen, dass eine Beobachtung neue Erkenntnisse bringt. Selbst wenn ein Mitglied des Forschungsteams ein Mitglied eines Fanclubs ist und somit schon einiges über die Rituale solcher Gruppen weiß, wird eine Beobachtung nicht nur einen Fanclub sondern mehrere umfassen und daher weiteren Einblick gewähren.

Eine teilnehmende Beobachtung ist in diesem Beispiel die Teilnahme an einer Schlachtenbummlerfahrt zu einem Auswärtsspiel.

Verdeckt ist die Beobachtung, wenn sich der Beobachter nicht als Beobachter zu erkennen gibt (Tarnung als Fan). Offen ist die Beobachtung, wenn er sich dem Fanclub (oder zumindest dessen Leitwölfen) als Beobachter zu erkennen gibt (und als solcher akzeptiert wird).

- 2)** Welche Schwierigkeiten hätte man bei obiger Studie (Fußballfans) mit schriftlichen Befragungen, welche bei Telefoninterviews?

Die erste und sicher eine große Schwierigkeit ist, an Adress- und Telefonlisten der Fans zu gelangen. Eine Möglichkeit wäre, über den Verein in Kontakt mit den Leitungen der Fanclubs zu treten und von diesen Listen zu erhalten. Ob diese aktuell und vollständig sind, ist eher zu bezweifeln.

- 3)** Welche Schwierigkeiten hätte man bei obiger Studie (Fußballfans) mit persönlichen Interviews?

Man könnte sicher im Schneeballsystem (man wird von einem interviewten Fan zum nächsten verwiesen) eine genügend große Stichprobe befragen. Das Problem ist hierbei die Repräsentativität, da es sicher keine Zufallsstichprobe ist.

5 Datenauswertung

5.1 Datenmatrix, Fehlerkontrolle und Fehlerbereinigung

Schon bei moderatem Datenumfang ist an eine händische Auswertung von Daten nicht zu denken. Für die Auswertung per Computer müssen die Daten in einem Datenfile vorliegen. Wenn nicht schon bei der Erhebung die Daten in einem solchen File angelegt wurden, ist die Übertragung dieser Informationen, ob aus Beobachtungsprotokollen, Fragebögen oder aus Kodierbögen der Inhaltsanalyse, in ein Datenfile notwendig. Die allgemein übliche Form eines solchen Files ist eine **Datenmatrix**. Das ist ein Rechteckschema, bei dem die Zeilen den Beobachtungseinheiten und die Spalten den Variablen entsprechen. In einem Spreadsheet könnten Daten etwa folgendermaßen abgelegt sein:

Fragebogen-ID	Sex	Einkommen	Wellness	Kultur	Sport
0039	1	21	4	3	1
0077	2	29	3	2	4
0079	2	43	4	3	3
0124	2	34	2		5
0155	2	31	3	3	3
0176	1	25	8	4	2
0275	1		4	2	3
0314	3	38	2	2	4
0317	11	34	3	3	3
0422	2	28000	1	4	4
0569	1	29	3	2	4
0810	1		2	0	4
0925	2	52	3	4	3
1182	1	45	5	3	4

Es gibt also fünf eigentliche Variablen: Sex, Einkommen, Wellness, Kultur und Sport, die in den Spalten 2 bis 6 stehen. Die Spalte 1 dient zur Kennzeichnung des Fragebogens. Dies ist von Bedeutung, um etwaigen Unklarheiten, die später entdeckt werden, im Originalfragebogen nachzugehen. Stellen wir uns vor, dass das Geschlecht mit 1 für Frau und 2 für Mann kodiert wurde, das Einkommen als Jahreseinkommen in 1000 Euro und die Angaben bei Wellness, Kultur und Sport das Interesse für diese Gebiete im Urlaub, gemessen auf einer Likert-Skala von 1 bis 5 bedeuten. Wie werden solche Daten kontrolliert?

Wenn die Daten als Datenmatrix vorliegen, ist die **Fehlerkontrolle**, also die Kontrolle der Daten auf Fehler, sehr erleichtert. Spaltenweise (also Variable nach Variable) wird die Datenmatrix untersucht, folgende Fälle sollten dabei entdeckt werden:

- **„Wild codes“:** Werte, die als Variablenwerte nicht denkbar sind, aber dennoch im Datenfile stehen. Häufigster Grund hierfür sind Tippfehler.
Im obigen Beispiel kommen als Werte der Variablen Sex auch die Zahlen 3 und 11 vor. Vermutlich ist 11 statt 1 eingetippt worden. Wofür 3 steht ist nicht klar. Bei Wellness kommt einmal der Wert 8 vor, bei Kultur der Wert 0. Da beide Variablen auf einer Likert-Skala mit Werten von 1 bis 5 gemessen wurden, sind beide Werte wild codes. Eine Kontrolle der Originalfragebögen ist notwendig.
- **„Missing values“:** es kommt immer wieder vor, dass bei einzelnen Beobachtungen nicht sämtliche Variablenwerte vorliegen. Das kann daran liegen, dass auf bestimmte Fragen die Auskunft verweigert wird, aber auch möglicherweise daran, dass bei der Dateneingabe ein Fehler passiert ist.
Im obigen Beispiel fehlen zwei Werte beim Einkommen und ein Wert bei Kultur. Eine Kontrolle der Originalfragebögen kann klären, ob die Werte tatsächlich nicht vorhanden sind oder ob nur bei der Eingabe ein Wert ausgelassen wurde.
- **Ausreißer:** Werte, die ihrer Größe nach weit außerhalb des Bereichs der anderen Beobachtungen liegen, sollten ebenfalls Verdacht wecken.
Im obigen Beispiel sticht natürlich der Wert von 28000 für das Einkommen ins Auge. Möglicherweise kommt tatsächlich ein Multimillionär in der Stichprobe vor, vermutlich wurde das Jahreseinkommen aber nicht in Einheiten von 1000 Euro sondern in Einheiten von 1 Euro notiert.

Die Klärung dieser Fälle, also eine Korrektur, wo möglich, bzw. eine Änderung auf Missing values, wo die ursprünglichen Werte nicht mehr eruierbar sind, nennt man die **Fehlerbereinigung**. Diese lästige und nicht sehr interessante Arbeit sollte noch durchgeführt werden, bevor ernsthafte Schritte in Richtung Datenauswertung unternommen werden.

5.2 Einfache deskriptive Statistiken

Sind Fehlerkontrolle und Fehlerbereinigung zumindest vorläufig abgeschlossen – oft werden noch Fehler in weiteren Schritten der Datenauswertung entdeckt – kann mit der Erstellung deskriptiver Statistiken begonnen werden. Das sind **Kennzahlen**, **Häufigkeitstabellen** und einfache **Grafiken**, die zur Beschreibung der einzelnen Variablen dienen (univariate Statistik), Genauerer dazu in den Abschnitten 8.1 und 10.1. Ziel dieser Arbeiten ist die Beschreibung (Deskription) der Stichprobe, nicht die Überprüfung von Hypothesen.

In einem weiteren Schritt können die gemeinsame Verteilung und der Zusammenhang von zwei Variablen untersucht werden (bivariate Statistiken). Allerdings sollen dabei nur interessante Variablenpaare herangezogen werden (bei 10 Variablen gibt es 45 Variablenpaare, bei 20 Variablen 190 Variablenpaare).

Im Zuge der Beschreibung des Datensatzes kann man zur Einsicht gelangen, dass gewisse Variablen neu kodiert, andere neu gebildet werden müssen. Wenn etwa das Alter von Personen in Klassen von 5 Jahren (10-14, 15-19, 20-24, 25-29, ..) eingeteilt wurde und man zur Erkenntnis kommt, dass Klassen mit 10 Jahren (10-19, 20-24, ..) besser sind, wird eine neue Variable mit der gewünschten Kodierung gebildet. Wenn mehrere Variable die verschiedenen Attraktionen eines Urlaubsortes bewerten, kann die Summe dieser Variablen als Gesamtattraktivitätsscore angesehen werden. Diesen Vorgang der Umkodierung und Neubildung von Variablen nennt man **Datentransformation**.

5.3 Inferenzstatistische Methoden

5.3.1 Ziele der Inferenzstatistik

Während deskriptive Methoden zur Beschreibung der Stichprobe dienen, wird mit den Methoden der **Inferenzstatistik (schließende Statistik)** versucht, von der Stichprobe auf die Grundgesamtheit zu schließen. Die Ziele, die dabei angestrebt werden, sind:

- **Überprüfung (Testen) von Fragestellungen (Hypothesen)**, die die Population betreffen.
Beispiele:
Unterscheiden sich Männer und Frauen in ihrer Einschätzung eines bestimmten Projekts?
Sind die Ausgaben für Besuch von Museen, Theatern, etc. höher als vor drei Jahren?
- Auffinden von Zusammenhängen und von Erklärungsmustern, genannt **Modellbildung**.
Beispiele:
Hängt die Zunahme von Buchungen aus einem bestimmten Land mit den Werbeaktivitäten zusammen, die in diesem Land gesetzt wurden?
Welchen Einfluss hat die Erhöhung von Mautgebühren auf die Nutzung der bemauteten Straßen?
- **Schätzung und Prognose** (Vorhersage) von Werten in einer Population.
Beispiele:
Wie viel Prozent der Österreicher sind für die Einführung des Ausländerwahlrechts?
Wie hoch sind die durchschnittlichen Tagesausgaben eines Österreichers für Restaurantbesuche?

Im Rahmen dieses Einführungskurses wird fast ausschließlich die Überprüfung von Fragestellungen, also das Testen von Hypothesen besprochen werden (in den Kapiteln 8 bis 12). Nur im Kapitel 11 wird die Modellbildung kurz gestreift.

5.3.2 Signifikanzniveau und p-Wert

Signifikanztests werden zur Überprüfung von Hypothesen eingesetzt. Einer **Nullhypothese (H_0)** wird die **Forschungshypothese (Alternativhypothese, H_A oder H_1)** gegenüber gestellt. Die Forschungshypothese ist interessant, sie besagt, dass ein Unterschied zwischen zwei Gruppen besteht, dass ein Effekt zu beobachten ist oder, dass eine Abweichung von einem vorgegebenen Wert existiert. Dem gegenüber besagt die Nullhypothese gerade, dass kein Unterschied zwischen zwei Gruppen besteht, dass kein Effekt besteht, dass keine Abweichung von einem vorgegebenen Wert existiert.

In der Stichprobe wird fast nie exakt die Situation vorliegen, die in der Nullhypothese formuliert ist. Kleine Abweichungen sind tolerabel, aber wie soll man mit der Situation umgehen, wenn eine starke Abweichung von der Nullhypothese zu beobachten ist? Prinzipiell gibt es zwei Möglichkeiten dafür:

- man zweifelt an der Stichprobe oder
- man zweifelt an der Nullhypothese.

Die Vorgangsweise in der Testtheorie ist die, dass die Nullhypothese angezweifelt wird.

Jetzt geht es noch darum, wie die Abweichung von der Situation in der Stichprobe von der Nullhypothese gemessen wird. Eigenes Rechnen würde erfordern, den Wert einer sog. Teststatistik zu berechnen und diesen Wert mit einem Wert zu vergleichen, den man aus einer Tabelle ermitteln muss. Wird mit einem Statistikprogramm gearbeitet, wird auch ein sog. **p-Wert** ausgegeben, eine Zahl zwischen 0 und 1. Je kleiner der Wert, desto mehr Skepsis ist gegenüber der Nullhypothese angebracht. Die Grenze, bis zu der die Nullhypothese beibehalten wird, bzw. ab der zur Alternativhypothese übergegangen wird, nennt man **Signifikanzniveau (α , Irrtumswahrscheinlichkeit)**. Der gängigste Wert für das Signifikanzniveau ist 5%; mitunter findet man noch 1%. Andere Werte für das Signifikanzniveau sollten nicht gewählt werden.

Ist der p-Wert also kleiner als das Signifikanzniveau, geht die Entscheidung zugunsten der Alternativhypothese aus. Die Sprechweise ist: „**Man verwirft die Nullhypothese**“. Man sagt auch, das Ergebnis sei **signifikant**.

Ist der p-Wert größer, gibt es keinen ausreichenden Grund, die Nullhypothese zu verwerfen. Die Alternativhypothese wurde durch die Stichprobe nicht ausreichend unterstützt.

Im Output des Statistikprogramms SPSS wird der p-Wert als Signifikanz oder als asymptotische Signifikanz bezeichnet.

5.3.3 Fehlermöglichkeiten bei Signifikanztests

Die Entscheidung nach einem Signifikanztest, egal ob zugunsten der Null- oder der Alternativhypothese, fußt zwar auf rationalen Überlegungen der Statistik, kann aber dennoch zu Fehlern führen. Die folgende Tabelle zeigt die potentiellen Fehlermöglichkeiten bei einem Signifikanztest:

	In der Realität trifft	
	H_0 zu	H_0 nicht zu
Entscheidung für H_0	Entscheidung richtig	Fehler 2.Art (β -Fehler)
Entscheidung gegen H_0	Fehler 1.Art (α -Fehler)	Entscheidung richtig

Beispiel: In einer Studie soll untersucht werden, ob Personen zunehmend später ihren Sommerurlaub buchen. Das ist die Forschungs- bzw. Alternativhypothese, die Nullhypothese besagt, dass die Buchung zur gleichen Zeit wie in früheren Jahren erfolgt. Angenommen ein Signifikanztest bestätigt die Alternativhypothese, so lautet ein Ergebnis der Studie, dass die Buchungen später erfolgen. Vielleicht waren in der Stichprobe unverhältnismäßig viele Spätbucher und haben das Ergebnis des Tests zugunsten der Alternativhypothese beeinflusst, obwohl in der Grundgesamtheit keine Tendenz zu späteren Buchungen besteht. Man ist also in Gefahr, die Nullhypothese zu verwerfen, obwohl sie richtig ist. In obiger Terminologie ist es die Gefahr, einen

Fehler 1.Art zu begehen. Hingegen besteht bei einer Entscheidung gegen die Nullhypothese keine Gefahr des Fehlers 2.Art.

Die Wahrscheinlichkeit für den Fehler 1.Art wird durch das Signifikanzniveau (α) kontrolliert, daher auch die Bezeichnung α -Fehler. Die Wahl des Signifikanzniveaus gibt die Risikobereitschaft an, sich dem Fehler 1.Art auszusetzen. Ist man hier sehr risikoscheu (z.B.: $\alpha = 1\%$, statt 5%), lehnt man nicht so leicht die Nullhypothese ab, setzt sich aber vermehrt der Gefahr aus, irrtümlich an der Nullhypothese festzuhalten, obwohl in der Realität die Alternativhypothese zutrifft. Diesen Fehler nennt man Fehler 2.Art. Er wird neben der Wahl des Signifikanzniveaus vor allem vom Stichprobenumfang bestimmt. Besonders bei geringem Stichprobenumfang ist das Risiko groß, einen tatsächlich existierenden Zusammenhang nicht zu entdecken.

5.4 Fragen

- 1) Kann ein Missing Value ein Wild Code sein?
- 2) Kann ein Wild Code ein Ausreißer sein?
- 3) Was ist die Aufgabe der deskriptiven Statistik?
- 4) Ein AIDS-Test ist „negativ“ (für die Testperson natürlich positiv), wenn im Blut der Testperson keine Antikörper nachgewiesen werden können. Kein Test ist ganz sicher. Ist eine Person nicht HIV-infiziert, das Testergebnis dennoch positiv, spricht man von einem falsch-positiven Ergebnis. Ist eine Person HIV-infiziert, das Testergebnis aber negativ, spricht man von einem falsch-negativen Ergebnis. Wenn die Nullhypothese lautet, dass eine Person nicht infiziert ist, was ist der Fehler 1.Art, was der Fehler 2.Art des AIDS-Tests.
- 5) In einer Untersuchung von 50 Männern und 50 Frauen stellt man fest, dass je größer die Schuhe, desto höher ist durchschnittlich das Einkommen der Personen. Trägt also die Schuhgröße etwas zur Erklärung des Einkommens bei?
- 6) Eine Politikerin meldete sich telefonisch in einer Radiosendung zum Thema „FSME-Impfung“ (vulgo: Zeckenschutzimpfung) mit der „statistischen Begründung“ zu Wort: In jenen Bundesländern Österreichs, in denen der Anteil der FSME-Geimpften hoch ist, sei auch der Anteil der Erkrankungen hoch. Daraus leitete sie implizit die Handlungsanweisung ab, sich nicht impfen zu lassen, weil dadurch die Gefahr, sich zu infizieren, reduziert werde. Ist das vernünftig?

Antworten:

- 1) Kann ein „Missing value“ ein „Wild Code“ sein?
Nein. Bei einem Missing value gibt es überhaupt keine Eintragung, bei einem Wild code ist die Eintragung außerhalb des Gültigkeitsbereichs für die Variable.
- 2) Kann ein „Wild Code“ ein Ausreißer sein?
Nein. Bei einem Wild code ist die Eintragung außerhalb des Gültigkeitsbereichs der Variablen, bei einem Ausreißer ist der Wert nur verdächtig, weil er stark von den anderen abweicht.
- 3) Was ist die Aufgabe der deskriptiven Statistik?
Die Beschreibung und Zusammenfassung der Stichprobe.

- 4) Ein AIDS-Test ist „negativ“ (für die Testperson natürlich positiv), wenn im Blut der Testperson keine Antikörper nachgewiesen werden können. Kein Test ist ganz sicher. Ist eine Person nicht HIV-infiziert, das Testergebnis dennoch positiv, spricht man von einem falsch-positiven Ergebnis. Ist eine Person HIV-infiziert, das Testergebnis aber negativ, spricht man von einem falsch-negativen Ergebnis. Wenn die Nullhypothese lautet, dass eine Person nicht infiziert ist, was ist der Fehler 1.Art, was der Fehler 2.Art des AIDS-Tests?

Der Fehler 1.Art tritt dann auf, wenn die Nullhypothese verworfen wird, obwohl sie gültig ist. In diesem Beispiel also, wenn die Entscheidung aufgrund des AIDS-Tests lautet, dass die Person HIV-infiziert ist, obwohl sie es nicht ist; also dann, wenn ein falsch-positiver Befund vorliegt. Analog entspricht der Fehler 2.Art einem falsch-negativen Befund.

- 5) In einer Untersuchung von 50 Männern und 50 Frauen stellt man fest, dass je größer die Schuhe, desto höher ist durchschnittlich das Einkommen der Personen. Trägt also die Schuhgröße etwas zur Erklärung des Einkommens bei?

Nein. Vermutlich hat die Variable Geschlecht zu dieser „Scheinkorrelation“ geführt. Männer tragen größere Schuhe und erzielen höhere Einkommen.

- 6) Eine Politikerin meldete sich telefonisch in einer Radiosendung zum Thema „FSME-Impfung“ (vulgo: Zeckenschutzimpfung) mit der „statistischen Begründung“ zu Wort: In jenen Bundesländern Österreichs, in denen der Anteil der FSME-Geimpften hoch ist, sei auch der Anteil der Erkrankungen hoch. Daraus leitete sie implizit die Handlungsanweisung ab, sich nicht impfen zu lassen, weil dadurch die Gefahr, sich zu infizieren, reduziert werde. Ist das vernünftig?

Nein. Man kann ja durchaus gegen Impfungen sein, die Argumentation der Politikerin (das ist nicht erfunden!) ist aber Schwachsinn. Wiederum handelt es sich um eine Scheinkorrelation. In den Bundesländern, in denen die Zecken verseucht sind, lassen sich mehr Personen impfen. Dort ist aber auch der Anteil an Erkrankten (vermutlich nicht geimpft) höher.

6 Berichterstattung

6.1 Projektbericht

Als sichtbares Ergebnis einer empirischen Untersuchung sollte ein **Projektbericht** vorliegen. Dieser kann bei einer kleinen Studie aus wenigen Seiten, bei großen Projekten aus mehreren Bänden bestehen. Der Projektbericht sollte Antworten auf folgende Fragen geben:

- „Was wurde untersucht?“
- „Wie wurde untersucht?“
- „Welches sind die Ergebnisse?“

Im Projektbericht ist daher an folgende Punkte zu denken:

- **Resultate:** In diesem Abschnitt sollen alle Resultate zur Diskussion gestellt werden, die im Rahmen der Studie erzielt wurden. Dabei geht es nicht um Details aus der Datenaufbereitung und Datenanalyse sondern um Ergebnisse, die mit der eigentlichen Fragestellung des Projekts in Bezug stehen. Auch Ergebnisse, die ursprüngliche Forschungshypothesen nicht bestätigen (nicht signifikante Ergebnisse), sollen hier angeführt werden. Eine rein technische Präsentation – etwa auf statistischer Ebene – ist zu vermeiden.
- **Methoden:** Neben den Resultaten sollten auch die einzelnen methodischen Schritte in nachvollziehbarer Weise dokumentiert werden. Was wurde wodurch operationalisiert, warum wurden Querschnittsdaten verwendet, wie wurde die Stichprobe gezogen, was war das Erhebungsinstrument (meist ein Fragebogen) etc.
- **Kurzfassung der Resultate:** Für eine Veröffentlichung in Medien und Fachzeitschriften ist eine Kurzfassung der Resultate nötig. Hier ist eine Konzentration auf jene Ergebnisse möglich, die als interessant erscheinen.
- **Fragebogen:** Auf jeden Fall sollte im Anhang der Fragebogen (oder ein anderes Erhebungsinstrument) enthalten sein.
- **Dokumentation der Daten:** Für Sekundäranalysen ist auch an eine Angabe und Dokumentation der Daten (natürlich in anonymisierter Form) zu denken. Bei Auftragsforschung oder bei sehr großen Datensätzen werden die Daten natürlich nicht veröffentlicht, die Daten sollten aber auf Diskette oder CD-ROM dem Bericht beigelegt sein.

6.2 Ethik und Politik in der Forschung

Als Abschluss des vor-statistischen Teils des Skriptums soll eine Sensibilisierung für ethische und politische Probleme in der Forschung erfolgen. Im Konkreten geht es in der Ethik um:

- Freiwilligkeit der Teilnahme
- Keine Schädigung der Teilnehmer
- Anonymität und Vertraulichkeit

- Täuschung der Teilnehmer (manchmal notwendig, um das Studienziel zu erreichen – Blindversuche, Doppelblindversuche)
- Vollständigkeit der wissenschaftlichen Darstellung (Vorgangsweise und Ergebnisse)

Politische Problempunkte sind nicht die Methoden sondern Inhalt und Verwendung der Studien. Die Forderung, dass die politische Orientierung des Auftraggebers und/oder des Forschers die Form und die Ergebnisse der Arbeit nicht beeinflussen darf („wertfreie Wissenschaft“), ist leicht erhoben, in der Realität aber nicht immer erfüllbar. Wissenschaft, vor allem die Sozialwissenschaften sind nie völlig losgelöst von Einflüssen der Politik.

6.3 Fragen

- 1) Was ist der Sinn eines Projektberichts?
- 2) Welche Punkte sollte er besprechen?
- 3) Worauf beziehen sich ethische, worauf politische Problempunkte wissenschaftlicher Forschung?

7 Eine kategoriale Variable

In empirischen Studien sind kategoriale Variable sehr häufig anzutreffen. Fragen nach Geschlecht oder Beruf liefern kategoriale Variable, genauso wie jede Frage nach einer Einstellung, die auf einer Likert-Skala gemessen wird.

7.1 Numerische und grafische Beschreibung

Die Auswertung der Stichprobe, wenn nur eine kategoriale Variable vorhanden ist, ist eine einfache Auszählung. Für jede Ausprägung der Variable wird gezählt, wie oft sie in der Stichprobe vorgekommen ist. Die numerische Präsentation erfolgt als Tabelle, in der die Häufigkeiten am besten sowohl als absolute als auch als relative Häufigkeiten oder Prozent dargestellt werden. Relative Häufigkeiten erhält man, wenn man die absoluten Häufigkeiten durch die Gesamtzahl an Beobachtungen dividiert; multipliziert man diese relativen Häufigkeiten mit 100 erhält man Prozent. Werden in der Tabelle nur relative Häufigkeiten oder Prozente angegeben, sollte eine Angabe gemacht werden, wie groß die Gesamtzahl an Beobachtungen war.

Beispiel: In einer Fremdenverkehrsregion wurde von 145 Gästen erhoben, woher sie stammen. Die Einteilung der Herkunft erfolgte nach folgender Aufteilung: Österreich, Deutschland, restliche Länder der EU, Nicht-EU-Staaten. Eine Tabelle könnte folgendes Aussehen haben (LANDCODE ist die Variable, in der das Herkunftsland der Urlaubsgäste gespeichert ist):

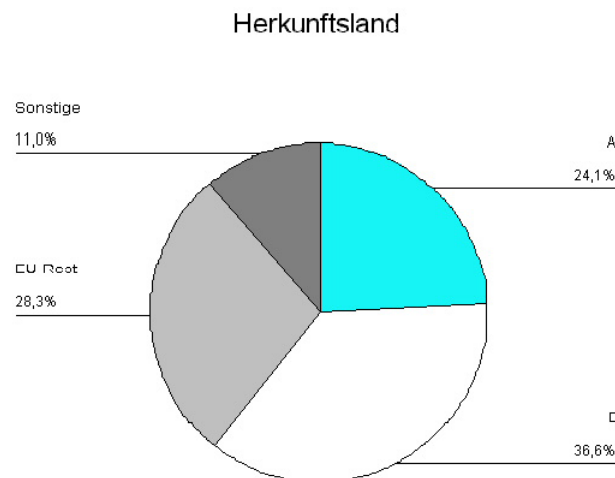
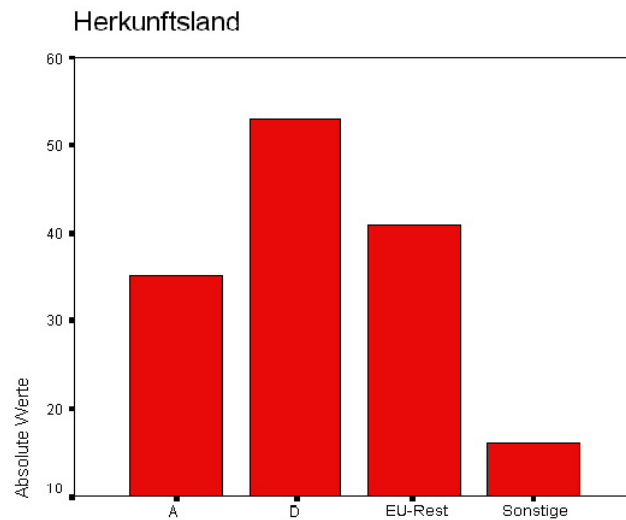
LANDCODE

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	A	35	24,1	24,1	24,1
	D	53	36,6	36,6	60,7
	EU-Rest	41	28,3	28,3	89,0
	Sonstige	16	11,0	11,0	100,0
	Gesamt	145	100,0	100,0	

Achtung: obige Tabelle ist ein SPSS-Output, in dem in der letzten Spalte kumulierte (aufaddierte) Prozente angegeben sind. Für nominale Variable wie hier ist diese Angabe sinnlos, für ordinale Variable (etwa Likert-Skalen) hat diese Spalte einen Sinn.

Als grafische Beschreibungen können Balken, Kreis- und Tortendiagramme dienen. Dabei werden Balken, Kreissektoren bzw. Tortenstücke entsprechend der beobachteten Häufigkeiten in der Stichprobe gezeichnet. Die Interpretation dieser Grafiken ist selbsterklärend. Wie bei Tabellen sollte auch bei den Grafiken eine Angabe über den Stichprobenumfang vorhanden sein.

Beispiel: obige Tabelle führt zu folgendem Balkendiagramm, hier mit absoluten Häufigkeiten als Höhe der Balken. Im Anschluss das zugehörige Kreisdiagramm



SPSS-Kommandos:

Tabellen: Analysieren -> Deskriptive Statistiken -> Häufigkeiten

Diagramme: Grafiken -> Balken...

Grafiken -> Kreis...

7.2 Test für den Anteilswert

Als inferenzstatistisches Verfahren für kategoriale Variablen wird der Test für den Anteilswert vorgestellt. Am besten man zieht die Methode an einem Beispiel auf.

Beispiel: In einer Stadt wird überlegt, die ganze Innenstadt vom Individualverkehr freizuhalten. In einer Befragung von 240 Personen waren 132 Personen für diesen Vorschlag, 108 dagegen. Kann man aufgrund dieser Stichprobe schon schließen, dass mehr als die Hälfte der Einwohner der Stadt für den Vorschlag sind?

Es wird einer Nullhypothese, dass der Anteil in der Grundgesamtheit 50% beträgt, eine Alternativhypothese gegenüber gestellt, dass der Anteil über 50% liegt.

EINSTELL			
	Beobachtetes N	Erwartete Anzahl	Residuum
Pro	132	120,0	12,0
Contra	108	120,0	-12,0
Gesamt	240		

Statistik für Test		EINSTELL
Chi-Quadrat		2,400
df		1
Asymptotische Signifikanz		,121

Interpretation dieses Ergebnisses:

Im ersten Teil (Beobachtetes N) gibt die Tabelle aus, wie viel Personen aus der Stichprobe für (132), wie viel sich gegen (108) das Projekt ausgesprochen haben. Dann wird in einer Spalte (Erwartete Anzahl) angeführt, wie viel Befürworter bzw. Gegner man in einer Stichprobe von 240 Personen erwarten würde, wenn genau 50 Prozent für und 50 Prozent gegen das Projekt (das entspricht der Nullhypothese) wären, nämlich jeweils 120. In der letzten Spalte (Residuum) ist die Differenz zwischen der beobachteten und erwarteten Anzahl angeführt.

In der zweiten und kleineren Tabelle (mit „Statistik für Test“ überschrieben) ist das eigentliche Testergebnis angeführt, für uns von Hauptinteresse ist der Wert 0,121 rechts unten. Das ist der p-Wert dieses Tests, in SPSS mit Signifikanz oder (wie hier) Asymptotische Signifikanz angegeben. Verglichen mit einem Signifikanzniveau von 5% (entspricht der Zahl 0,05) ist er größer, das bedeutet, dass die Nullhypothese (der Anteil der Befürworter ist 50%) beibehalten wird. Die Abweichung in der Stichprobe ist nicht signifikant.

Anmerkungen:

- Die anderen ausgegebenen Werte sind: Chi-Quadrat = 2,400 ist der sog. Wert der Teststatistik. Er misst, wie stark die Abweichung in der Stichprobe von der Nullhypothese ist. df = 1 (Freiheitsgrade, degrees of freedom) gibt an, in welcher Tabelle man einen Wert zu suchen hätte, mit dem der Chi-Quadrat-Wert zu vergleichen ist. Wenn man sich auf den p-Wert konzentriert, sind diese beiden Werte nicht von großer Bedeutung.

- Eigentlich bezieht sich der p-Wert (0,121) auf den Test mit der Alternativhypothese, dass der Anteil der Befürworter ungleich 50 Prozent ist (man nennt dies auch einen **zweiseitigen** Test, weil Abweichungen von der Nullhypothese nach oben, mehr als 50%, und nach unten, weniger als 50%, in der Hypothese zugelassen werden). Unsere eigentliche Frage, ob mehr als 50% Befürworter sind, beinhaltet aber nur die Abweichung nach oben. Wir sollten eigentlich einen sog. **einseitigen** Test durchführen. Für einen Einführungskurs belassen wir es aber bei zweiseitigem Testen!

SPSS-Kommandos:

Der Anteilstest ist nur umständlich zu rechnen, man muss die unter der Nullhypothese erwarteten Häufigkeiten händisch eingeben.

Analysieren -> Nichtparametrische Tests -> Chi-Quadrat..

7.3 Fragen

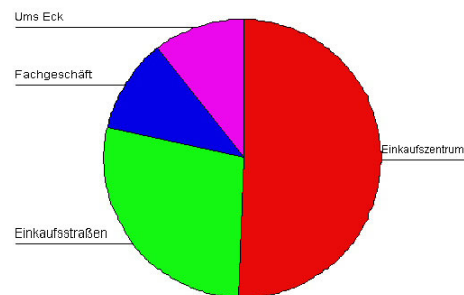
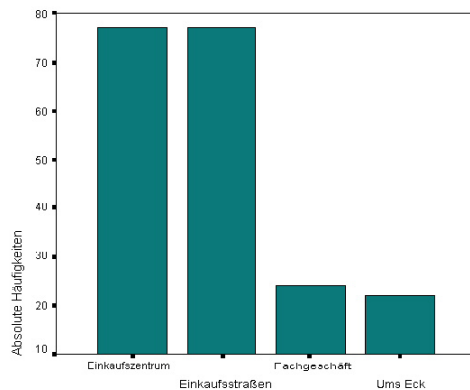
1) Ein Mitarbeiter eines städtischen Hallenbades führt eine Untersuchung zum Besuch dieser Einrichtung durch. Dazu notiert er sich über einen Zeitraum von 70 Tagen (=10 Wochen) täglich mehrere Werte, welche davon sind kategoriale Variablen?

- Wie viel Besucher Insgesamt
- Wie viel Schulklassen
- War schulfreier Tag
- War Trainingstag für den Schwimmklub
- Welcher Wochentag
- War es ein Regentag
- Wie hoch war die Außentemperatur
- Wie viel Besucher waren im angeschlossenen Cafe-Restaurant

2) 200 Personen wurden befragt, wo sie am liebsten einkaufen. Folgende Tabelle wurde nach einer Auszählung erhalten:

	Häufigkeit
Einkaufszentrum	83
Einkaufsstraßen	66
Fachgeschäfte	28
Ums Eck	23

Betrachten Sie das folgende Balken- und das folgende Tortendiagramm:



Welches der beiden Diagramme passt zu obiger Tabelle?

3) In einer Umfrage gaben 210 von 400 Befragten an, die Spitzenkandidatin einer Landespartei zu kennen. Ist sie damit schon so bekannt wie ihr Vorgänger, dessen Bekanntheitsgrad bei 60% lag? Dazu wurde ein Test mit folgendem Ergebnis gerechnet.

BEKANNT

	Beobachtetes N	Erwartete Anzahl	Residuum
Bekannt	210	240,0	-30,0
Nicht bekannt	190	160,0	30,0
Gesamt	400		

Statistik für Test

	BEKANNT
Chi-Quadrat	9,375
df	1
Asymptotische Signifikanz	,002

- Formulieren Sie Null- und Alternativhypothese des Tests.
- Für welche Hypothese entscheiden Sie sich nach dem Test?

Antworten

1) Ein Mitarbeiter eines städtischen Hallenbades führt eine Untersuchung zum Besuch dieser Einrichtung durch. Dazu notiert er sich über einen Zeitraum von 70 Tagen (=10 Wochen) täglich mehrere Werte, welche davon sind kategoriale Variablen?

- Wie viel Besucher Insgesamt

- Wie viel Schulklassen
- War schulfreier Tag (kategorial)
- War Trainingstag für den Schwimmklub (kategorial)
- Welcher Wochentag (kategorial)
- War es ein Regentag (kategorial)
- Wie hoch war die Außentemperatur
- Wie viel Besucher waren im angeschlossenen Cafe-Restaurant

2) 200 Personen wurden befragt, wo sie am liebsten einkaufen.

Welches der beiden Diagramme passt zu obiger Tabelle?

Weder das Balkendiagramm (gleich hohe Balken für Einkaufszentren und Einkaufsstraßen sind ein Widerspruch) noch das Kreisdiagramm (mehr als 50% für Einkaufszentren sind zuviel) passen zur Häufigkeitstabelle.

3) In einer Umfrage gaben 210 von 400 Befragten an, die Spitzenkandidatin einer Landespartei zu kennen. Ist sie damit schon so bekannt wie ihr Vorgänger, dessen Bekanntheitsgrad bei 60% lag? Dazu wurde ein Test mit folgendem Ergebnis gerechnet.

- Formulieren Sie Null- und Alternativhypothese des Tests. Die Nullhypothese lautet: der Bekanntheitsgrad der Politikerin liegt bei 60% (das entspricht einer erwarteten Häufigkeit von 240 bei 400 befragten Personen), die Alternativhypothese besagt, dass der Bekanntheitsgrad nicht 60% beträgt.
- Für welche Hypothese entscheiden Sie sich nach dem Test? Für die Alternativhypothese, weil die asymptotische Signifikanz kleiner als gängige Signifikanzniveaus ist.

8 Zwei kategoriale Variablen

An folgendem Beispiel sollen sowohl die Verfahren zur Beschreibung der Stichprobe als auch ein statistischer Test für diese Datensituation vorgestellt werden.

Beispiel: 198 Personen, 102 Frauen und 96 Männer wurden über ihre Einstellung zu Wellnessurlaub befragt. Die Einstellung wurde ursprünglich auf einer 5-stufigen Likert-Skala gemessen. Für die Betrachtung jetzt genügt es, die Einstellung in die drei Kategorien „dafür“, „neutral“ und „dagegen“ einzuteilen. Es liegen somit zwei kategoriale Variablen vor, Geschlecht und Einstellung zu Wellnessurlaube.

Mit den Methoden des vorigen Kapitels sind wir in der Lage jede Variable für sich zu beschreiben. Wie beschreibt man die gemeinsame Verteilung der beiden Variablen?

8.1 Numerische und grafische Beschreibung

Da Geschlecht zwei Ausprägungen und die Einstellung zu Wellnessurlaube drei Ausprägungen hat, sind insgesamt 6 ($=2 \cdot 3$) Ausprägungskombinationen der beiden Variablen denkbar. Die einfache Auszählung, wie oft diese Ausprägungskombinationen in der Stichprobe vorgekommen sind, führt zu einer sog. **Kreuztabelle**.

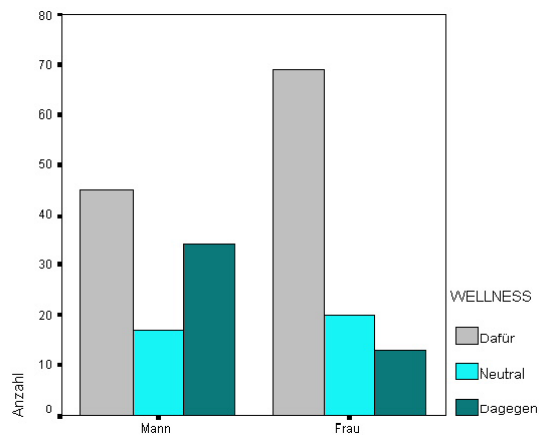
SEX * WELLNESS Kreuztabelle

			WELLNESS			Gesamt
			Dafür	Neutral	Dagegen	
SEX	Mann	Anzahl	45	17	34	96
		% der Gesamtzahl	22,7%	8,6%	17,2%	48,5%
	Frau	Anzahl	69	20	13	102
		% der Gesamtzahl	34,8%	10,1%	6,6%	51,5%
Gesamt		Anzahl	114	37	47	198
		% der Gesamtzahl	57,6%	18,7%	23,7%	100,0%

8.1.1 Gesamtprozent und gruppierte Balkendiagramme

Obige Tabelle ist eine Zusammenfassung der Stichprobe. Die Eintragung links oben im Zahlenbereich bedeutet, dass 45 Männer sich für Wellnessurlaube geäußert hatten, etc. In diesem Beispiel sind nicht nur die absoluten Häufigkeiten angegeben sondern auch relative. Die Prozentangabe unter 45 bedeutet, dass sich 22,7% der Befragten Männer und positiv zu Wellnessurlaube eingestellt waren. Diese Prozentangabe bezieht sich also auf die Gesamtzahl der Befragten, man nennt diese Prozentangabe daher auch **Gesamtprozent**.

Eine grafische Beschreibung kann über **gruppierte Balkendiagramme** erfolgen. Die folgende Grafik ist ein solches Diagramm, die Gruppen sind durch den Faktor SEX gebildet und die Höhen der Balken entsprechen absoluten Häufigkeiten.



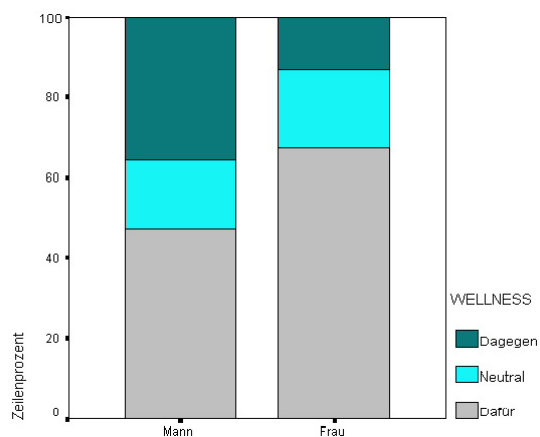
8.1.2 Zeilen- und Spaltenprozent, gestapelte Balkendiagramme

Werden Prozente für Männer bzw. Frauen separat berechnet, erhält man sog. **Zeilenprozent**. In der folgenden Tabelle sind Zeilenprozent für das obige Beispiel angeführt, die Zeilensumme ergibt jeweils 100%. So gibt die erste Eintragung an, dass unter den Männern 46,9% positiv zu Wellnessurlaube eingestellt ist, die Interpretation der anderen Eintragungen verläuft analog.

SEX * WELLNESS Kreuztabelle
% von SEX

		WELLNESS			Gesamt
		Dafür	Neutral	Dagegen	
SEX	Mann	46,9%	17,7%	35,4%	100,0%
	Frau	67,6%	19,6%	12,7%	100,0%
Gesamt		57,6%	18,7%	23,7%	100,0%

Gestapelte Balkendiagramme geben diese Information am besten wieder. Das folgende Diagramm enthält die Information der vorigen Tabelle.

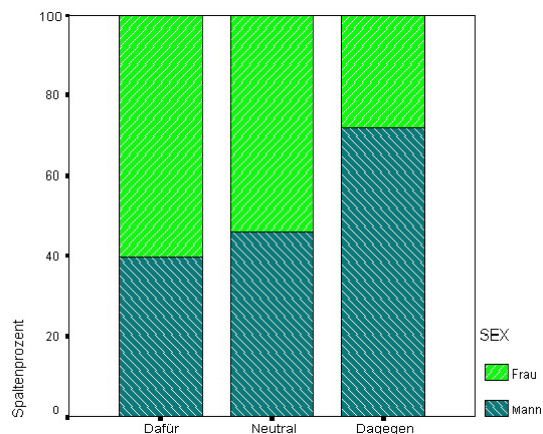


Nach demselben Prinzip sind auch **Spaltenprozent** berechenbar. Unten stehende Tabelle gibt diese Auswertung an.

SEX * WELLNESS Kreuztabelle
% von WELLNESS

		WELLNESS			Gesamt
		Dafür	Neutral	Dagegen	
SEX	Mann	39,5%	45,9%	72,3%	48,5%
	Frau	60,5%	54,1%	27,7%	51,5%
Gesamt		100,0%	100,0%	100,0%	100,0%

Die Interpretation läuft jetzt spaltenweise. Von den Personen, die positiv zu Wellnessurlaube eingestellt sind, waren 39,5% Männer und 60,5% Frauen. Für die weiteren Spalten ist die Interpretation analog durchzuführen. Dieser Tabelle entspricht das folgende gestapelte Balkendiagramm.



SPSS-Kommandos:

Kreuztabellen mit Gesamt-, Spalten- oder Zeilenprozent erhält man durch:

Analysieren -> Deskriptive Statistiken -> Kreuztabellen

Gestapelte Balkendiagramme, die Spalten- oder Zeilenprozent entsprechen, sind leider nur umständlich zu erhalten. Die automatisch ausgewählten Farben sind überdies fürchterlich grell und sollten beim Bearbeiten der Grafik umgestellt werden.

Grafiken -> Interaktiv -> Balken...

8.2 Homogenitätstest

Als inferenzstatistisches Verfahren dieses Kapitels wird der sog. Homogenitätstest vorgestellt. Mit ihm kann die Frage beantwortet werden, ob die Verteilung einer kategorialen Variablen in zwei oder mehreren Gruppen unterschiedlich ist. Die Nullhypothese besagt immer, dass kein Unterschied in der Verteilung des kategorialen Merkmals zwischen den Gruppen besteht. Die Alternativhypothese lautet: es gibt einen Unterschied zwischen den Gruppen.

Rechnerisch liefert der Pearson-Chi-Quadrat-Test die Antwort auf diese Frage. Im SPSS-Output findet man noch weitere Tests, für uns ist nur die Zeile mit dem Pearson-Chi-Quadrat-Test relevant.

Beispiel: Im Beispiel der Einstellung zu Wellnessurlaube kann etwa die Frage gestellt werden, ob sich Frauen und Männer in ihrer Einstellung dazu unterscheiden. Ein SPSS-Output dazu hat folgendes Aussehen:

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	14,510	2	,001
Likelihood-Quotient	14,875	2	,001
Zusammenhang linear-mit-linear	13,293	1	,000
Anzahl der gültigen Fälle	198		

a. 0 Zellen (.0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 17,94.

Die markierte Zelle enthält den p-Wert, der zur Interpretation des Testergebnisses dient. Dieser Wert ist hier (0,001) kleiner als das übliche Signifikanzniveau von 5% ($=0,05$). Das heißt, dass die Nullhypothese (Frauen und Männer unterscheiden sich nicht) verworfen werden muss. Die Entscheidung nach dem Test lautet also: Männer und Frauen unterscheiden sich in der Einstellung zu Wellnessurlaube.

Anmerkungen:

- Die anderen ausgegebenen Werte sind: Wert = 14,510 ist der sog. Wert der Teststatistik. Er misst, wie stark die Abweichung in der Stichprobe von der Nullhypothese ist. $df = 2$ (Freiheitsgrade, degrees of freedom) gibt an, in welcher Tabelle man einen Wert zu suchen hätte, mit dem der Chi-Quadrat-Wert zu vergleichen ist. Wenn man sich auf den p-Wert konzentriert, sind diese beiden Werte nicht von großer Bedeutung.
- Die anderen Zeilen beziehen sich auf Tests, die in diesem Rahmen nicht besprochen werden können.
- Die Fußnote der Tabelle gibt an, dass dieser Test auf „guten“ Daten beruht. Je weniger Zellen (in diesem 2x3-Schema) erwartete Häufigkeiten (wie man sie berechnet, ist für Sie nicht von Interesse) aufweisen, die kleiner als 5 sind, desto besser ist es. Hier gibt es keine Probleme, der kleinste Wert ist 17,94.

SPSS-Kommandos:

Analysieren -> Deskriptive Statistiken -> Kreuztabellen (bei Statistik Chi-Quadrat ankreuzen)

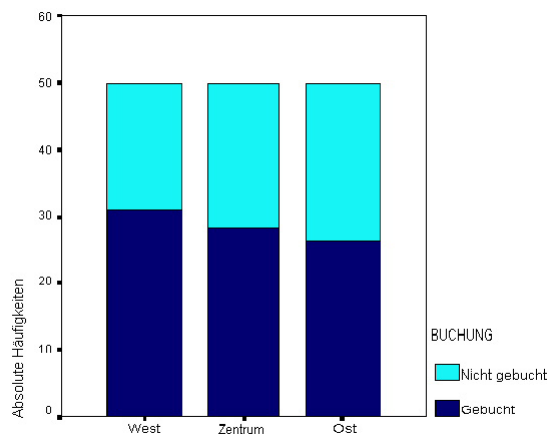
8.3 Fragen

In einer Stadt werden Anfang Mai in den drei Stadtteilen je 50 Personen befragt, ob schon ein Urlaub im Sommer nicht nur geplant, sondern schon gebucht ist. Die Auszählung für die drei Stadtteile ergibt folgende Tabelle:

	West	Zentrum	Ost
Gebucht	31	28	26
Nicht gebucht	19	22	24

1) Wenn für jeden Stadtteil der Anteil der Personen, die schon einen Sommerurlaub gebucht haben, beschrieben werden soll, sollen Spalten- oder Zeilenprozent berechnet werden?

2) Betrachten Sie das folgende Balkendiagramm!



Welche Aussagen kann man daraus ableiten?

- Da die Balken nicht auf eine Höhe von 100% gehen ist es kein gestapeltes Balkendiagramm
- In jedem Stadtteil haben in der Stichprobe mindestens die Hälfte der Personen ihren Sommerurlaub schon gebucht
- Da in jedem Stadtteil gleich viel Personen befragt wurden, hätte ein gestapeltes Balkendiagramm mit relativen Häufigkeiten fast das gleiche Aussehen.
- Im Stadtteil Ost ist die Buchungsrate am geringsten.

3) Ein Homogenitätstest, ob sich die Stadtteile in ihrer Buchungsfrequenz für den Sommerurlaub unterscheiden wurde gerechnet.

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (2-seitig)
Chi-Quadrat nach Pearson	1,032	2	,597
Likelihood-Quotient	1,035	2	,596
Zusammenhang linear-mit-linear	1,011	1	,315
Anzahl der gültigen Fälle	150		

- Was sind Null- und Alternativhypothese dieses Tests?
- Wird die Nullhypothese nach dem Test verworfen?

Antworten

1) Wenn für jeden Stadtteil der Anteil der Personen, die schon einen Sommerurlaub gebucht haben, beschrieben werden soll, sollen Spalten- oder Zeilenprozent berechnet werden?

Spaltenprozent. Die Angaben für die Stadtteile sind in den Spalten angeordnet.

2) Betrachten Sie das folgende Balkendiagramm!

Welche Aussagen kann man daraus ableiten?

- Da die Balken nicht auf eine Höhe von 100% gehen ist es kein gestapeltes Balkendiagramm. Falsch, hier ist ein gestapeltes Balkendiagramm mit absoluten Häufigkeiten gegeben. Das macht Sinn, weil in jedem Stadtteil gleich viel Personen befragt wurden.
- In jedem Stadtteil haben in der Stichprobe mindestens die Hälfte der Personen ihren Sommerurlaub schon gebucht. Richtig, der untere Teil der Balken steht jeweils für Buchung und ist nie kleiner als der obere Teil.
- Da in jedem Stadtteil gleich viel Personen befragt wurden, hätte ein gestapeltes Balkendiagramm mit relativen Häufigkeiten fast das gleich Aussehen. Richtig, man müsste nur die Beschriftung der Skala statt von 0 bis 50 durch eine Beschriftung von 0 bis 100 ersetzen.
- Im Stadtteil Ost ist die Buchungsrate am geringsten. Richtig, der untere Balkenteil ist dort am kleinsten.

3) Ein Homogenitätstest, ob sich die Stadtteile in ihrer Buchungsfrequenz für den Sommerurlaub unterscheiden wurde gerechnet.

- Was sind Null- und Alternativhypothese dieses Tests? Die Nullhypothese besagt, dass kein Unterschied in der Buchungsrate zwischen den Stadtteilen besteht, in der Alternativhypothese werden solche Unterschiede behauptet.
- Wird die Nullhypothese nach dem Test verworfen? Nein, dazu ist die asymptotische Signifikanz für den Chi-Quadrat-Test nach Pearson (0,597) viel zu hoch.

9 Eine metrische Variable

Metrische Variable können sinnvoll mit Zahlen beschrieben werden. Beispiele dafür sind Größen, die mit physikalischen Messinstrumenten ermittelt werden (Körpergröße, Gewicht, Wartezeiten, etc.), viele ökonomische Größen (Einkommen, Reisekosten, etc.), Zählvariable (Anzahl Kinder, Anzahl Wochenendurlaube, etc.), aber auch Gesamtscores von vielen Einzelitems (etwa als Zusammenfassung von Matrixfragen) können als metrische Variable aufgefasst werden.

Metrische Variablen erlauben den Einsatz von mehr Methoden als kategoriale, sie erfordern aber auch mehr Sorgfalt bei deren Einsatz, da nicht alle Methoden für jede metrische Variable geeignet sind.

9.1 Numerische und grafische Beschreibung

Die Auflistung, wie oft jeder Wert vorgekommen ist, führt – anders als bei kategorialen Variablen – selten zu mehr Übersicht. Eine Einteilung der Werte in Klassen ist notwendig, um zu einer Datenreduktion zu gelangen. So werden etwa Alter (in Jahren) zu Altersklassen zu jeweils 5 oder 10 Jahren zusammengefasst.

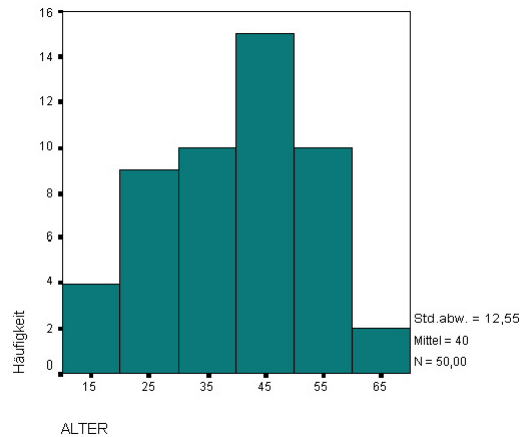
Beispiel: Die Altersverteilung der Gäste eines Thermalbades wird anhand einer Stichprobe von 50 Gästen durch folgende Tabelle beschrieben:

Alter	Absolute Häufigkeit	Relative Häufigkeit
Unter 20	4	8%
20 – 29	9	18%
30 – 39	10	20%
40 – 49	15	30%
50 – 59	10	20%
Über 60	2	4%

9.1.1 Histogramm

Aus so einer Tabelle kann leicht ein **Histogramm** abgeleitet werden. Bei einem Histogramm sind im Unterschied zu einem Balkendiagramm keine Abstände zwischen den Balken, die Balken berühren sich also. Sind die Klassen gleich breit (der Fall, den wir behandeln), gibt die Höhe des Balkens an, wie viel Beobachtungen in der jeweiligen Klasse liegen. Dort wo die Balken hoch sind, sind Beobachtungen also häufiger als dort, wo die Balken niedriger sind.

Beispiel: Für das obige Beispiel mit der Altersverteilung der Gäste eines Thermalbades erhält man folgendes Histogramm, die Klassen sind durch die jeweiligen Klassenmittelpunkte beschriftet. Die Beschriftung links ist in absoluten Häufigkeiten erfolgt, manchmal geschieht dies auch in relativen Häufigkeiten. Die Altersklasse von 40 – 49 Jahren ist also am häufigsten als Gast anzutreffen.



Bietet ein Histogramm einen guten optischen Überblick über die Verteilung der Daten, so will man gewisse Charakteristika der Verteilung durch die Angabe einer Zahl beschreiben. Man nennt solche charakteristischen Werte auch Maßzahlen.

9.1.2 Lagemaße

Lagemaße sollen angeben, wo typische Vertreter der Stichprobe sind. Zur Anwendung kommen folgende Maße:

- **Mittelwert:** Das bekannteste Lagemaß ist wohl der Mittelwert, berechnet als das arithmetische Mittel aller Werte der Stichprobe.
- **Median:** Zur Berechnung des Medians werden die Werte zuerst der Größe nach sortiert. Der Median ist der Wert, der nach diesem Sortieren in der Mitte zu finden ist. Er teilt den Datensatz in die untere Hälfte (der kleinen Werte) und die obere Hälfte (der großen Werte).
- **Modus:** Der Modus (oder Modalwert) ist jener Wert, der in der Stichprobe am häufigsten vorgekommen ist. Für metrische Daten ist dies ein eher seltenes Lagemaß, es kann aber auch bei kategorialen Daten eingesetzt werden.

Der Mittelwert kann durch Ausreißer in den Daten stark verfälscht werden, der Median nicht. Man sagt: „der Median ist **robust** gegenüber Ausreißern“.

Beispiel: Alter der Gäste eines Thermalbades.

Mittelwert: 40,02

Median: 41,00

Modus: 41,00

Mittelwert und Median unterscheiden sich nicht sehr stark, das deutet darauf hin, dass die Verteilung nicht stark von der Symmetrie abweicht.

9.1.3 Streuungsmaße

Streuungsmaße sollen angeben, wie stark die Daten von einander abweichen können. Man will also wissen, wie stark sie variieren, wie stark sie streuen. Weichen die Werte nicht stark vom Lagemaß ab, soll das Streuungsmaß klein sein, weichen sie stark ab, soll das Streuungsmaß einen hohen Wert anzeigen. Anders als Lagemaße können Streuungsmaße nie negative Werte annehmen.

Das händische Berechnen dieser Werte ist schon umständlich, auf die Angabe von Formeln wird also verzichtet. Streuungsmaße sind:

- **Varianz:** Berechnet als das arithmetische Mittel der quadrierten Abweichungen vom Mittelwert, ist es das bekannteste Streuungsmaß.
- **Standardabweichung:** Sie ist die Wurzel der Varianz
- **Interquartilsabstand:** Berechnet als Differenz zwischen drittem und erstem Quartil der Daten (siehe Boxplot)

Der Interquartilsabstand ist robust gegenüber Ausreißern, im Gegensatz zu Varianz und Standardabweichung.

Die Größe der Werte selbst wird kaum interpretiert, zum Vergleich von zwei Gruppen werden sie aber herangezogen, um etwa zu überprüfen, ob die Streuung in beiden Gruppen in etwa gleich groß ist.

Beispiel: Alter der Gäste eines Thermalbades.

Varianz: 157,57

Standardabweichung: 12,55

Interquartilsabstand: 20,25

9.1.4 Boxplot

Die Zusammenführung mehrerer bisher besprochener Konzepte ist der **Boxplot**. Ihm liegt die Idee zugrunde, den Datensatz in vier Viertel einzuteilen. Der Median trennt die Daten ja in eine untere und eine obere Hälfte. Bestimmt man von der unteren (bzw. oberen) Hälfte wiederum den Median, so erhält man die Grenze zwischen dem Viertel der kleinsten (bzw. größten) Werte und dem Rest der Daten. Die Einteilung des Datensatzes ist damit wie folgt gegeben:

Minimum – 1. Viertel der Daten (niedrigste Werte) –

1.Quartil – 2. Viertel der Daten –

Median (= 2.Quartil) – 3. Viertel der Daten –

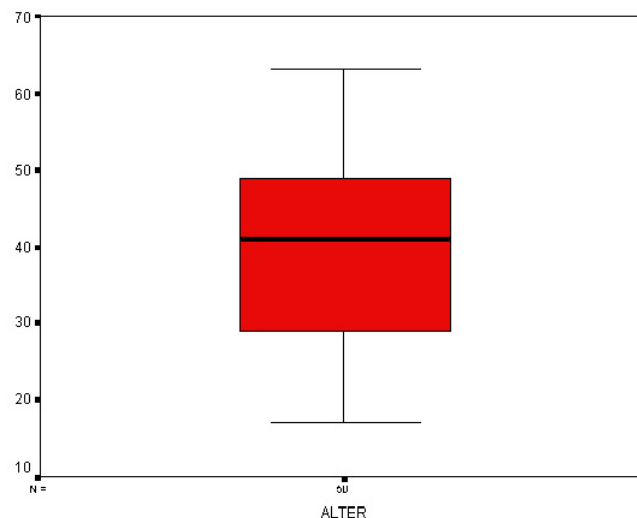
3. Quartil – 4. Viertel der Daten (größte Werte) –

Maximum

Diese fünf Werte (das Minimum und Maximum und die drei Quartile) bestimmen das Aussehen des Boxplots. Eine Box (Schachtel) wird vom ersten zum dritten Quartil gezeichnet, die Box ist dort

unterteilt, wo der Median (2.Quartil) ist. Zusätzlich werden Linien von der Box nach unten zum Minimum und nach oben zum Maximum gezogen.

Beispiel: Alter der Gäste eines Thermalbades.



Klar erkennbar ist die Box, sie geht von einem Wert knapp unter 30 (siehe Skala links) bis ca. 50 und ist durch den Median (knapp über 40) unterteilt. Das Minimum liegt unter 20, das Maximum über 60.

9.1.5 Vergleich mehrerer Verteilungen

Um ein Gefühl für die Maßzahlen, Histogramme und Boxplots zu erhalten werden einige hypothetische Datensätze präsentiert, die mit dem vorherigen einen gemeinsam haben, einen Mittelwert von 40. Zur Wiederholung, der bisher immer präsentierte Datensatz hatte folgende Kennzahlen:

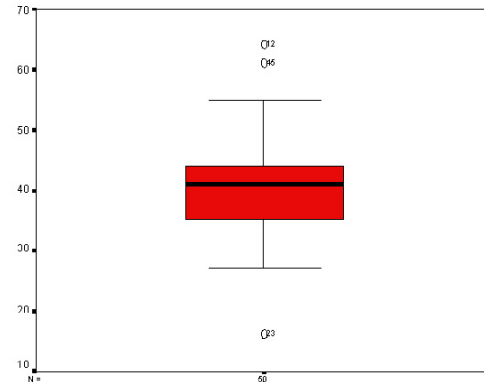
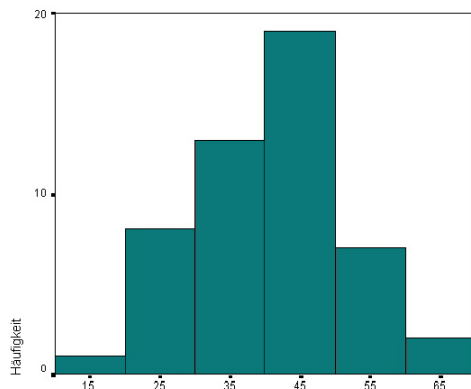
Mittelwert	Median	Varianz	Interquartilsabstand
40	41	157,57	20,25

Verteilung mit Ausreißern

Zuerst eine Verteilung, bei der die Daten stärker um Mittelwert und Median konzentriert sind, einige Werte liegen jedoch nach wie vor weit vom Lagemaß entfernt. Die stärkere Konzentration führt zu kleineren Werten bei den Streuungsmaßen.

Mittelwert	Median	Varianz	Interquartilsabstand
40	41	86,71	9,25

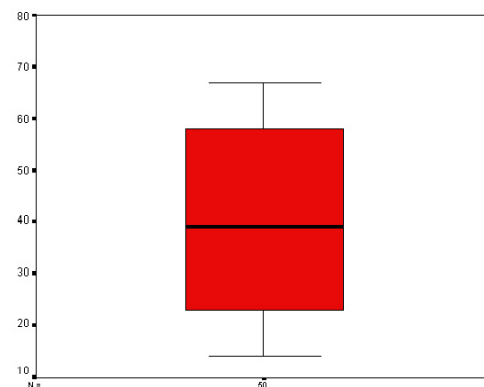
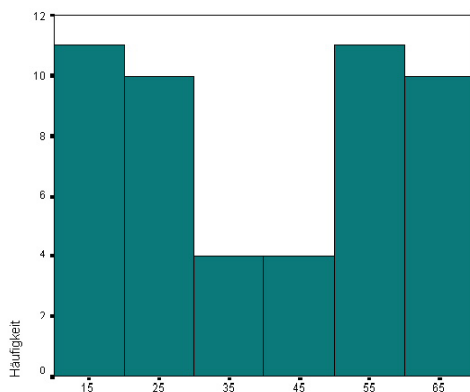
Histogramm und Boxplot haben folgendes Aussehen:



Im Boxplot sind einige Werte durch kleine Kreise markiert. Das bedeutet, dass das Statistikprogramm diese Werte als Ausreißer einstuft, sie liegen zu weit vom Zentrum der Daten entfernt.

U-förmige Verteilung

Als nächstes eine Veränderung, bei der die Werte am Anfang und Ende des Intervalls liegen, dazwischen kaum vorkommen. Man spricht auch von U-förmigen Verteilungen, wenn man die Grafiken sieht, wird gleich klar warum.

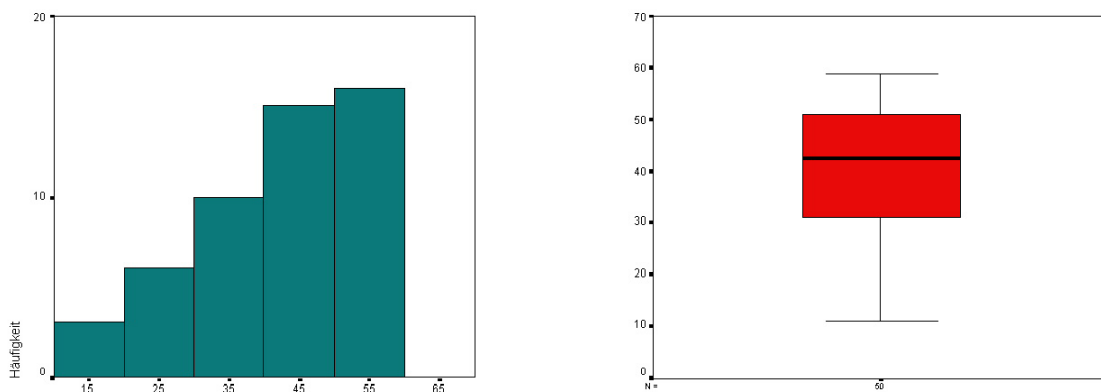


Die Box ist jetzt wesentlich breiter, die Werte am Anfang und Ende des Intervalls werden nicht mehr als Ausreißer klassifiziert. In den Streuungsmaßen kommt jetzt auch zum Ausdruck, dass die Werte im Schnitt weiter vom Zentrum der Daten entfernt sind.

Mittelwert	Median	Varianz	Interquartilsabstand
40	39	332,39	35,75

Schiefe Verteilung

Als letztes eine Veränderung zu einer schiefen Verteilung, einer Verteilung, die etwa im Histogramm eindeutig nicht symmetrisch ist. Zwar waren schon die bisher vorgestellten Verteilungen nicht symmetrisch, man wird bei einer Stichprobe aber fast nie exakte Symmetrie antreffen. In den folgenden Grafiken ist das Abweichen von der Symmetrie aber weit klarer.



Im Histogramm ist keine Symmetrie erkennbar, im Boxplot ist das Abweichen von der Symmetrie in diesem Beispiel nicht ganz so deutlich. Bei Symmetrie sollte die Box (in etwa) in der Mitte durch den Median unterteilt sein, die Linien zum Minimum und Maximum sollten (in etwa) gleich lang sein. Die Unterteilung der Box ist zwar nicht exakt in der Mitte, eine starke Abweichung ist es aber nicht. Die Linien zu den Extrema deuten schon eher ein Abweichen von der Symmetrie an.

Mittelwert	Median	Varianz	Interquartilsabstand
40	42,5	155,00	20,00

SPSS-Kommandos:

Häufigkeiten: Analysieren -> Deskriptive Statistiken -> Häufigkeiten

Histogramm: Analysieren -> Deskriptive Statistiken -> Explorative Datenanalyse

Boxplot: Analysieren -> Deskriptive Statistiken -> Explorative Datenanalyse

Maßzahlen: Analysieren -> Deskriptive Statistiken -> Explorative Datenanalyse

oder Analysieren -> Deskriptive Statistiken -> Häufigkeiten

9.2 Test für den Mittelwert

Als inferenzstatistische Methode behandeln wir den Vergleich des Mittelwerts aus der Stichprobe gegen einen vorgegebenen Wert.

Beispiel: Alter der Gäste eines Thermalbades. Die Leitung des Thermalbades hat in einer Werbekampagne versucht, verstärkt auch jüngere Gäste anzusprechen. Einen Monat nach Beendigung der Werbekampagne ist die schon bekannte Stichprobe gezogen worden und dabei das Alter der Gäste erhoben worden. Deutet dies auf eine Verjüngung bei den Besuchern hin, wenn vor der Werbekampagne der Altersschnitt bei 44 Jahren lag?

Es soll also die Stichprobe mit einem vorgegebenen Wert (hier 44 Jahre) verglichen werden. Die statistische Antwort auf diese Fragestellung ist ein Mittelwertsvergleich, wenn die Daten nicht zu schräg sind, ein sog. t-Test. Die Nullhypothese lautet: Der Mittelwert in der Grundgesamtheit entspricht dem vorgegebenen Wert. Die Alternativhypothese ist die Verneinung der Nullhypothese, der Mittelwert in der Grundgesamtheit ist ein anderer Wert.

Die Formeln sparen wir uns, wie schaut ein Output aus dem SPSS aus?

Test bei einer Stichprobe

Testwert = 44		df	Sig. (2-seitig)	Mittlere Differenz	95% Konfidenzintervall der Differenz	
T					Untere	Obere
ALTER	-2,242	49	,030	-3,98	-7,55	-,41

Oben steht der Wert, gegen den die Stichprobe verglichen wird (Testwert = 44). Das wichtige Ergebnis ist in der Spalte: Sig. (2-seitig) zu finden, es ist dies der sog. Signifikanzwert (in anderen Statistikpaketen auch p-Wert genannt). Er hat hier den Wert 0,030. Dieser Wert muss mit einem Signifikanzniveau (üblich 5%) verglichen werden. Da 0,030 kleiner als das Signifikanzniveau ist, wird die Nullhypothese verworfen und die Alternativhypothese angenommen. Die Daten deuten also darauf hin, dass es zu einer Änderung in der Altersstruktur gekommen ist. Da der Mittelwert in der Stichprobe 40 war, kann man schließen, dass die Besucher durchschnittlich jünger wurden.

SPSS-Kommandos:

Analysieren -> Mittelwerte vergleichen -> T-Test bei einer Stichprobe

9.3 Fragen

1) Ein Mitarbeiter eines städtischen Hallenbades führt eine Untersuchung zum Besuch dieser Einrichtung durch. Dazu notiert er sich über einen Zeitraum von 70 Tagen (=10 Wochen) täglich mehrere Werte, welche davon sind metrische Variable?

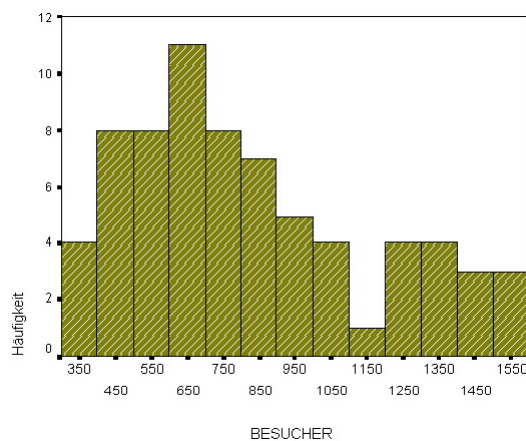
- Wie viel Besucher Insgesamt
- Wie viel Schulklassen
- War Trainingstag für den Schwimmklub
- Welcher Wochentag
- War es ein Regentag
- Wie hoch war die Außentemperatur

- Wie viel Besucher waren im angeschlossenen Cafe-Restaurant
- Wie hoch war der Umsatz im Cafe-Restaurant
- Wie viel Besucher in der Sauna

2) Er errechnet sich einen Mittelwert der Tagesumsätze im Cafe-Restaurant von 1.100.- Euro und einen Median von 1.050.- Euro. Aus welcher Angabe kann er auf den Gesamtumsatz schließen und wie hoch ist dieser?

3) Nur für die 10 Sonntage errechnet er für die Besuchszahlen eine Standardabweichung von 210, analog für die 10 Mittwoche eine Standardabweichung von 140. Was bedeutet dies für den Vergleich von Sonntag und Mittwoch.

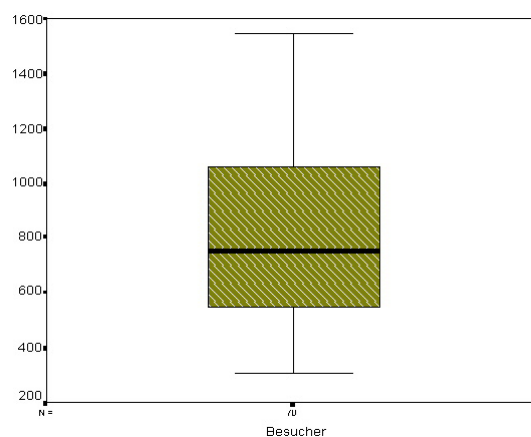
4) Betrachten Sie das folgende Histogramm für die Besucherzahlen.



Welche der folgenden Aussagen dazu sind richtig?

- Es kamen nie mehr als 1600 Besucher.
- Die Verteilung ist in etwa symmetrisch.
- Viermal kamen genau 300 Besucher.
- 19-mal kamen mindestens 1000 Besucher.
- Am häufigsten wurden Besucherzahlen in der Klasse von 600 - 700 beobachtet.

5) Betrachten Sie den folgenden Boxplot für die Besucherzahlen. Welche der folgenden Aussagen dazu sind richtig?



Welche der folgenden Aussagen dazu sind richtig?

- Der Median ist kleiner als 800.
- Ein Viertel der Beobachtungen ist größer als 1000.
- Das Maximum der Besucherzahlen liegt zwischen 1500 und 1600
- Die Hälfte der Beobachtungen ist größer als 800.
- Der Interquartilsabstand ist kleiner als 400.
- Es gibt keine Ausreißer in den Beobachtungen.

6) Um einen Vergleich zum Vorjahr herzustellen, rechnet sich der Mitarbeiter einen t-Test, dessen SPSS-Output hier folgt.

Test bei einer Stichprobe

Testwert = 720			
	T	df	Sig. (2-seitig)
BESUCHER	2,704	69	,009

- Gegen welchen Wert wurde der beobachtete Mittelwert getestet?
- Die lauten Null- und Alternativhypothese des Tests?
- Ist ein signifikanter Unterschied beobachtet worden?

Antworten

1) Metrisch ist eine Variable dann, wenn ihren Werten sinnvoll Zahlen zugeordnet werden können

- Wie viel Besucher Insgesamt (metrisch)
- Wie viel Schulklassen (metrisch)
- War Trainingstag für den Schwimmklub
- Welcher Wochentag
- War es ein Regentag
- Wie hoch war die Außentemperatur (metrisch)
- Wie viel Besucher waren im angeschlossenen Cafe-Restaurant (metrisch)
- Wie hoch war der Umsatz im Cafe-Restaurant (metrisch)
- Wie viel Besucher in der Sauna (metrisch)

2) Er errechnet sich einen Mittelwert der Tagesumsätze im Cafe-Restaurant von 1.100.- Euro und einen Median von 1.050.- Euro. Aus welcher Angabe kann er auf den Gesamtumsatz schließen und wie hoch ist dieser?

Natürlich kann aus dem Mittelwert durch einfache Multiplikation mit 70 der Gesamtumsatz in diesen 70 Tagen berechnet werden. Mit dem Median geht das nicht!

3) Nur für die 10 Sonntage errechnet er für die Besuchszahlen eine Standardabweichung von 210, analog für die 10 Mittwoche eine Standardabweichung von 140. Was bedeutet dies für den Vergleich von Sonntag und Mittwoch.

Es bedeutet, dass am Sonntag die Besucherzahlen stärker variieren (streuen) als am Mittwoch.

4) Betrachten Sie das folgende Histogramm für die Besucherzahlen.

Welche der folgenden Aussagen dazu sind richtig?

- Es kamen nie mehr als 1600 Besucher. Richtig: die letzte Klasse geht von 1500 bis 1600, also gibt es keine Beobachtung über 1600.
- Die Verteilung ist in etwa symmetrisch. Falsch: das Histogramm zeigt kein symmetrisches Bild
- Viermal kamen genau 300 Besucher. Falsch: Es sind 4 Beobachtungen in der Klasse von 300 bis 400. Das heißt aber nicht, dass alle Beobachtungen exakt 300 waren.
- 19-mal kamen mindestens 1000 Besucher. Richtig: die Klassen über 1000 haben die Häufigkeiten: 4 + 1 + 4 + 4 + 3 + 3; zusammen sind das 19.
- Am häufigsten wurden Besucherzahlen in der Klasse von 600 - 700 beobachtet. Richtig: das ist der höchste Balken.

5) Betrachten Sie den folgenden Boxplot für die Besucherzahlen. Welche der folgenden Aussagen dazu sind richtig?

Welche der folgenden Aussagen dazu sind richtig?

- Der Median ist kleiner als 800. Richtig: der Median ist als Unterteilung der Box eingezeichnet. Er ist eindeutig unterhalb von 800 (Skala links) eingezeichnet
- Ein Viertel der Beobachtungen ist größer als 1000. Richtig: die obere Begrenzung der Box (das 3.Quartil) liegt eindeutig über 1000. Aber zwischen dem 3.Quartil und dem Maximum liegt ein Viertel der Beobachtungen.
- Das Maximum der Besucherzahlen liegt zwischen 1500 und 1600. Richtig: der Boxplot endet mit dem Maximum der Daten; dieses ist irgendwo zwischen 1500 und 1600.
- Die Hälfte der Beobachtungen ist größer als 800. Falsch: der Median liegt unter 800. Also reichen ein paar Werte unter 800.
- Der Interquartilsabstand ist kleiner als 400. Falsch: der Interquartilsabstand ist die Differenz zwischen drittem und erstem Quartil. Das 1.Quartil ist kleiner als 600, das 3.Quartil ist größer als 1000. Somit ist die Differenz größer als 400 (der Interquartilsabstand ist die Länge der Box).
- Es gibt keine Ausreißer in den Beobachtungen. Richtig: im Boxplot sind keine Beobachtungen als Ausreißer (durch Kreise oder Sterne) markiert.

6) Um einen Vergleich zum Vorjahr herzustellen, rechnet sich der Mitarbeiter einen t-Test, dessen SPSS-Output hier folgt.

- Gegen welchen Wert wurde der beobachtete Mittelwert getestet? 720
- Die lauten Null- und Alternativhypothese des Tests? Nullhypothese: der Mittelwert in der Grundgesamtheit beträgt 720. Alternativhypothese: der Mittelwert in der Grundgesamtheit beträgt nicht 720.
- Ist ein signifikanter Unterschied beobachtet worden? Ja, die Signifikanz (Sig.) ist 0,009 also kleiner als die üblichen Signifikanzniveaus.

10 Zwei metrische Variablen

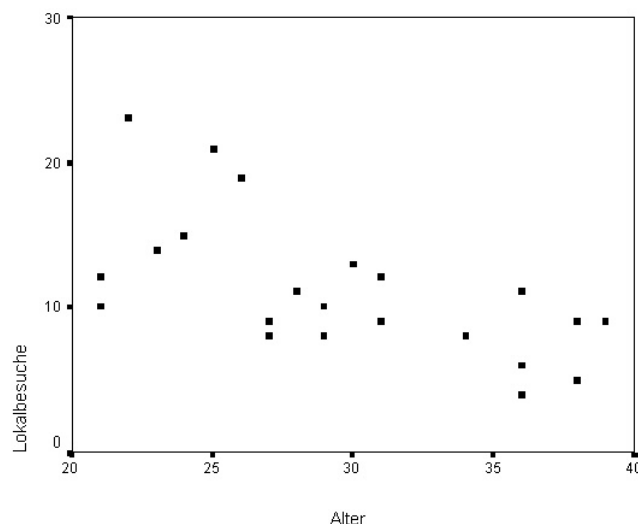
Natürlich können zwei metrische Variable getrennt untersucht werden. In diesem Kapitel interessiert uns aber die Untersuchung des Zusammenhangs zwischen den beiden Variablen. Damit ist gemeint, ob eine Beziehung besteht in der Art: „je größer die Werte der einen Variablen, desto größer im Schnitt auch die Werte der anderen Variablen“ bzw. „je größer die Werte der einen Variablen, desto kleiner im Schnitt die Werte der anderen Variablen“. Im ersten Fall würde man von einem positiven Zusammenhang, im zweiten Fall von einem negativen Zusammenhang sprechen.

10.1 Numerische und grafische Beschreibung

10.1.1 Streudiagramm

Bevor lange gerechnet wird, sollte ein Blick auf die Daten geworfen werden. Diesen Blick richtet man am besten auf ein **Streudiagramm** (engl.: scattergram). Dies ist eine Darstellung der Daten in einem x-y-Diagramm, die Beobachtungen sind als Punkte erkennbar.

Beispiel: 25 Singles im Alter zwischen 20 und 40 wurden über 2 Monate beobachtet. Es wurde registriert, wie oft sie am Abend Lokale besuchen. Ein Streudiagramm der Variablen Alter und Lokalbesuche ergibt folgenden Plot:



Generell besteht ein negativer Zusammenhang: höheres Alter ist durchschnittlich mit weniger Lokalbesuchen verbunden.

10.1.2 Korrelationskoeffizient

Wenn das Streudiagramm in etwa eine lineare Beziehung zwischen den zwei Variablen ableiten lässt, kann der sog. Korrelationskoeffizient nach Pearson berechnet werden. Er misst die Stärke des Zusammenhangs, seine Interpretation basiert auf folgenden Eigenschaften:

- Der Korrelationskoeffizient nimmt nur Werte zwischen -1 und 1 an.
- Das Vorzeichen des Korrelationskoeffizienten gibt an, ob der Zusammenhang positiv oder negativ ist.
- Die Extremwerte -1 und +1 werden nur angenommen, wenn die Punkte im Streudiagramm exakt auf einer Geraden liegen.
- Je größer der Absolutbetrag des Korrelationskoeffizienten (man lässt also das Vorzeichen weg), desto konzentrierter liegen die Punkte um eine (gedachte) Gerade.
- Ist der Korrelationskoeffizient nahe bei 0, ist der Zusammenhang zwischen den Variablen nur schwach.

Rechentechnisch ist die Berechnung schon für wenige Beobachtungen ein Langzeitjob. Wir verzichten auf Formel und widmen uns der Interpretation.

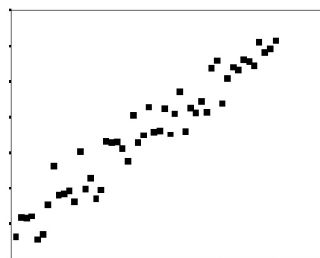
Beispiel: Bei den 25 Singles und deren Lokalbesuchen errechnet sich ein Korrelationskoeffizient von $r = -0,651$. Das Vorzeichen überrascht nicht, es passt zum negativen Zusammenhang. Die absolute Größe von 0.651 ist weder sehr hoch noch sehr niedrig.

10.1.3 Weitere Beispiele

Es werden noch einige Beispiele vorgestellt, um ein Gefühl für die Anwendbarkeit der hier vorgestellten Methoden zu entwickeln.

Stark positiver linearer Zusammenhang

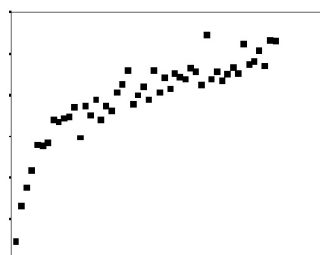
Ideale Bedingungen, $r = 0,963$.



Stark positiver nichtlinearer Zusammenhang

Der Zusammenhang ist nicht linear.

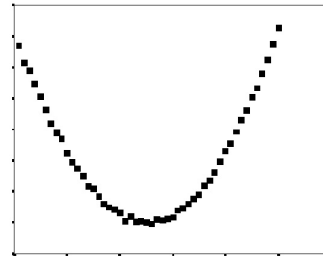
Dennoch hoher Wert für r , $r = 0,866$



Nichtlinearer Zusammenhang

Der Zusammenhang ist nicht linear.

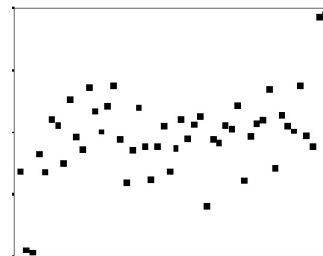
Korrelationskoeffizient nicht passend ($r = 0,073$)



Pseudozusammenhang durch Ausreißer

2 Ausreißer links unten, 2 Ausreißer rechts oben.

Ohne diese Ausreißer kein Zusammenhang.



SPSS-Kommandos:

Streudiagramm: Grafiken -> Streudiagramm

Korrelationskoeffizient: Analysieren -> Korrelation -> Bivariat

10.2 Test für den Korrelationskoeffizienten

Der Korrelationskoeffizient misst den (linearen) Zusammenhang zwischen zwei Variablen in der Stichprobe. Ist dieser Wert nicht ganz nahe bei 0, stellt sich die Frage, ob der Wert signifikant von 0 abweicht.

Bei diesem Test lautet die Nullhypothese, dass der Korrelationskoeffizient 0 ist. Die Alternativhypothese besagt, dass er von 0 abweicht.

Beispiel: Das Beispiel mit den Lokalbesuchen von Singles bringt folgenden SPSS-Output.

Korrelationen

		ALTER	LOKALB
ALTER	Korrelation nach	1,000	-,651
	Pearson		
	Signifikanz (2-seitig)		,000
N		25	25
LOKALB	Korrelation nach	-,651	1,000
	Pearson		
	Signifikanz (2-seitig)	,000	
N		25	25

Interpretation:

- Im Output ist der Wert des Korrelationskoeffizienten ablesbar ($r=0,651$).
- Der Signifikanzwert ($=0,000$) ist sehr klein. Der Wert ist nicht exakt 0, aber auf drei Nachkommastellen unterscheidet er sich nicht von 0.
- Bei einem Signifikanzniveau von 5% ($=0,05$) wird die Nullhypothese verworfen, weil der Signifikanzwert (p-Wert) kleiner als Signifikanzniveau ist.
- Die Entscheidung lautet somit: Der Korrelationskoeffizient weicht signifikant von 0 ab. Da er in der Stichprobe positiv ist, dürfen wir schließen, dass der Korrelationskoeffizient signifikant positiv ist.
- Der Wert von $1,000$ als Korrelationskoeffizient der Variablen ALTER und LOKALB mit sich selbst darf nicht überraschen. Überlegen Sie, wie das Streudiagramm einer Variablen mit sich selbst aussieht.

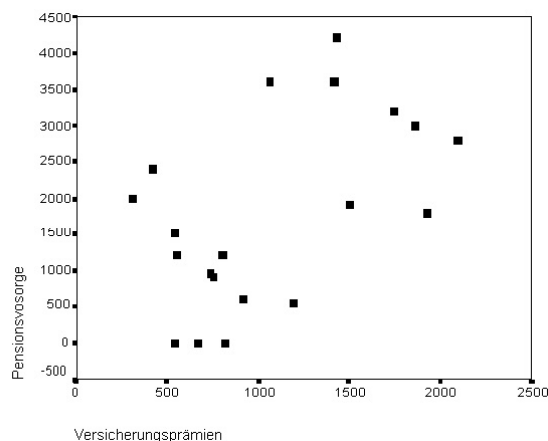
Bemerkung: Natürlich wäre es gerade auch bei diesem Beispiel interessant, konkret die Gerade zu bestimmen, die den Zusammenhang zwischen dem Alter und der Anzahl von Lokalbesuchen beschreibt. Damit wäre eine Prognose aus dem Alter auf die Anzahl von Lokalbesuchen möglich. Allerdings überschreitet dies den Rahmen dieses Kurses. Zur Anwendung kämen Methoden der Regressionsrechnung.

SPSS-Kommandos:

Analysieren -> Korrelation -> Bivariat

10.3 Fragen

Eine Mitarbeiterin einer Versicherung sammelt Daten über einige Kunden, sie summiert deren jährliche Versicherungsprämien und deren jährliche Privatpensionsvorsorge. Von 20 Kunden kann sie folgendes Streudiagramm ableiten:



1) Welche Aussagen kann man aus dem Diagramm ableiten?

- Im Wesentlichen besteht ein positiver Zusammenhang zwischen den beiden Variablen.
- Mehrere Kunden aus der Stichprobe zahlen keine Versicherungsprämien sondern leisten sich nur eine Pensionsvorsorge.
- Der Korrelationskoeffizient wird negativ sein, weil der Zusammenhang nur schwach ist.
- Der Kunde mit den höchsten Zahlungen für Versicherungsprämien zahlt auch am meisten für private Pensionsvorsorge ein.
- Der Kunde mit den niedrigsten Zahlungen für Versicherungsprämien zahlt jährlich ca. 2000.-€ für private Pensionsvorsorge ein.

2) Ein Test für den Korrelationskoeffizienten wird durchgeführt:

Korrelationen

		VERSICHERUNG	PENSION
VERSICHERUNG	Korrelation nach Pearson	1,000	,536
	Signifikanz (2-seitig)	,	,015
	N	20	20
PENSION	Korrelation nach Pearson	,536	1,000
	Signifikanz (2-seitig)	,015	,
	N	20	20

- Wie lauten Null- und Alternativhypothese dieses Tests?
- Welchen Wert hat der Korrelationskoeffizient?
- Ist das Ergebnis signifikant?

Antworten

1) Welche Aussagen kann man aus dem Diagramm ableiten?

- Im Wesentlichen besteht ein positiver Zusammenhang zwischen den beiden Variablen. Richtig, je höher die Versicherungsprämien desto höher durchschnittlich auch die private Pensionsvorsorge.
- Mehrere Kunden aus der Stichprobe zahlen keine Versicherungsprämien sondern leisten sich nur eine Pensionsvorsorge. Falsch, die kleinsten Versicherungszahlungen betragen ca. 300€.
- Der Korrelationskoeffizient wird negativ sein, weil der Zusammenhang nur schwach ist. Falsch, auch wenn der Zusammenhang nicht sehr stark ist, er ist auf jeden Fall positiv. Daher ist auch der Korrelationskoeffizient positiv.
- Der Kunde mit den höchsten Zahlungen für Versicherungsprämien zahlt auch am meisten für private Pensionsvorsorge ein. Falsch, dieser Kunde zahlt zwar ca. 3000 € als

Pensionsvorsorge ein, es gibt aber mehrere, die noch höhere Beträge als Pensionszahlungen leisten.

- Der Kunde mit den niedrigsten Zahlungen für Versicherungsprämien zahlt jährlich ca. 2000.-€ für private Pensionsvorsorge ein. Richtig, das ist der Punkt, der am weitesten links liegt.

2) Ein Test für den Korrelationskoeffizienten wird durchgeführt:

- Wie lauten Null- und Alternativhypothese dieses Tests? In der Nullhypothese wird behauptet, dass es keinen Zusammenhang zwischen den beiden Variablen gibt. Die Alternativhypothese verneint dies, es gibt einen Zusammenhang.
- Welchen Wert hat der Korrelationskoeffizient? 0,536
- Ist das Ergebnis signifikant? Bei einem Signifikanzniveau von 5% schon, bei einem von 1% nicht. Die Signifikanz von 0,015 wird mit dem Signifikanzniveau verglichen. Ist sie größer, wird die Nullhypothese beibehalten, sonst wird sie verworfen.

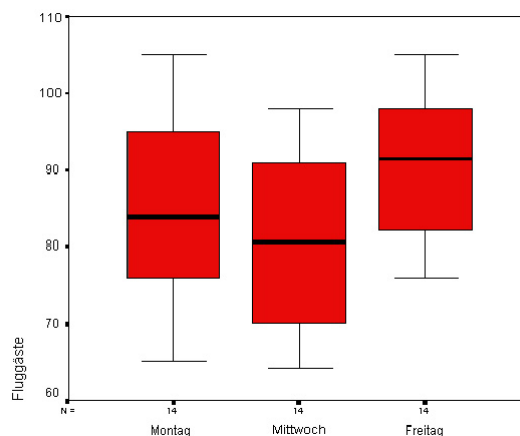
11 Eine metrische und eine kategoriale Variable

Diese Datensituation liegt immer dann vor, wenn eine metrische Variable in mehreren Gruppen beobachtet wird. Die kategoriale Variable dient zur Definition der Gruppen.

11.1 Grafische Beschreibung

Die numerische Beschreibung jeder Gruppe kann mit den üblichen Maßzahlen für metrische Variablen erfolgen. Allerdings ist bei vielen Gruppen kaum ein Überblick über die Daten zu erzielen. Weit besser geht dies, wenn für jede einzelne Gruppe ein Boxplot erstellt wird und diese Boxplots nebeneinander gestellt werden.

Beispiel: Eine Fluggesellschaft fliegt eine Destination seit 14 Wochen an. Die Auslastung für die drei Wochentage, an denen Flüge erfolgen, ist in den drei Boxplots zusammen gefasst.



Als Repräsentant für jeden Wochentag kann der Median, also die Unterteilung der Box, herangezogen werden. Man sieht, dass für diese Stichprobe der Freitag am besten, der Mittwoch am schlechtesten abschneidet. Ein Streuungsmaß, nämlich der Interquartilsabstand, ist als Länge der Box aus der Grafik ablesbar. Die Unterschiede zwischen den drei Wochentagen sind nicht sehr groß.

SPSS-Kommandos:

Analysieren -> Deskriptive Statistiken -> Explorative Datenanalyse

oder direkt über

Grafiken -> Boxplot

11.2 Vergleich der Mittelwerte von zwei Gruppen

Natürlich interessiert uns die Frage, ob die Unterschiede zwischen Gruppen signifikant sind; also die Frage, ob die Unterschiede, die an der Stichprobe beobachtet wurden so groß sind, dass man daraus auch auf Unterschiede in der Grundgesamtheit schließen kann. Wir beschränken uns auf den Mittelwertsvergleich bei zwei Gruppen. Die Nullhypothese besagt, dass kein Unterschied im Mittelwert zwischen den Gruppen besteht, in der Alternativhypothese wird dies verneint, die Mittelwerte sind nicht gleich. Das statistische Verfahren, mit dem der Vergleich durchgeführt wird, nennt sich t-Test.

Beispiel: Im Beispiel mit der Fluggesellschaft liegen drei Gruppen (Wochentage) vor, wir untersuchen, ob sich die Auslastung am Montag von der am Mittwoch signifikant unterscheidet. Im SPSS-Output gibt es zunächst ein paar Kennzahlen für jede Gruppe (Gruppenstatistiken), dann die Ergebnisse des eigentlichen Tests.

Gruppenstatistiken

	WANN	N	Mittelwert	Standardabweichung	Standardfehler des Mittelwertes
FLUGGAST	Montag	14	85,14	12,95	3,46
	Mittwoch	14	81,29	11,78	3,15

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit		df Sig. (2-seitig)	
		F	Signifikanz	T			
FLUGGAST	Varianzen sind gleich	,226	,639	,824	26		,417
	Varianzen sind nicht gleich			,824	25,772		,417

Interpretation:

- Zuerst wird ein Test (Levene-Test) durchgeführt, ob die Varianzen in den beiden Gruppen gleich sind. In diesem Fall (Signifikanz=0,639) kann man davon ausgehen, dass sich die Varianzen nicht unterscheiden.
- Dann wird der Test durchgeführt, ob sich die Mittelwerte unterscheiden. Einmal unter der Annahme gleicher Varianzen in den Gruppen (obere Zeile), einmal unter der Annahme unterschiedlicher Varianzen (untere Zeile). Hier sind die Ergebnisse fast gleich, jedenfalls ist der Signifikanzwert (Sig.) mit 0,417 derselbe für beide Fälle.
- Dieser Signifikanzwert (in anderen Statistikpaketen auch p-Wert genannt) wird mit dem Signifikanzniveau verglichen. Legt man das üblichste Signifikanzniveau zugrunde, nämlich 5%, wird die Nullhypothese beibehalten, da diese 0,417 größer als das Signifikanzniveau sind.
- Aus Formatierungsgründen für dieses Skriptum wurden vom Standard SPSS-Output die vier letzten Spalten weggelassen; diese sind für die Interpretation des Ergebnisses nicht von Bedeutung.

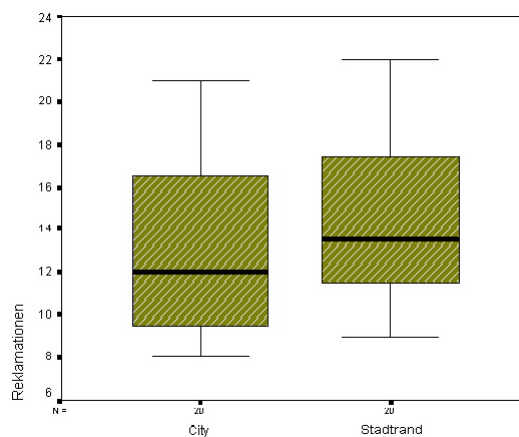
SPSS-Kommandos:

Analysieren -> Mittelwerte vergleichen -> T-Test bei unabhängigen Stichproben

11.3 Fragen

Eine Handelfirma unterhält in einer Stadt zwei Filialen, eine in der City, die andere am Stadtrand. Über 20 Wochen wurden die eingehenden Reklamationen in beiden Filialen registriert.

1) Der Boxplot zeigt eine Zusammenfassung der gesammelten Daten.



Welche Aussagen können aus dem Boxplot abgeleitet werden?

- In der City-Filiale gab es in diesen 20 Wochen durchschnittlich weniger Reklamationen.
- Die Streuung unterscheidet sich deutlich zwischen den zwei Filialen.
- Der Median der Reklamationen in der City-Filiale liegt bei 12.
- Am Stadtrand gab es nie weniger als 10 Reklamationen.
- In beiden Filialen gab es Tage mit mehr als 20 Reklamationen.

2) Ein Test, ob unterschiedlich viel Reklamationen in den beiden Filialen einlangen wurde durchgeführt.

Test bei unabhängigen Stichproben

	Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit		
	F	Signifikanz	t	df	Sig. (2-seitig)
REKLAMAT Varianzen sind gleich	,066	,799	-1,197	38	,239
Varianzen sind nicht gleich			-1,197	37,931	,239

- Wie lauten Null- und Alternativhypothese des Tests?
- Unterscheiden sich die beiden Filialen signifikant?

Antworten

1) Der Boxplot zeigt eine Zusammenfassung der gesammelten Daten.

Welche Aussagen können aus dem Boxplot abgeleitet werden?

- In der City-Filiale gab es in diesen 20 Wochen durchschnittlich weniger Reklamationen. Richtig, die Boxplots haben ein ähnliches Aussehen, nur ist der Plot für die City weiter unten angesiedelt.
- Die Streuung unterscheidet sich deutlich zwischen den zwei Filialen. Falsch, als Streuungsmaß ist der Interquartilsabstand als Länge der Box ablesbar. In der City ist diese Box etwas länger aber nicht viel länger.
- Der Median der Reklamationen in der City-Filiale liegt bei 12. Richtig, das ist die Unterteilung der Box.
- Am Stadtrand gab es nie weniger als 10 Reklamationen. Falsch, es gab zumindest einmal nur 9 Reklamationen.
- In beiden Filialen gab es Tage mit mehr als 20 Reklamationen. Richtig, die Linien zum Maximum gehen bei beiden Boxplots über die 20er-Markierung

2) Ein Test, ob unterschiedlich viel Reklamationen in den beiden Filialen einlangen wurde durchgeführt.

- Wie lauten Null- und Alternativhypothese des Tests? Die Nullhypothese besagt, dass es keine Unterschiede beim Mittelwert der Reklamationen zwischen den zwei Filialen gibt. Die Alternativhypothese besagt, dass es Unterschiede gibt.
- Unterscheiden sich die beiden Filialen signifikant? Nein, die Signifikanz von 0,239 ist größer als die üblichen Signifikanzniveaus (5% oder 1%). Daher wird die Nullhypothese beibehalten.

12 Weitere statistische Verfahren

In dieser Einführung sind neben beschreibenden statistischen Verfahren einige einfache inferenzstatistische Verfahren vorgestellt worden. Es kann aber leicht passieren, dass Fragen auftauchen, die mit diesen wenigen Werkzeugen nicht mehr adäquat behandelt werden können. Zwar ersparen Statistikprogramme wie SPSS das eigenhändige Rechnen und erleichtern den Zugang zu fortgeschrittenen Analyseverfahren. Aber allein die Auswahl unter den vielen Verfahren ist oft nur nach professionellem Rat richtig zu treffen.

Eine kurze Übersicht über Verfahren, die relativ oft im Zuge empirischer Studien zum Einsatz kommen, ist in folgender Tabelle gegeben (eine dichotome Variable ist eine kategoriale Variable mit zwei Ausprägungen). Kursiv geschriebene Methoden wurden zumindest teilweise in diesem Skriptum besprochen.

Abhängige Variable(n)	Unabhängige Variable(n)	Methoden
2 kategoriale	-	Chi-Quadrat Test
Mehrere kategoriale	-	Loglineare Modelle
1 kategoriale	1 kategoriale	<i>Homogenitätstest</i>
1 dichotome	kategoriale und/oder metrische	Logistische Regression
1 metrische	1 kategoriale	<i>t-Test</i> , U-Test, Varianzanalyse
1 metrische	mehrere kategoriale	Höhere Varianzanalysen
1 metrische	kategoriale und metrische	<i>Korrelation</i> , lineare Regression

Daneben gibt es noch eine Reihe weiterer Verfahren, mit denen Beziehungen innerhalb von Beobachtungen und unter den Variablen aufgedeckt werden können.

13 Bibliographie

Babbie E.(1998): The Practice of Social Research. Wadsworth, Belmont CA

Diekmann A.(2002): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. 9.Auflage. Rowohlt, Reinbek

Bortz J., Döring N.(1995): Forschungsmethoden und Evaluation für Sozialwissenschaftler (2.Auflage). Springer, Berlin

Hatzinger R.(2003): Statistik für Sozial- und Wirtschaftswissenschaften. Skriptum an der WU Wien

Keller G., Warrack B.(2000): Statistics for Management and Economics (5th ed). Duxbury, Pacific Grove CA

Jahoda M., Lazarsfeld P., Zeisel H. (1975): Die Arbeitslosen von Marienthal. Erstauflage 1933. Suhrkamp, Frankfurt am Main

Noelle-Neumann E., Petersen T. (1996): Alle, nicht jeder. Einführung in die Methoden der Demoskopie. dtv, München

Die angegebene Literatur ist knapp gehalten, lange Literaturlisten verursachen nur ein schlechtes Gewissen.

Als Grundtext über empirische Sozialforschung kann ich Diekmann empfehlen. Das Buch von Babbie ist in amerikanischer Manier sehr ausführlich, das von Bortz und Döring ist ein deutsches Pendant dazu. Im Buch von Noelle-Neumann und Petersen sind Fragebogen, Stichproben und Befragung interessant dargestellt, das Buch selbst ist aber nicht als Text für empirische Sozialforschung gedacht.

Das Skriptum von Hatzinger ist ein guter Einführungstext in die Statistik, ich habe mich beim Verfassen dieses Skriptums an seine Einteilung des Stoffs gehalten. Das Buch von Keller und Warrack ist ein sehr gutes Statistik-Lehrbuch, vom Stoff ausreichend für einen zweisemestrigen Statistik-Kurs.

Die Studie von Jahoda, Lazarsfeld und Zeisel ist eine der berühmtesten Sozialstudien, auch 70 Jahre nach ihrem Erscheinen noch interessant zu lesen.