

Methodenlehre II

Mitschrift der Vorlesung von
Prof. Dr. Hans-Peter Krüger
im SS 07

Roland Pfister

Bayerische Julius-Maximilians-Universität Würzburg

Inhaltsverzeichnis

0. Vorwort	3
1. Überblick	4
1.1. Erklären vs. Verstehen	4
1.1.1. Verstehen durch.....	4
1.1.2. Bewertung der Hermeneutik	5
1.1.3. Beispiel.....	5
1.2. Erklären durch Gesetze	6
1.2.1. Arten von Gesetzmäßigkeiten	6
1.2.2. Erklären als Antwort auf die „Warum“-Frage?	7
1.2.3. Bedingungsanalyse der Kausalität	8
1.2.4. Kausale und Assoziative Gesetze	8
2. Forschungstypen	10
2.1. Die Problematik des reduktiven Schlusses	10
2.1.1. J. St. Mill: Canon of Induction	11
2.1.2. Das MAX-Kon-Min-Prinzip.....	13
2.2. Die Problematik des regressiven Schlusses	17
2.2.1. Epidemiologie	17
2.2.2. Fall-Kontroll-Studie	18
2.2.3. Verantwortlichkeitsanalyse	19
2.2.4. Kausalschluss: Kriterien	19
2.2.5. Interaktionen.....	19
3. Das methodische Problem der Zeit	20
3.1. Entwicklungspsychologische Methoden	20
3.1.1. Längsschnittsequenz.....	20
3.1.2. Querschnittsequenz.....	20
3.2. Zeitreihen	21
3.2.1. Statistischer Regressionseffekt	21
3.2.2. Ausgangswertgesetz von Wilder	21
3.2.3. Cross-over-Design.....	21
4. Erklären mit Statistik	24
4.1. Einführung: Wahrscheinlichkeit	24
4.1.1. Herstellung von Zufall	24
4.1.2. Prüfung von Zufall	24
4.1.3. Vom Zufall zur Wahrscheinlichkeit.....	25
4.1.4. Rechnen mit Wahrscheinlichkeiten.....	25

4.2. Definitionen von Wahrscheinlichkeit	26
4.2.1. Klassische Wahrscheinlichkeitstheorie	26
4.2.2. Kolmogorov	26
4.3. Grundlagen der Inferenzstatistik	27
4.3.1. Enumeration	27
4.3.2. Zuordnen von Wahrscheinlichkeiten.....	27
4.3.3. Definition der Zufallsvariablen	28
4.3.4. Wahrscheinlichkeitsverteilung der Zufallsvariablen	28
4.3.5. Entscheidungsregel	28
4.3.6. Zusammenfassung	29
4.4. Exakte Tests	31
4.5. Asymptotische Tests	34
4.6. Die Ausgangslage	34
5. Von der Stichprobe zur Population.....	36
5.1. Grundgesamtheiten	36
5.2. Stichproben und Repräsentativität.....	36
5.3. Statistiken und Parameter	37
5.3.1. Momenten-Methode	37
5.3.2. Prinzip der kleinsten Quadrate	37
5.3.3. Maximum Likelihood Methode	38
6. Modelle der Spieltheorie	39

0. Vorwort

Dozent: Prof. Dr. H.-P. Krüger

Termin: Montag, 16:00 – 17:30, Kühle-Hörsaal

Klausur: - Termin: 28.09.07
- Stoff: 60% Forschungsmethoden, je 20% Statistik und Vorlesung

Web: www.izvw.de
User: student
PW: Fahrtauglichkeit

1. Überblick

Der erste Teil der Vorlesung bezog sich vor allem auf die Einordnung der Psychologie in die Ansätze der Adaptation, Präformation und Repräsentation sowie Grundlagen der Messtheorie.

Methoden können in diesem Zusammenhang als Vorgang des Messens und daraus resultierender Verhaltensvorhersage oder -manipulation aufgefasst werden. Oder anders ausgedrückt: Die Wissenschaft soll Sachverhalten 1.) beschreiben und 2.) erklären. Der zweite Teil der Vorlesung bezieht sich auf die Suche nach Gesetzmäßigkeiten, die eine solche Erklärung möglich machen.

Die wissenschaftliche Erklärung beruht nach Hempel und Oppenheimer auf der Einordnung des Besonderen in das Allgemeine über ein allgemeines Gesetz:

Prämisse	Es gibt	→	$\exists x P(x)$
Gesetz	Für alle	→	$\forall x [P(x) \rightarrow C(x)]$
<hr/>			
Consecutio			$\exists x C(x)$

*)

– was die Frage aufwirft, inwiefern in der Psychologie allgemeine Gesetzmäßigkeiten überhaupt existieren.

*) Beispiel: Prämisse: Es gibt Psychologiestudenten. Gesetz: Alle Psychologiestudenten lieben den Bortz. Consecutio: Der Bortz wird geliebt.

1.1. Erklären vs. Verstehen

Wilhelm Dilthey führte hierzu den Unterschied zwischen idiographischen verstehenden (Geistes-) und nomothetischen erklärenden (Natur-)Wissenschaften ein. Die Methode der idiographischen Wissenschaften muss sich demnach von der Messung der nomothetischen Wissenschaften abheben – ein Verstehen als eigener Prozess unabhängig vom Erklären: die Hermeneutik („Kunst der Auslegung“).

Die Hermeneutik besteht aus der Deutung von Sachverhalten aus ihrer Einmaligkeit heraus und umfasst zwei Vorgehensweisen:

- 1.) Bilde keine Hypothesen und suche Quellen (alle Quellen sind zunächst gleichwertig).
- 2.) Verstehen: das Bild fügt sich plötzlich zusammen (Hermeneutischer Sprung), sodass klar wird, welche Informationen richtig und welche falsch waren (Zirkelbezug; Analogie zur nomothetischen vs. idiographischen Methode: junger und alter Kommissar).

In diesem Zusammenhang scheint es notwendig, den Begriff des Verstehens genauer zu definieren. Dabei kann man vier Bedeutungen unterscheiden:

1.1.1. Verstehen durch....

1.1.1.1. ...Kenntnis eines deduktiven Systems

Diese Form des Verstehens entspricht der nomothetischen Methode: Verstehen durch Erklären. Durch Erfahrung entstehen hierbei Regeln, welche automatisch verwendet werden (Faden, Durchmesserlängerung).

1.1.1.2. ...Vorkennnis der Elemente

Ein Beispiel ist eine Übersetzung aus dem Lateinischen: Zwar können einzelne Wörter verschiedene Bedeutungen innehaben, werden jedoch alle bekannten Bedeutungen miteinander kombiniert, so ergibt sich der Sinn des Satzes. Diese Form des Verstehens beruht also ebenfalls nur auf einer einfachen Anwendung von Wissen.

1.1.1.3. ...Verstehen durch Kenntnis der Umstände

Beispiel: Tacitus-Übersetzung zur Kaiserkrönung von Augustus. Nur durch andere (hier: historische) Informationen wird ein Verstehen des Satzes möglich – aber auch hier findet sich kein qualitativ unterschiedlicher Verstehensprozess.

1.1.1.4. „Atmosphäre“, „Stil“, „Charakter“

Das Auswerten projektiver Tests wie des TAT kommt am ehesten an das Verstehen im Sinne der Hermeneutik heran. Doch auch hier sind Hypothesen im Spiel (z.B. Murrays Theorie von Needs und Presses, deren Kombination das Leitthema ergibt) – es kann also auch kein wirklich idiographisches Verstehen angenommen werden.

1.1.2. Bewertung der Hermeneutik

Bei allen angeführten Arten des Verstehens gibt es keinen Hermeneutischen Sprung – Verstehen erfolgt also niemals von selbst. Außerdem beruht der postulierte Hermeneutische Sprung wie erwähnt auf einem Zirkelschluss: Informationen aus verschiedenen Quellen ergeben ein Bild auf dessen Grundlage Informationen die nicht hinein passen abgelehnt werden.

Heidegger führt an, dass eben dieser Zirkelschluss den Kern des hermeneutischen Verstehens ausmacht. Dagegen steht jedoch Stegmüllers radikale Ablehnung des hermeneutischen Zugangs: Dieser geht offensichtlich ex post facto vor und liefert damit höchstens Pseudoerklärungen. Eine tragfähige wissenschaftliche Erklärung muss hingegen in die Zukunft gerichtet sein und eine (testbare) Vorhersage ermöglichen.

1.1.3. Beispiel

Eine (historisch) wichtige Fragestellung für die Psychiatrie ergab sich aus der Debatte um idiographische und nomothetische Methoden: Ist die Diagnose in der Psychiatrie nur über einen Psychiater (idiographisch) möglich oder auch über standardisierte Tests (nomothetisch)?

Meehl (1956) veröffentlichte zu diesem Thema eine Untersuchung (Wanted – a good cookbook), bei der klinische, persönliche und standardisierte Tests systematisch miteinander verglichen wurden. Dabei zeigte sich, dass der Mensch – und damit die idiographische Methode – ein guter Datenerheber ist (klinische Daten verbesserten die Diagnostizität von standardisierten Tests), jedoch ein schlechter Datenverwerter im Vergleich zu diesen Tests.

1.2. Erklären durch Gesetze

Wie schon anfangs erläutert, besteht die wissenschaftliche Erklärung nach Hempel und Oppenheimer in der Einordnung des Besonderen ist das Allgemeine über die Kenntnis eines geltenden Gesetzes.

Prämisse	$\exists x P(x)$
Gesetz	$\forall x [P(x) \rightarrow C(x)]$
Consecutio	$\exists x C(x)$

Erklären kann demnach auf zwei Arten erfolgen: Ex ante (deduktiv; Es liegt eine Prämisse vor und ein allgemeines Gesetz gilt, also ist die Consecutio vorhersagbar) und Ex post (induktiv; Es liegt eine Consecutio vor, ein allgemeines Gesetz gilt, also muss die Prämisse gelten).

Diese Suche nach Gesetzmäßigkeiten wurde unter anderem von Laplace angeführt, um über die Erfassung der Natur in einer Funktion Vorhersagen treffen zu können.

1.2.1. Arten von Gesetzmäßigkeiten

Ein Gesetz ist eine Verknüpfung von Variablen durch ein gemeinsames Prinzip. Es können dabei verschiedene Arten von Gesetzmäßigkeiten unterschieden werden, beispielsweise nach der Skalenqualität der beteiligten Variablen, Allgemeinheitsgrad und Unerschöpflichkeit.

1.2.1.1. Skalenqualität der beteiligten Variablen

Ein Beispiel für ein quantitatives Gesetz ist das Fechnersche Gesetz $R = a * \log(S) + b$ oder auch die Stevens-Potenzfunktion.

Ein ordinales Gesetz ist beispielsweise der Einfluss von kognitiver Dissonanz nach Festinger: Je größer die Dissonanz, desto größer der Druck, sie zu reduzieren.

Ein nominales Gesetz liegt beispielsweise der postdeziSIONalen Dissonanz zugrunde: Wenn eine Person eine Entscheidung zwischen zwei Alternativen getroffen hat, dann entsteht bei ihr Dissonanz.

1.2.1.2. Allgemeinheitsgrad

In wissenschaftlichen Gesetzen muss mindestens ein Allquantor auftreten. Die allgemeinste Formulierung

$$\forall x: U(x) \rightarrow V(x)$$

lässt sich jedoch einschränken auf bestimmte Personen

$$\forall p \in P: U(p) \rightarrow V(p)$$

oder auch andere Variablen, wie Personen, Situationen und Zeitpunkte:

$$\forall (p, s, t) \in P \times S \times T: U(p, s, t) \rightarrow V(p, s, t).$$

Daraus folgt direkt, dass ein Gesetz umso leichter zu widerlegen ist, je allgemeiner es ist. Wird ein Gesetz auf die beobachtete Menge eingegrenzt, erklärt er nichts mehr (Prinzip der Unerschöpflichkeit).

1.2.1.3. Unerschöpflichkeit

Der Allquantor muss sich auf eine Menge beziehen, die prinzipiell unendlich ist. Diese Unendlichkeit kann sich auf die Gruppe der Personen P, die Menge der Situationen S und die Menge der Zeitpunkte T beziehen.

1.2.1.4. Schlussfolgerung

In einem Gesetz werden also UV und AV durch eine funktionale Beziehung miteinander verknüpft. Dabei gelten folgende Forderungen:

- Die Skalenqualität der beteiligten Variablen muss definiert sein.
- Es muss mindestens ein Allquantor im Gesetz enthalten sein.
- Das Gesetz muss sich auf eine prinzipiell unendliche Grundgesamtheit beziehen.

Es stellt sich jedoch die Frage, woraus diese funktionale Beziehung zwischen zwei Variablen resultiert – bzw. ob die Feststellung eines funktionalen Zusammenhangs ein Nachweis einer Kausalität („Warum?“) ist.

1.2.2. Erklären als Antwort auf die „Warum“-Frage?

Leibnitz¹ definierte das existentielle Warum als die alles entscheidende Frage. Dabei führt jede Frage nach einem Warum immer zu einem weiteren, bis der allererste Grund gefunden ist. [„Warum gibt es überhaupt etwas und nicht nichts?“]. Diese Frage nach der ersten Ursache ist nach Heidegger jedoch nur auf Gott zurückzuführen; ein existentielles Warum ist der Wissenschaft also verschlossen.

Die Überlegungen Galileis zu diesem Problem können als der Beginn der modernen Wissenschaft angesehen werden: Er forderte, Dinge/Eigenschaften ohne deren Ursache zu beschreiben² („Wie?“ statt „Warum?“).

Diese neue Ausrichtung wird auch an Galileis Beschreibung des Fallgesetzes deutlich (Schiefe Ebene; Zeitmessung über einen kontinuierlich tropfenden Eimer; Abstände der Tropfen als relatives Zeitmaß).

Das existentielle Warum ist also keine Frage der Wissenschaft, die nach einem „Wie?“ fragt, in dem jedoch ein verstecktes „Warum?“ in Form der angenommenen Kausalität enthalten ist. Im Folgenden soll die Struktur der Kausalität betrachtet werden.

1.2.2.1. Locke, Hume: Empirismus

Nach Auffassung Humes (tabula rasa) ist Kausalität ein Gegenstand der Erfahrung („Nihil est in intellectu quod non prior fuerit in sensu.“). Kausalität als solche ist also nicht sichtbar, sondern wird durch das wiederholte beobachten von „erst das eine, dann das andere“ konstruiert. Kausalität ist also kein notwendiges Prinzip.

Ein Beleg für Kausalität als empirischer Erfahrungsbegriff ist der bedingte Reflex nach Pavlov.

¹ Ratio est in natura, cur aliquid potius existat quam nihil. Id consequens est magni illius principii, quod nihil fiat sine ratione. [Es gibt einen Grund in der Natur, warum etwas existiert und nicht nichts. Dies ist eine Konsequenz des Grundsatzes, der sagt, nichts werde ohne Grund.]

² „Für jetzt verlangt unser Autor nicht mehr, als dass wir einsehen, wie er uns einige Eigenschaften der beschleunigten Bewegung untersucht und erläutert (ohne Rücksicht auf die Ursache der letzteren)...“.

1.2.2.2. Kant: Kausalität als Eigenschaft des Verstandes

Alle Anschauung geschieht in der Zeit, was es notwendig macht, ein Prinzip zu entwerfen, um Zufälligkeit von Notwendigkeit zu trennen. Kausalität wird von Kant daher als Eigenschaft des „Verstandes“ definiert, die eine Funktion der zeitlichen Abstimmung von Ereignissen ist (Michotte: autochthone Kausalität). Sie ist also erfahrungsunabhängig und wird ohne vorherige Erfahrung und wider besseres Wissen wahrgenommen.

Belege aus der Psychologie finden sich beispielsweise in Computersimulationen zu Kugelstößen.

1.2.3. Bedingungsanalyse der Kausalität

Die wichtigste Voraussetzung zur Wahrnehmung von Kausalität ist, dass Ursache und Wirkung in einer eindeutigen zeitlichen Folge aufeinander folgen. Dies ist insbesondere bei Regelkreisen von Bedeutung (also für alle biologischen Systeme), da man hier keine eindeutige Struktur, z.B. im Sinne von UV und AV definieren kann: Einmal ist eine Größe X die Ursache, einmal ist dieselbe Größe X die Wirkung.

1.2.3.1. Die INUS-Bedingung

Eine Ursache ist „an insufficient but necessary part of an unnecessary but sufficient condition“.

Eine Ursache U eines Ereignisses V ist also ein Teil einer Menge S von Bedingungen, wobei U selbst notwendig aber nicht hinreichend für V ist. Die Gesamtbedingung S hingegen ist nicht notwendig aber insgesamt hinreichend für das Auftreten von V.

Beispiel: Die Manipulation einer UV im Labor wirkt sich auf die AV aus. Diese Manipulation ist notwendig, aber ohne den situationalen Kontext der Untersuchung nicht hinreichend (zum Verhalten braucht man nicht nur Frustration, sondern auch einen Boden auf dem man stehen kann, wenn man frustriert bzw. aggressiv ist).

1.2.3.2. Die ceteris paribus - Bedingung

Jede kausale Erklärung muss die Voraussetzungen enthalten, dass alle anderen Bedingungen konstant gehalten wurden (ceteris paribus = unter sonst gleichen Umständen):

$$\forall x: U(x)_{cp} \rightarrow V(x)$$

Dies führt zu den experimentellen Techniken der Konstanthaltung, Kontrolle der Sekundärvarianz und Randomisierung. Prinzipiell gilt jedoch die manipulative Regel: U kann nur dann Ursache von V sein, wenn V ohne U nicht auftritt (Kontrollgruppe).

1.2.4. Kausale und Assoziative Gesetze

Zur Beantwortung der Wie-Frage gibt es kausale und assoziative (korrelative) Gesetzmäßigkeiten. Kausale Gesetze umschließen dabei immer die Zeitfolge Ursache-Wirkung. Diese muss deshalb auch gesondert geprüft werden (Ex ante – Forschung; Experiment).

Assoziative bzw. korrelative Bedingungsstrukturen beschreiben hingegen einen „zeitfreien“ Zusammenhang zwischen Variablen (Ex post – Forschung).

Anmerkung: Wie bereits erwähnt ist die kausale Bedingungsstruktur (U vor V) des Experiments i. d. R. zu einfach, da alle biologischen Systeme Regelkreise sind. Von daher müsste die systemische Betrachtungsweise in die wissenschaftliche Psychologie einfließen.

2. Forschungstypen

Je nach Lesart des Schemas der wissenschaftlichen Erklärung von Hempel und Oppenheimer

Prämisse	$\exists x P(x)$
Gesetz	$\forall x [P(x) \rightarrow C(x)]$
Consecutio	$\exists x C(x)$

entstehen vier Forschungsansätze:

- Deduktiv: Das allgemeine Gesetz ist bekannt (Angewandte Forschung; aus dem Allgemeinen wird das Besondere abgeleitet).
- Reduktiv/Induktiv: Das allgemeine Gesetz wird gesucht (Grundlagenforschung).
- Progressiv: Die Ursache wird gesetzt, der Effekt abgewartet (ex ante – Forschung, Experiment).
- Regressiv: Die Wirkung wird festgestellt, die Ursache wird gesucht (Ex post facto – Forschung).

Dabei verhalten sich deduktive und induktive sowie progressive und regressive Forschung genau entgegengesetzt, sodass durch die Kombination jeweils eines Teils der beiden Paare mit dem eines anderen vier Forschungstypen entstehen:

	Progressiv	Regressiv
Reduktiv	Typ I Labor-/Feldexperimente	Typ II Feldforschung / ex post facto – Forschung (z.B.: Fernsehen und Aggression)
Deduktiv	Typ III Prognostische Diagnostik, z.B. Schulreifetest	Typ IV Vergangenheitsforschung, Anamnese, Juristerei (Kriminalfall), Hermeneutik

2.1. Die Problematik des reduktiven Schlusses

Das Problem des induktiven Schlusses besteht aus der rein logisch nicht zulässigen Verallgemeinerung von einer Anzahl einzelner Beobachtungen auf ein allgemeines Gesetz (gehaltserweiternder Schluss). Dieser Allsatz ist grundsätzlich nur falsifizierbar aber nicht verifizierbar. Bei einem empirischen Beweis kann nur eine und-Beziehung gezeigt werden („Ich tue etwas und dann passiert etwas anderes.“), auf das wenn-dann muss allerdings geschlossen werden (induktiver Schluss).

Popper forderte aufgrund dieser Unmöglichkeit, einen Allsatz zu verifizieren, ein „stetes Bemühen um Bewährung“ der erschlossenen allgemeinen Gesetzmäßigkeit. Es ist unter diesem Gesichtspunkt selbstverständlich von großer Bedeutung, Regeln für einen plausiblen Schluss zu finden.

John Stuart Mill führte daher in seinem Canon of Induction 5 Bedingungen auf, unter denen ein induktiver Schluss zulässig ist. Wichtig ist bei allen 5 Kanons die Unterscheidung zwischen Ursachen (Causes) und Begleitumständen (vgl. INUS-Bedingung).

2.1.1. J. St. Mill: Canon of Induction

2.1.1.1. First Canon: Method of agreement

Die Methode der Übereinstimmung besagt, dass eine Ursache in demjenigen Merkmal einer Situation zu suchen ist, dass in allen Situationen vorhanden ist, die das interessierende Phänomen hervorrufen.

If ...	variable condition	... produce	phenomenon
	A B C D		X
	A E F G		X
	A H I J		X

Then „A“ is causally connected with „X“

Es werden also Situationen miteinander verglichen, bei denen der Effekt immer vorhanden ist. Um möglichst sicher sein zu können, dass das konstante Merkmal tatsächlich verantwortlich ist, sollte die Varianz der Begleitumstände möglichst hoch sein.

2.1.1.2. Second Canon: Method of difference

Wenn eine Situation 1 sich von einer Situation 2 in exakt einem Merkmal A unterscheidet und der gewünschte Effekt in Situation 1 auftritt und in Situation 2 nicht, so kommt das Merkmal A als Ursache in Frage (ceteris paribus).

If ...	variable condition	... produce	phenomenon
	A B C D		X
	B C D		not X

Then „A“ is causally connected with „X“

Um hier exakte Aussagen treffen zu können, sollte die Begeleitvarianz möglichst gering sein (Differenzmethode: Vortest-Nachtest = Behandlungswirkung).

2.1.1.3. Third Canon: Joint method of agreement and difference

Durch die Kombination der ersten beiden Kanons wird das Prinzip der Randomisierung verwirklicht. Die Methode der Differenz ist dann unzulänglich, wenn die Sekundärvarianz hoch ist. Wenn aber ein Merkmal A konstant in einer Reihe von Situation zu finden ist und in diesen Situationen das Phänomen auftritt, in anderen ebenso heterogenen Situationen jedoch nicht, so kann das Merkmal als Ursache interpretiert werden.

If ...	variable condition	... produce phenomenon
	A B C D E	X
	A F G H I	X
	J K L M	not X
	N O P Q	not X
Then „A“ is causally connected with „X“		

Wenn also viele Unterschiedlichkeiten in beiden Gruppen vorliegen, dann ist das Gruppenergebnis als solches durchaus vergleichbar (Prinzip der Erwartungswertgleichheit). Dies wird empirisch über das Prinzip der Randomisation gewährleistet.

Die Erwartungswertgleichheit gilt hierbei nur für die gesamte Gruppe, von daher sind nur Gruppenvergleiche möglich. Je größer die Gesamtgruppe ist, desto geringer ist dabei der Schätzfehler.

2.1.1.4. Fourth Canon: Method of residues

Die Methode der Residuen ist der Grundstein für jede mehrfaktorielle Varianzanalyse. Sind in einer Situation 3 Merkmale vorhanden, wobei bekannt ist, dass zwei davon für andere Phänomene verantwortlich sind, so ist das dritte Merkmal mit großer Wahrscheinlichkeit die Ursache für das interessierende Phänomen.

If ...	variable condition	... produce phenomenon
	A B C	X (Y Z)
And B is known to be the case of Y And C is known to be the case of Z		
Then „A“ is causally connected with „X“		

Der Kanon der Residuen ist jedoch nur dann korrekt, wenn keine Wechselwirkungen zwischen den Merkmalen vorliegen – ihre Effekte also additiv sind.

2.1.1.5. Fifth Canon: Method of concomitant variation

Das Merkmal einer Situation, welches sich analog zum Effekt verändert kommt als Ursache in Frage (bzw. Gezeitenstärke in Abhängigkeit vom Mondzyklus).

If ...	variable condition	... produce phenomenon
	A B C	X
	2A B C	2X
	3A B C	3X
	4A B C	4X
Then „A“ is causally connected with „X“		

Aufgrund dieses Kanons ist also die Untersuchung von funktionellen Zusammenhängen von großer Bedeutung, was durch eine Verwendung von mehrstufigen (>2) UVn sichergestellt werden kann.

2.1.1.6. Fallen des Kanons

Zwar sind die Regeln des Canons of Induction ein guter Anhaltspunkt, es muss jedoch immer beachtet werden, dass ein induktiver Schluss immer fehlschlagen kann (sei es aufgrund von zu wenigen Beobachtungen oder aber auch durch systematische Fehler).

Am Beispiel der Methode der Übereinstimmung:

If... Wodka + Wasser then...betrunken
 If... Scotch + Wasser then...betrunken
 If... Bourbon + Wasser then...betrunken
 If... Brandy + Wasser then...betrunken
 → Wasser ist kausal mit „betrunken“ verknüpft.

2.1.1.7. Die Systematik des Canons

Wird die Ursache variiert, wird der Schluss umso besser, je größer die Begleitvarianz ist (auch wenn ein existierender Effekt dann mit niedrigerer Wahrscheinlichkeit gefunden wird).

Empirisches Handeln ist also gekennzeichnet durch a) die Variation des Interessanten (Method of difference) mit dem Ziel eines möglichst deutlichen Effekts (MAX), b) der Elimination von Störgrößen mit dem Ziel, nur das notwendige zu untersuchen (Method of residues; MIN) und c) dem Konstanthalten des Nicht-Eliminierbaren (KON) mit dem Ziel maximaler Kontrolle.

	Begleit- varianz	Ursachen- variation
Method of agreement	Minimal	keine
Method of difference	Maximal	ja
Joint method	Maximal	ja
Method of residues	Minimal	Keine
Concomitant variation	Maximal	keine

2.1.2. Das MAX-Kon-Min-Prinzip

Das Max-Kon-Min-Prinzip (Kerlinger, 1973) umfasst eine Reihe von Techniken für verschiedene Aspekte des Experiments, die dazu dienen, das Verhältnis von erklärter zu Fehlvarianz zu verbessern.

2.1.2.1. Variation: Maximiere die Primärvarianz

Grundsätzlich sollte die Primärvarianz so weit wie möglich maximiert werden. Dabei gibt es jedoch einige Einschränkungen:

- 1.) Vom Maximum zur Trivialität ist es nur ein kleiner Schritt. Pbn unter dem Einfluss von 3 Litern Bier reagieren z.B. anders als nüchterne.
- 2.) In einem Zwei-Gruppen-Versuchsplan unterstellt die Maximierung einen monotonen Verlauf der UV – viele Zusammenhänge sind jedoch nicht-linear, wie z.B. der U-förmige Zusammenhang von Aktivierung und Leistung im Yerkes-Dodson-Gesetz (1908). In einem Standardexperiment mussten dabei Tanzmäuse ein Drahtgitter (Schocks mit variabler Stromstärke zur Aktivierung) überqueren, um an Futter zu kommen. Sowohl Stromstärke als auch Nahrungsdeprivation beeinflussen hierbei die Leistung der Versuchstiere.

- 3.) Bei der Untersuchung von Funktionalitäten zwischen UV und AV sollten daher immer mehr als zwei Ausprägungen der UV verwendet werden.
- 4.) Ebenfalls ist es wichtig zu kontrollieren, ob die Behandlung tatsächlich in der gewünschten Art und Weise angekommen ist (z.B. kleinerer Gehörgang von Frauen verursacht eine größere subjektive Lautstärke von Lauten über Kopfhörer).

2.1.2.2. Elimination: Minimiere die Fehlvarianz

Alle Techniken der Elimination beziehen sich auf das Ausrichten der Vpn auf das Ziel des Versuches (Untersuchungsfrage). Dazu gilt es insbesondere, Störfaktoren zu lokalisieren und auszuschalten und somit die Gesamtvarianz zu minimieren.

- 1.) Situation: Ort, Zeit, Atmosphäre
- 2.) Vpn: Motivation (soziale Erwünschtheit, Vp-Stunden, „gute“ Vp, Bewertungsangst), Erwartung (Theorie über Zweck des Versuchs, Placebo und Täuschung, Hawthorne-Effekt) und Prozesse (Aktivation, Ermüdung, Lernen).
Der Hawthorne-Effekt kommt aus der ABO-Psychologie (Majo) und wurde in einer Studie zu Arbeitsumgebungen in den Hawthorne-Näheren erstmals beschrieben. Dort zeigte sich, dass die Pbn allein durch das Wissen um die Beobachtung ihr Verhalten drastisch veränderten (Arbeitsleistung steigt bei zunehmend schlechteren Lichtverhältnissen).
- 3.) VL: Erwartung (Pygmalion-Effekt), Vp-VL-Interaktion (Sicherheit, Nervosität, Mann-Frau).

Die zentrale Rolle bei der Elimination nimmt die Instruktion ein, wobei die Instruktion nicht nur den (verbalen) Anweisungsteil sondern auch jeden anderen Versuchsumstand umfasst. Wichtigster Grundsatz: „Bleibe normal, wenn du mit anderen Menschen umgehst – auch im Labor.“

2.1.2.3. Kontrolltechniken

Kontrolltechniken versuchen die Gleichheit der Gruppen zu unterstützen, indem sie die Gesamtvarianz um die Personenvarianz reduzieren. Die Gesamtvarianz nach der Kontrolle ist also gleich der Gesamtvarianz vor dem Anwenden der Kontrolltechniken minus die Personenvarianz.

Bei unabhängigen Gruppen unterscheidet man die Techniken der Blockbildung (experimentelle Zwillinge), Bildung von Schichten und die Einführung von Kovariaten. Bei abhängigen Gruppen (Problem der Übertragungseffekte!) lassen sich folgende Techniken verwenden, um die Faktor Zeit wenigstens bestmöglich konstant zu halten: Permutation (vollständige und unvollständige Zufallspermutation) sowie das Lateinische Quadrat.

Blockbildung: Über einen Vortest werden Pärchen gleicher Merkmalsausprägungen gebildet. Blocken eliminiert also die Unterschiede hinsichtlich des Merkmals anhand dessen geblockt wird.

Schichtbildung (Stratifizierung): In jeder Schicht finden sich im Gegensatz zur Blockbildung mehrere Probanden pro Bedingung. In Schichten können dadurch Wechselwirkungen betrachtet werden: Trennung von Merkmal x Treatment und restlicher Personenvarianz (Trennung von Behandlungs- und Merkmals-/Personenvarianz).

Hieraus resultiert, dass Blocken zwar sehr effektiv ist und einen vorhandenen Effekt mit großer Wahrscheinlichkeit entdeckt, die Ergebnisse sind allerdings nicht einwandfrei interpretierbar. Bei Schichten findet sich weniger Kontrolle und somit ist auch die Wahrscheinlichkeit einen Effekt zu entdecken geringer. Wenn jedoch ein solcher gefunden wird, so ist dieses Ergebnis wesentlich repräsentativer („Aber die Ergebnisse sind lecker.“).

Es kann also ein Kontinuum von Effektentdeckung (Kontrolle, Blockbildung) über Stratifizierung bis hin zu Randomisierung angenommen werden (bzw. von „keine Aussagekraft“ bis „viel“).

Kovariation: Durch die Einführung von Kovariaten kann die Kovarianz zwischen Kovariater und AV bestimmt werden (Regression). Die Gesamtvarianz wird also um die Kovarianz vermindert, sodass in die Analyse der Behandlungsunterschiede nur noch die Abweichungen von der Regressionsgeraden eingehen. Dies führt zum statistischen Verfahren der Kovarianzanalyse.

Diese erfolgt in mehreren Schritten: Zunächst wird die Regressionsgerade berechnet (Kontrollvariable: Abszisse, AV: Ordinate). Der Behandlungseffekt ist derjenige Effekt, der die Abweichung von der Gerade verursacht (+Fehler). Daher werden die Schätzwerte berechnet die aus der Ausprägung der Kovariate geschätzt werden können:

$$\text{z.B.: } AV = 0,96 * K + 1,19$$

Von diesen Werten wird die gemessene Ausprägung abgezogen:

$$X_{\text{neu},i} = \hat{X}_i - X_i$$

Im Anschluss wird ein Signifikanztest mit den entsprechenden Gruppenwerten durchgeführt, der den eventuell vorhandenen Effekt mit höherer Sicherheit aufzeigt als vor der Kovarianzanalyse („Es gibt einen Effekt, wenn ich um die Intelligenz kontrolliere“). Größenordnung: Vor der Kovarianzanalyse ergibt ein exemplarischer t-Test $p = 0.45$, danach $p = 0.07$.

Probleme bei abhängigen Blöcken (Messwiederholung): Wird eine Person mit sich selbst geblockt muss die Messwiederholung einen zeitlichen Verlauf voraussetzen, sodass Zeiteffekt (Übertragung, Lernen...) niemals eliminiert, sondern höchstens konstant gehalten werden können [Problem: „Was ist ein Behandlungs- und was ein Zeiteffekt?“].

Bei einer **vollständigen Permutation** sind die Häufigkeiten der Behandlungen, aber auch die Häufigkeiten aller möglichen Teilfolgen von Behandlungen gleich, sodass Zeiteffekte bestmöglich konstant gehalten werden. Ist $N < k!$ (mit $k = \text{Anzahl der Behandlungen}$) muss auf eine **unvollständige Permutation** zurückgegriffen werden. Hierbei werden aus einer Liste der vollständigen Permutationen per Zufall Permutationen ausgewählt, woraus zumindest eine Gleichheit der Erwartungswerte für alle Häufigkeiten von Behandlungsfolgen resultiert.

Lateinisches Quadrat: Auch das Lateinische Quadrat stellt einen Versuch dar, mit wenigen Vpn auszukommen. Dabei wird ein Quadrat aus $k \times k$ Behandlungen aufgespannt, wobei sowohl die erste Zeile als auch Spalte mit der richtigen Reihenfolge (A, B, C, D) ausgefüllt wird. Bei vier Behandlungen ergeben sich daraus vier Standardquadrate.

Nun wird per Zufall bestimmt, welches Quadrat, welche Spalten- und Zeilenfolge sowie welche Behandlungsfolge verwendet wird. Dabei werden höhere Ordnungen (AB gleich häufig wie BA) jedoch nicht kontrolliert.

1) Aufsuchen der Standardquadrate für ein bestimmtes k

Q1	Q2	Q3	Q4
A B C D	A B C D	A B C D	A B C D
B A D C	B C D A	B D A C	B A D C
C D A B	C D A B	C A D B	C D B A
D C B A	D A B C	D C B A	D C A B

2) Zufall 1: Welches Quadrat? Hier: Q3

3) Zufall 2: Welche Spaltenfolge?
 Hier: 3 2 1 4

C B A D
 A D B C
 D A C B
 B C D A

3) Zufall 3: Welche Zeilenfolge?
 Hier: 1 4 3 2

C B A D
 B C D A
 D A C B
 A D B C

3) Zufall 4: Welche Behandlungsfolge?
 Hier: B3 B4 B2 B1

B2 B4 B3 B1
 B4 B2 B1 B3
 B1 B3 B2 B4
 B3 B1 B4 B2

2.1.2.4. Kontrolltechniken: Beurteilung

Die Stärken und Schwächen der einzelnen Kontrolltechniken (Messwiederholung, Block, Schichten und Randomisieren) sollen kurz dargestellt werden. Bei allen Techniken ist die Stärke der Kontrolle der Personenvarianz direkt proportional zur Wahrscheinlichkeit, einen vorhandenen Effekt zu entdecken, schränkt jedoch umgekehrt die Reproduzierbarkeit der Ergebnisse stark ein.

- Bewertung der Technik nach Kriterium:**
- A) Kontrolle der Personvariablen
 - B) Reproduzierbarkeit in anderen Untersuchungen, externe Validität
 - C) Chance, vorhandene Effekte zu entdecken
 - D) Zahl der Voraussetzungen bei der statistischen Überprüfung

	A	B	C	D	E	F
Messwiederholung	+++	---	+++	+++	+++	+++
Blocken	++	-	++	+	0	0
Schichten	+	0	+	+	0	0
Randomisation	0	++	0	---	-	0

- E) Sparsamkeit an Versuchspersonen, für Blocken/Schichten sind zusätzliche Vortests nötig
- F) Einflüsse aus mehrfacher Applikation von Behandlungen, etwa Lernen, Ermüdung, carry over

2.2. Die Problematik des regressiven Schlusses

Während der induktive Schluss das Problem aufwirft, dass von einer Reihe empirischer Gegebenheiten nicht sicher auf ein allgemeingültiges Gesetz geschlossen werden kann, ist der regressiv Schluss (von der Consecutio zur Prämisse) keine Tautologie (logisch nicht stringent):

$$[C \wedge (P \rightarrow C)] \rightarrow P$$

Die Schlussrichtung des regressiven Schlusses ist also entgegen der Wirkrichtung. Man unterscheidet insgesamt zwei Standardfälle für regressiv Schlüsse:

- Die behauptete Consecutio ist so negativ, dass sie nicht experimentell herbeigeführt werden darf.
- Über Ursache der Consecutio ist noch so wenig bekannt, dass eine experimentelle Manipulation noch nicht möglich ist (Heuristik).

Ein bekanntes Beispiel für eine regressiv Fragestellung stellt die Framingham-Studie dar, die nach der Kleinstadt in New England benannt ist, im Laufe derer beinahe die gesamte Bevölkerung alle 2 Jahre untersucht wurde. Dabei sollten Risikofaktoren für koronare Herzkrankheiten (KHK) aufgedeckt werden. [Die Framingham-Studie ist eine Fall-Kontroll-Studie; siehe 2.2.2].

Ein großes Problem der Framingham-Studie (und anderer epidemiologischer Untersuchungen) ist, dass die Pbn nicht randomisiert den Bedingungen „Erkrankung ja“ vs. „nein“ zugeordnet werden können, sodass eine Erklärung über Drittvariablen niemals entkräftet werden kann.

Ein wichtiger psychologischer Befund dieser Studie ist, dass eine Typ-A-Persönlichkeit einen Risikofaktor für die Ausbildung von KHK darstellt.

2.2.1. Epidemiologie

Die Epidemiologie – Lehre von der Verbreitung von Krankheiten – beruht fast ausschließlich auf regressiv Untersuchungen. Bei manchen Fragestellungen sind zwar auch experimentelle Designs möglich, mit UV: Exposition ja/nein, AV: Inzidenz ja/nein (=relative Häufigkeit).

	Erkrankt	Nicht erkrankt
Exposition ja	a	b
Exposition nein	c	d

Aus den Ergebnissen zu Inzidenz bei Exposition und Inzidenz ohne Exposition kann das relative Risiko errechnet werden:

$$I(ja) = \frac{a}{a+b} \quad I(nein) = \frac{c}{c+d}$$

$$\text{Relatives Risiko: } \frac{I(ja)}{I(nein)}$$

Eine derartige Berechnung des relativen Risikos ist nur bei Zufallsstichproben möglich (progressiv Plänen). Wird eine regressiv Untersuchung durchgeführt, so berechnet man anstelle des relativen Risikos das sog. odds ratio.

In diesem Bereich haben sich zwei verschiedene Untersuchungsdesigns entwickelt: die Fall-Kontroll-Studie sowie die Verantwortlichkeitsanalyse.

2.2.2. Fall-Kontroll-Studie

Die Fall-Kontroll-Studie stellt den häufigsten Fall der Epidemiologie dar. Hier soll von bekannten Consecutios (Erkrankt ja/nein) mit Hilfe einer Theorie über die Verursachung ($P \rightarrow V$) auf die Prämisse (Risikofaktoren) geschlossen werden: regressiver Schluss.

Da die Probanden bei Fall-Kontroll-Studien nicht mehr aus einer definierten Population stammen (sich also theoretisch auch aus verschiedenen Populationen zusammensetzen können) lässt sich das relative Risiko wie bei progressiven Studien nicht mehr berechnen. Daher greift man auf das odds ratio zurück.

2.2.2.1. Odds Ratio

Für das Eintreten eines Ereignisses sei die Wahrscheinlichkeit p gegeben. Dann ist $(1-p)$ die Wahrscheinlichkeit für das Nicht-Eintreten des Ereignisses.

Die Chance (odds) für das Eintreten des Ereignisses ist $\text{odds} = \frac{p}{1-p}$. Für

$\text{odds}=1,5$ lässt sich also folgern, dass die Wahrscheinlichkeit des Eintretens eines Ereignisses 1,5-mal höher ist als die Wahrscheinlichkeit des Nicht-Eintretens.

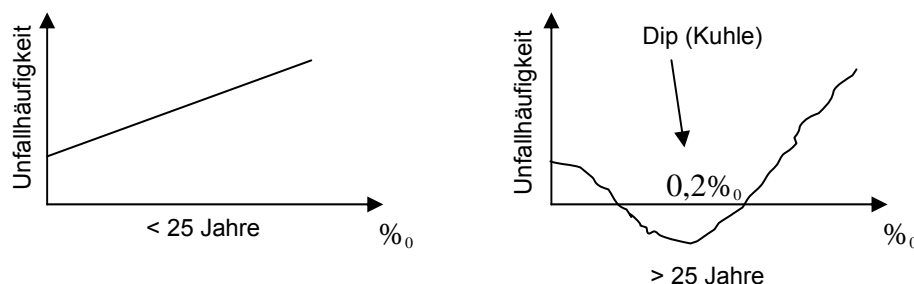
Auf Basis der obigen Vierfeldertafel lassen sich die odds dafür berechnen, dass ein Erkrankter exponiert war (a/c) bzw. ein Nicht-Erkrankter exponiert war (b/d). Das Verhältnis der beiden odds wird als odds ratio bezeichnet.

$$OR = \frac{a \cdot d}{b \cdot c}$$

Das odds ratio ist ein umso besserer Schätzer für das relative Risiko,

- Je eher die untersuchten Fälle repräsentativ sind,
- Je eher die untersuchten Kontrollen repräsentativ sind und
- Je seltener die untersuchte Krankheit auftritt.

2.2.2.2. Exkurs: Promillekurven



2.2.2.3. Attributables Risiko

Das Attributable Risiko (h) beschreibt den Teil des Risikos, der wirklich auf die Tatsache der Exposition zurückzuführen ist. Beispielsweise ist ein Grundunfallrisiko von 19% auch bei Nüchternen zu finden. Wird der Faktor h in jede Promilleklasse eingeführt erhält man den sog. Überschuss, der angibt, wie stark der „bereinigter“/„kontrollierter“ Einfluss der Exposition ist.

2.2.3. **Responsibilitätsanalyse**

Auch die Responsibilitätsanalyse dient der Bestimmung des odds ratio. Es wird hier am Beispiel eines Verkehrsunfalls davon ausgegangen, dass einer der beteiligten Fahrer der Verursacher ist, der andere der Unschuldige. Beide Personen befinden sich natürlich zur gleichen Zeit am gleichen Ort in der gleichen Situation, sodass sie als experimentelle Zwillinge angesehen werden können.

Das odds ratio berechnet sich durch den Vergleich von relativen Häufigkeiten gegebener Exposition bei Verursachern und Unschuldigen:

$$OR = \frac{Verursacher(mit)}{Unschuldige(mit)}$$

Für Marihuanakonsum findet sich durchaus ein Zusammenhang mit Unfallrisiken, wenn zwischen aktiver Substanz (THC) und metabolisiertem Rest (THCOOH) unterschieden wird. Zudem sind Wechselwirkungen zwischen verschiedenen Substanzen zu beachten.

2.2.4. **Kausalschluss: Kriterien**

Trotz der diversen Probleme, die ein regressiver Schluss unausweichlich beinhaltet, lassen sich bestimmte Kriterien aufstellen, bei deren Vorliegen, eine kausale Interpretation möglich wird.

- 1.) Die Ursache muss zeitlich vor der Wirkung liegen
- 2.) Je stärker ein Zusammenhang (relatives Risiko/odds ratio), desto wahrscheinlicher ist eine kausale Beziehung.
- 3.) Beziehung zwischen Dosis/Stärke und Wirkung
- 4.) Replizierbarkeit (eine Studie ist keine Studie)
- 5.) Biologische Plausibilität
- 6.) Berücksichtigung alternativer Erklärungen
- 7.) Das Erkrankungsrisiko sollte sinken, wenn die Exposition beendet ist.
- 8.) Spezifität des Zusammenhangs: Ausschluss anderer Ursachen
- 9.) Konsistenz: Übereinstimmung mit anderen Ergebnissen

2.2.5. **Interaktionen**

Liegen mehrere Ursachen vor und unterscheidet sich die tatsächliche von der erwarteten Wirkung, so liegt eine Interaktion vor. Ist die Wirkung größer als erwartet, so spricht man von einer positiven Interaktion (Synergismus), ist sie kleiner als erwartet, so spricht man von einer negativen Interaktion (Antagonismus).

3. Das methodische Problem der Zeit

Alle Untersuchungen werden durch den Faktor Zeit beeinflusst. Besonders deutlich sieht man dies in der Entwicklungspsychologie, die sich sozusagen der Erforschung dieses Faktors verschrieben hat.

Ein Beispielsbefund beschreibt die Entwicklung der Tragehaltung von Kindern. Es wird dabei zwischen Typ-I (Brustträger) und Typ-II (unter dem Arm) unterschieden. Es zeigt sich, dass Jungen immer Typ-II-Träger sind, Mädchen ebenfalls als solche beginnen, aber ab dem 6./7. Lebensjahr zum Typ-I-Träger werden.

3.1. Entwicklungspsychologische Methoden

Nach Schaie benötigt man zu Interpretation entwicklungspsychologischer Befunde immer drei Arten von Informationen: Alter, Kohorte und Messzeitpunkt. Da sich jedoch immer einer der drei Faktoren durch die anderen beschreiben lässt, genügt die Erfassung von zwei Faktoren (Baltes): $X_t=f(A,G)$. [G = Generation]

Hieraus ergeben sich die drei typischen Designs der Entwicklungspsychologie: Querschnitt-, Längsschnitt- und Zeitwandelmethode. Letztere eignet sich vor allem zur Erfassung von Kohorteneffekten: Personen einer bestimmten Altersgruppe werden in jeder Kohorte untersucht („immer alle 80-Jährigen“).

Ein Beispiel für einen Kohorteneffekt stellt Emile Durkheims (1897) Studie „Der Selbstmord“ dar (Beginn der empirischen Soziologie). Hier zeigte sich, dass die Selbstmordraten drastisch sinken, wenn die Gemeinschaft durch Kriege oder Krisen gefordert wird.

Eine Möglichkeit, die Nachteile von Quer- und Längsschnittmethode zu kompensieren stellen Quer- und Längsschnittsequenzen dar.

Kohorte	1880	1880	1900	1920	1940	1960
	1900	1900	1920	1940	1960	1980
	1920	1920	1940	1960	1980	2000
	1940	1940	1960	1980	2000	2020
	1960	1960	1980	2000	2020	2040
	1980	1980	2000	2020	2040	2060
	2000	2000	2020	2040	2060	2080
	0	20	40	60	80	
	Alter					

Kohorte	1880	1880	1900	1920	1940	1960
	1900	1900	1920	1940	1960	1980
	1920	1920	1940	1960	1980	2000
	1940	1940	1960	1980	2000	2020
	1960	1960	1980	2000	2020	2040
	1980	1980	2000	2020	2040	2060
	2000	2000	2020	2040	2060	2080
	0	20	40	60	80	
	Alter					

3.1.1. Längsschnittsequenz

Eine Längsschnittsequenz besteht aus zwei oder mehr unabhängigen Längsschnittuntersuchungen, sodass hier neben Entwicklungseffekten auch der Einfluss der Kohorte erfasst wird.

3.1.2. Querschnittssequenz

Zu mindestens zwei Zeitpunkten werden mindestens zwei Kohorten querschnittlich untersucht.

Ein großes Problem beider Sequenzmethoden ist der mit ihnen verbundene extrem hohe Aufwand.

3.2. Zeitreihen

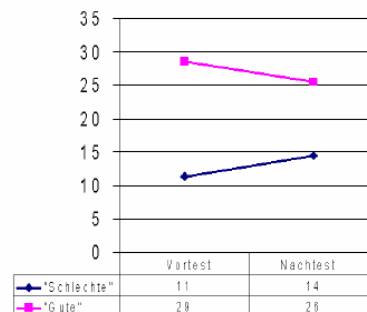
Einige besondere Probleme bei der Untersuchung von Zeitreihen sollen im Folgenden vorgestellt werden. [Interessanter Befund: Menschen sterben eher kurz nach ihrem Geburtsmonat als davor]. Hier sollen vor allem der statistische Regressionseffekt sowie das Wildersche Ausgangswertgesetz betrachtet werden.

3.2.1. Statistischer Regressionseffekt

Der statistische Regressionseffekt (=Regression zur Mitte) tritt auf, wenn Versuchsgruppen nach einer Messung in hohe/niedrige Merkmalsausprägungen unterteilt werden. Wird eine erneute Messung durchgeführt, so zeigen die „besseren“ Probanden eine „Verschlechterung“ und umgekehrt.

Dieser Effekt lässt sich auch demonstrieren, wenn statt realen Messwerten zwei Reihen von Zufallszahlen verwendet werden (anwendbar v.a. auf Ordinaldaten).

	T1	T2	
m=8 ←	8	7	→ m=5,8
	3	3	
m=3 ←	2	4	→ m=5,2
	9	2	
	5	10	
	1	8	
	10	6	
	7	5	
	4	1	
	6	9	



→ Einziger Schutz: Kontrollgruppe und Differenz bilden

3.2.2. Ausgangswertgesetz von Wilder

[Exkurs: Wilder war ein Österreicher und ist in der USA ausgewandert].

Dem Wilderschen Ausgangswertgesetz liegt die Auffassung von biologischen Systemen als Regelsysteme zugrunde, und damit die Annahme einer höheren Regelung bei stärkerer Abweichung bzw. Auslenkung.

Das Ausgangswertgesetz von Wilder beschreibt also eine negative Korrelation von Ausgangswert und Differenzwert. Somit sieht der Effekt genauso aus wie der statistische Regressionseffekt, es liegt aber ein echter empirischer Effekt zugrunde.

Zur Kontrolle beider Effekte muss also eine Kontrollgruppe eingeführt werden. Der Effekt wird dann gemessen, indem man pro Schicht die Veränderung durch die Behandlung (Behandlung – Kontrollgruppe) berechnet und zwischen den Schichten vergleicht.

3.2.3. Cross-over-Design

Das Cross-over-Design besteht aus der Variation der Behandlungsfolgen, um Zeiteffekte möglichst gut kontrollieren zu können. Es werden als zwei unabhängige Versuchsgruppen eingesetzt, an denen jeweils zwei abhängige Messungen vorgenommen werden.

	Periode 1	Periode 2
Behandlungsfolge 1 - 2	Mittel B1/P1 = A	Mittel B2/P2 = B
Behandlungsfolge 2 - 1	Mittel B2/P1 = C	Mittel B1/P2 = D

Jeder Messwert Y wird demnach von verschiedenen Faktoren beeinflusst:

$$Y = M + P + B + R + e$$

Mit M = allgemeines Mittel

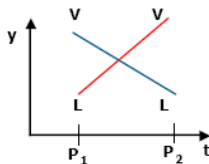
$$P = \text{Periodeneffekt} = [(A-C) + (B-D)] / 2$$

$$B = \text{Behandlungseffekt} = [(A-B) + ((D-C))] / 2$$

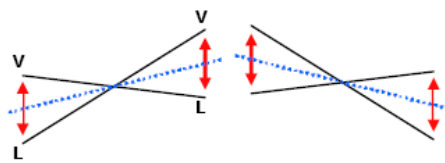
$$R = \text{Residualeffekt} = (A+B) - (C+D)$$

e = Fehler

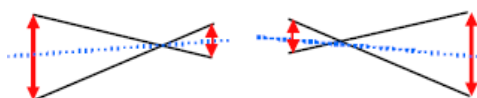
Werden nun die Behandlungen L und V und die Perioden P1 und P2 verwendet, können sich folgende Ergebnisse einstellen:



Nur Behandlungseffekt, kein Perioden- und kein Residualeffekt



Behandlungs- und Periodeneffekt, kein Residualeffekt (der Behandlungsunterschied ist in beiden Perioden gleich).



Behandlungs-, Perioden- und Residualeffekt mit verringertem Behandlungsunterschied (links) und vergrößertem Behandlungsunterschied (rechts).

3.2.3.1. Periodenunterschiede

Periodenunterschiede können zurückgehen auf gleiche externe Einflüsse (z.B. Verschlimmerung der Krankheit) oder aber auch auf gleiche interne Effekte (gleicher Lernzuwachs in beiden Gruppen).

3.2.3.2. Residualunterschiede

Residualunterschiede bedeuten, dass Behandlungsunterschiede in der ersten Periode anders ausfallen als in der zweiten. Dabei kann sich der Unterschied verringern (positiver Residualeffekt) oder vergrößern (negativer Residualeffekt).

3.2.3.3. Behandlungsunterschiede

Behandlungsunterschiede bedeuten, dass die Wirkungen der beiden Behandlungen unabhängig von der Periode sind. Liegen keine Residualunterschiede vor, so ist der Behandlungsunterschied interpretierbar.

Liegen positive Residualunterschiede vor (Behandlungseffekt in der zweiten Periode geringer als in der ersten), wird der Behandlungseffekt unterschätzt. Das Verfahren ist konservativ aber erlaubt.

Liegen negative Residualunterschiede vor, wird der Behandlungseffekt überschätzt. Das Verfahren ist radikal und nicht erlaubt.

4. Erklären mit Statistik

4.1. Einführung: Wahrscheinlichkeit

Eine Prüfung mit statistischen Verfahren (Inferenzstatistik) baut auf der Schlussweise des Modus tollens auf: Aus dem Nicht-Vorhandensein einer Wirkung wird auf das Nicht-Vorliegen einer Ursache geschlossen. Hierbei werden jedoch immer probabilistische Aussagen getroffen, sodass definiert werden muss, was unter einer Wahrscheinlichkeit bzw. dem Zufall allgemein zu verstehen ist.

Popper definierte einen Zufallsprozess als Prozess, dessen Ergebnis nicht vorhersagbar ist (messtechnische Definition). Dabei wird der mikroskopische Zufall in Form (fast) unendlich vieler (fast) unendlich kleiner Prozesse in endlich viele makroskopische Realisationen abgebildet. Der einzelne Ausgang ist dabei niemals bestimmbar, wohl aber die Menge aller möglichen Ausgänge.

Der Molekularbiologe Jaques Monod argumentierte gegen diese Art von völligem Zufall. Aus seiner Sicht kann die Kausalität, die in der Welt allgegenwärtig ist, nicht auf Zufallsprozessen beruhen (Beispiel: 2 70jährige finnische Zwillinge werden am gleichen Tag fast an derselben Stelle beim Radfahren von einem Laster erfasst). Alle Kausalketten sind nach Monod so unwahrscheinlich, dass an einem völligen Zufall gezweifelt werden muss.

Es liegen also zwei unterschiedliche Definitionen des Zufalls vor, eine post-hoc-Definition (Monod) sowie eine ante-hoc-Definition (Popper). Die Frage nach der Wahrscheinlichkeit eines empirischen Ereignisses („Wie wahrscheinlich ist es, dass auf der Erde Leben entstanden ist?“) verbietet sich jedoch selbst: Empirische Ereignisse gibt es sicher; damit beträgt ihre Wahrscheinlichkeit 1.

[Nebenbei: Natürlich ist die Aussage „Das empirische Ergebnis hat eine Wahrscheinlichkeit von $p = \dots$ “ als Resultat eines Hypothesentests genauso unsinnig. Vielmehr wird eine Wahrscheinlichkeit dafür berechnet, welche Wahrscheinlichkeit das Ergebnis unter Annahme der H_0 hätte.]

4.1.1. Herstellung von Zufall

Zufall ist, wofür keine Regel angegeben werden kann. Somit ist die Herstellung von Zufallsprozessen praktisch unmöglich. Es gibt jedoch einige Versuche, hinreichend komplexe Strukturen zugrunde zu legen, sodass zumindest keine Vorhersage mehr möglich ist.

Ein Beispiel ist die Mitte der Quadratzahl-Methode von John von Neumann, die auch hinter der RND-Funktion von Taschenrechner steht.

Letztlich ist die Herstellung von echtem Zufall jedoch nicht möglich.

4.1.2. Prüfung von Zufall

Zwar ist die Herstellung von Zufall nicht möglich, wenn jedoch ein Prozess abgelaufen ist, lässt sich prüfen, ob dieser zufällig war. Eine zufallserzeugende Funktion ist dabei so lange zufällig, bis durch einen Test ihre Nicht-Zufälligkeit nachgewiesen ist. Möglichkeiten sind nach Owen:

- Häufigkeitstest: Jede Zahl gleich häufig?
- Pärchentest: Jedes Pärchen gleich häufig?
- Poker-Test: Bildung von Quartetten

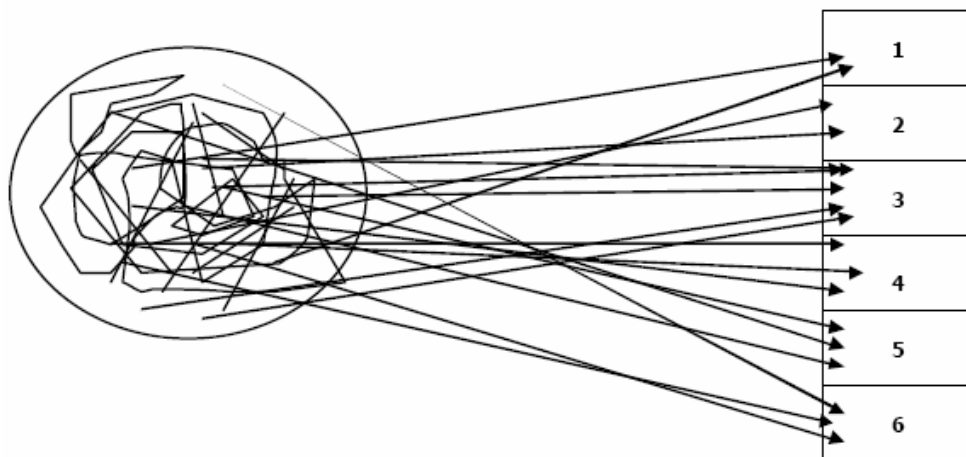
Es liegt also eine operationale Zufallsdefinition vor: „Bisher ist es noch nicht gelungen, eine Regel festzustellen.“

Exkurs: Über den Häufigkeiten-, Pärchen-, Drillings-Test usw. lässt sich auch die Intelligenz einer Person prüfen. Hier sollten alle möglichen Kombinationen von $n \cdot (0,1)$ in einer Reihe von n Nullen und Einsen gleich häufig vorkommen. Je länger die Reihe ist, desto besser sollten sich die Ereignisse an eine Gleichverteilung anpassen:

010101010101111011011100000010101101001011...

0: 50%	00: 25%	000: 12,5%	...
1: 50%	01: 25%	001: 12,5%	...
	10: 25%
	11: 25%	111: 12,5%	...

4.1.3. Vom Zufall zur Wahrscheinlichkeit



Ein System unbekannter Einflüsse oder prinzipiell unbestimmbarer Natur (Zufall) wird über eine unbekannt Abbildungsregel in eine Menge von empirischen Realisationen abgebildet, wodurch sich eine bestimmte Wahrscheinlichkeit dieser Realisationen ergibt.

[Nebenbei: In Berlin existiert eine Firma, die ideale Würfel herstellt.]

4.1.4. Rechnen mit Wahrscheinlichkeiten

„Zufall ist das, was nicht vorausgesagt werden kann (ex ante). Ex post ist Zufall aber wunderbar regelhaft und damit prüfbar.“

Um Wahrscheinlichkeiten zu berechnen, können für

n = Anzahl der zu kombinierenden Elemente und

k = Ordnung oder Klasse der Kombinationen

die nebenstehenden Formeln angewendet werden. Die beiden entscheidenden Faktoren sind hier Berücksichtigung der Anordnung und Berücksichtigung von Wiederholungen (Zmz vs. ZoZ).

		Wiederholungen zulässig?	
		Ja	Nein
Berücksichtigung der Anordnung?	Ja	n^k	$\frac{n!}{(n-k)!}$
	Nein	$\binom{n+k-1}{k}$	$\binom{n}{k}$

4.2. Definitionen von Wahrscheinlichkeit

Es lassen sich drei Definitionen von Wahrscheinlichkeit unterscheiden: Induktive Wahrscheinlichkeit sowie klassische und axiomatische deduktive Wahrscheinlichkeit.

Induktive Wahrscheinlichkeit	Deduktive Wahrscheinlichkeit	
	klassisch	axiomatisch
Wahrscheinlichkeit als Grenzwert der relativen Häufigkeit (Gesetz von Mises).	Wahrscheinlichkeit als Verhältnis von Günstigen und Möglichen Fällen (Pascal)	Zuordnung von Funktionswerten zu Teilmengen des Stichprobenraumes (Kolmogorov)
$P(A) = \lim_{n \rightarrow \infty} Q(A)$	$P(A) = \frac{G}{M}$	$P(A): A \rightarrow f(A)$

4.2.1. Klassische Wahrscheinlichkeitstheorie

Die Wahrscheinlichkeit nach Pascal besteht aus dem Verhältnis der Anzahlen von günstigen (g) und möglichen Fällen (Enumeration; m): $p = g/m$.

Insgesamt beruht Pascals Wahrscheinlichkeitstheorie auf dem Prinzip der Indifferenz (auch: Prinzip vom unzureichenden Grund), dass für $n > 1$ unterscheidbare und sich gegenseitig ausschließende Ereignisse die Wahrscheinlichkeit jedes Ereignisses als $1/n$ annimmt, sofern keine weiteren Informationen verfügbar sind.

Diese sog. Laplace-Wahrscheinlichkeit nimmt also eine diskrete Gleichverteilung aller Ereignisse an. Dieses Prinzip führt aber schnell zu Irrtümern, wenn die betrachteten Ereignisse nicht unabhängig voneinander sind oder den Ereignisraum nicht vollständig abdecken.

Diese Definition ist deduktiv, sodass auch der Begriff „logische Wahrscheinlichkeit“ verwendet wird. Voraussetzung ist hierbei, dass alle Fälle gleich wahrscheinlich sind.

Probleme treten auf, wenn die (Elementar-)Ereignisse nicht gleich wahrscheinlich³ sind, oder die Menge der möglichen Fälle nicht abzählbar ist.

4.2.2. Kolmogorov

Die Probleme des klassischen Ansatzes werden durch Kolmogorovs axiomatische Definition gelöst. Hier werden drei Voraussetzungen genannt, um von einer Wahrscheinlichkeit sprechen zu können:

- 1.) Festlegung des Stichprobenraumes Ω , in dem jedes Ereignis enthalten sein muss.
- 2.) Definition von Teilmengen A_i in Ω , die im Fall der klassischen Wahrscheinlichkeit disjunkt sind und den Stichprobenraum ausfüllen.
- 3.) Zuweisung von Funktionswerten zu den Teilmengen, sodass gilt:

$$P(A) \geq 0$$

Nicht-Negativität

³ Anmerkung: Was heißt überhaupt „gleich wahrscheinlich“? → Zirkeldefinition!

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \text{bzw.}$$

$$P(A \cup B) = P(A) + P(B) \text{ wenn } A \text{ und } B \text{ disjunkt: } \textit{Sigma-Additivität}$$

$$P(\Omega) = 1 \quad \textit{Normierung}$$

Durch die Kolmogorov-Axiome können auch abzählbar große Mengen (Stichprobenräume) in Teilmengen unterteilt und so mit einer Wahrscheinlichkeit versehen werden.

4.3. Grundlagen der Inferenzstatistik

Die Inferenzstatistik bildet das empirische Handeln nach und fragt, welche Ergebnisse unter der Annahme eines Zufallsprozesses anstatt von realen Personen, Treatments, etc. zu erwarten wären.

Dazu wird ein Modell konstruiert, das alle möglichen Ergebnisse umfasst, dabei aber nur unter Zufallsbedingungen zustande gekommen ist. In der Konsequenz heißt das, dass jedes empirische Ergebnis auch Resultat eines Zufallsprozesses sein kann.

Mit Hilfe der Inferenzstatistik soll nun geprüft werden, wie wahrscheinlich das gegebene empirische Ereignis oder ein extremeres durch den angenommenen Zufallsprozess zustande gekommen ist. Ist diese Wahrscheinlichkeit zu klein verliert die Zufallsannahme ihre Glaubwürdigkeit (die Nullhypothese wird verworfen).

Am Beispiel der englischen Lady soll das Vorgehen der schließenden Statistik beschrieben werden.

Insgesamt lassen sich fünf Stufen unterscheiden: Enumeration (Stichprobenraum), Zuordnen von Wahrscheinlichkeiten, Definition der Zufallsvariablen, Erstellung der Wahrscheinlichkeitsverteilung der Zufallsvariablen sowie das Festlegen einer Entscheidungsregel.

4.3.1. Enumeration

Zunächst muss bestimmt werden, welche Ereignisse überhaupt auftreten können. Dies wird als Enumeration bezeichnet [Pascal: Auflistung der möglichen Fälle; Kolmogorov: Festlegen des Stichprobenraumes].

Bei der Enumeration tauchen alle nur denkbaren Möglichkeiten auf. Das heißt, dass das empirische Ergebnis mit Sicherheit unter den aufgelisteten Möglichkeiten ist. Voraussetzung der statistischen Prüfung ist also, dass jedes empirische Ergebnis auch durch Zufall entstanden sein kann.

Wichtig: Für die Erstellung eines Zufallsmodells braucht man immer Informationen aus der Empirie. Also beispielsweise: „Wie viele Tassen Tee gebe ich der englischen Lady?“.

4.3.2. Zuordnen von Wahrscheinlichkeiten

Unter der Annahme, dass es kein Kriterium gibt, das die Fälle unterscheidbar macht, sind alle Fälle des Zufallsmodells gleich wahrscheinlich (Prinzip der Indifferenz).

4.3.3. Definition der Zufallsvariablen

Alein auf Basis der initialen Fragestellung muss überlegt werden, was den Untersucher an den möglichen Fällen interessiert (Zahl der Richtigen oder Falschen, Reihenfolgen etc.).

Ist dies geklärt, wird die Menge der möglichen Ereignisse auf eine neue Variable reduziert. Diese Zufallsvariable ist dabei weder zufällig (weil sie schließlich willkürlich festgelegt wird) noch zwingend variabel ☺. [Synonyme: Prüfgröße oder Teststatistik]

Die Wahl der Zufallsvariable (und damit die initiale Fragestellung) führt dabei zu unterschiedlichen statistischen Tests.

4.3.4. Wahrscheinlichkeitsverteilung der Zufallsvariablen

Die Wahrscheinlichkeitsverteilung der Zufallsvariablen wird auch „Verteilung unter H_0 “ oder „Nullverteilung“⁴ bezeichnet.

Der Würfel (Zufallsmodell) bringt also nicht ein Ergebnis, sondern gleich eine ganze Verteilung.

4.3.5. Entscheidungsregel

Eine Entscheidungsregel lässt sich beispielsweise finden, indem ein **Kriterium** aus der Prüfgröße definiert wird (z.B. die englische Lady muss in mindestens 90% der Fälle richtig liegen).

Bei diesem Vorgehen wird eine neue Variable (groß) X gebildet, welche die Zufallsvariable in einen Annahme- (Für-) und einen Ablehnungsbereich (Gegen-Bereich) abbildet. Für diese Variable X ist dann eine Wahrscheinlichkeit anzugeben.

Eine andere Möglichkeit, zu einer Entscheidungsregel zu kommen, bezieht sich auf das **Modell** selbst. Auf der Menge der möglichen Ausgänge (= Ergebnisse) wird ein Bereich definiert, der dann zur Ablehnung des Modells führt, wenn in ihm auch das empirische Ergebnis liegt.

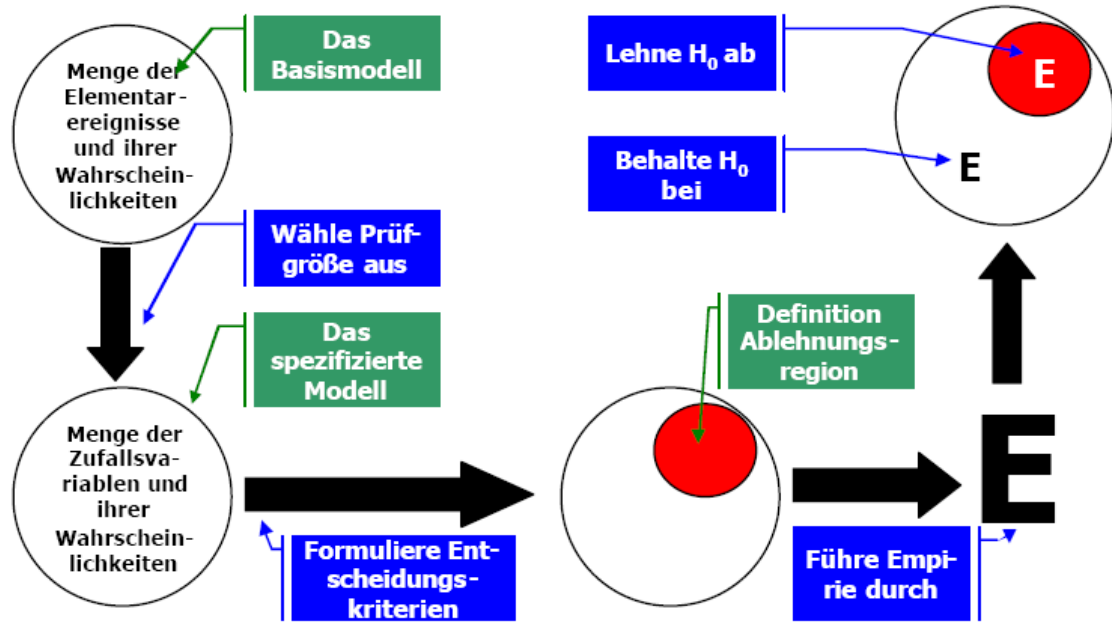
Man ermittelt also einen Ablehnungsbereich. Liegen empirische Ergebnisse in diesem Ablehnungsbereich, so sprechen diese stark gegen die Nullhypothese und führen dazu, dass diese verworfen wird (Modus tollens).

Nun stellt sich jedoch die Frage, welche $\alpha\%$ der Ergebnisse den Ablehnungsbereich bilden. Zum einen wird (willkürlich!) ein Kriterium festgelegt, wie breit der Ablehnungsbereich sein soll (also z.B. 5%), zum anderen muss auf Basis der Fragestellung bzw. der Zufallsvariablen ermittelt werden, wo in der Nullverteilung der Ablehnungsbereich lokalisiert ist.

Zusammenfassend lässt sich also sagen, dass die H_0 zur Erstellung der Verteilung verwendet wird, die H_1 hingegen zur Definition des Ablehnungsbereichs.

⁴ Die Bezeichnungen Nullverteilung oder auch Nullhypothese machen durchaus Sinn, da sie schließlich unter der Voraussetzung eines Zufallsprozesses (= kein Einfluss) erstellt wurden.

4.3.6. Zusammenfassung



Das inferenzstatistische Vorgehen lässt sich also in folgende Punkte unterteilen:

- Formulierung der Nullhypothese
- Basismodell: Definition der Menge der Elementarereignisse
- Definition einer Zufallsvariablen
- Bestimmung der Dichteverteilung der Zufallsvariablen (spezifiziertes Modell)
- Entscheidungskriterium: Annahme- und Ablehnungsbereich
- Durchführung der Empirie
- Prüfung, in welchem Bereich das empirische Ergebnis liegt

Eigentlich müsste das statistische Modell schon vor jeder Empirie spezifiziert werden. Um dies zu ermöglichen muss jedoch der Stichprobenumfang festgelegt sein. Hier stellt sich das große Problem, dass die optimale Stichprobengröße von der Effektstärke des interessierenden Effekts abhängt, die normalerweise am Anfang der Untersuchung nicht bekannt ist.

Hierzu noch einige Anmerkungen (Problembereiche).

4.3.6.1. Enumeration: Festlegen der Ereignisse

Ereignisse sind idealisierte Ausgänge eines Zufallsexperiments [es wird also z.B. nicht berücksichtigt, dass sich bei einem Münzwurf theoretisch auch die die Erde auftun und die Münze verschlucken könnte].

4.3.6.2. Zuordnung von Wahrscheinlichkeiten

Die Zuordnung von Wahrscheinlichkeiten kann auf zwei Arten erfolgen: theoriebezogen (a priori) und empirie-basiert.

Die theoriebezogene Zuordnung kann etwa nach dem Zufalls- oder Rate-modell erfolgen oder auch aufgrund von theoretischen Modellannahmen hergeleitet werden.

Die empirische Zuordnung kann entweder aus zuvor gewonnenem Wissen geschehen (z.B. in der Epidemiologie) oder auch aus der vorliegenden Untersuchung abgeleitet werden.

Hier ist jedoch immer zu beachten, dass die Empirie keine Wahrscheinlichkeiten kennt – hier gibt es nur Häufigkeiten. Der Grenzwert dieser Häufigkeiten ist also nur ein Schätzer für tatsächliche Wahrscheinlichkeiten.

Allerdings ist dabei zu beachten, dass durch die empiriebezogene Zuordnung von Wahrscheinlichkeiten die Anzahl der Freiheitsgrade des statistischen Modells verringert wird: Jede Information, die sich das Modell aus der Empirie holt kostet es einen Freiheitsgrad.

Am Beispiel der Vierfeldertafel: Ohne empirische Informationen besitzt die 4-Felder-Tafel logischerweise 4 df. Wird jedoch die Anzahl der Pbn spezifiziert ($n = a+b+c+d$), so kostet dies einen df. Wird jetzt eine Randsumme festgelegt (z.B. $n_{1.} = a + b$, wobei $c+d$ ebenfalls festgelegt wird, da $c+d = n - (a + b)$) kostet dies wieder einen df. Wird nun die letzte mögliche Randsumme spezifiziert, so besitzt das Modell nur noch einen df.

		nachher		
		+	-	
vorher	+	a	b	
	-	c	d	
				n

↑
3. wie 2.

← 2. Wieder ein df weniger, wobei die zweite Randsumme ebenfalls festgelegt ist
 ← 1. Das kostet 1 df.

Genauso: Der Varianzschätzer hat $n-1$ df, weil \bar{X} aus der Empirie genommen wird. Eine Korrelation hat $n - 2$ df, weil für ihre Berechnung zwei Mittelwerte (je einer pro Variablen) aus der Empirie genommen werden.

Eine letzte Möglichkeit der empirischen Definition von Wahrscheinlichkeiten ist die Mediandichotomisierung, die selbstverständlich zuverlässig zu $p = 0,5$ führt.

4.3.6.3. Definition der Zufallsvariablen

Die Definition der Zufallsvariablen hängt nur von der inhaltlichen Fragestellung ab (s.o.). So können Summen und Mittelwerte (Lokation), Quadratsummen und Varianzen (Dispersion) oder andere deskriptive Größen herangezogen werden.

Die Definition der Zufallsvariablen ist der Schritt in der Modellentwicklung, der unterschiedliche Tests produziert.

Vergleicht man beispielsweise Zwillinge hinsichtlich Geburtsgewicht und IQ (Differenzbildung $IQ_{\text{dick}} - IQ_{\text{dünn}}$), so könnte als Zufallsvariable die Summe der positiven Differenzen interessieren (\rightarrow Nominaldaten: Anzahl „+“). Hieraus resultiert der Vorzeichentest.

Allerdings lassen sich die Differenzen auch nach ihrem Absolutbetrag rangordnen, sodass große Differenzen stärker gewichtet werden. Nun sprechen hohe Rangsummen der positiven Differenzen für die Alternativhypothese: Ist die Rangsumme der positiven Differenzen hoch, so sind die

dickeren Zwillinge auch die intelligenteren (Wilcoxon-Test für abhängige Stichproben).

Verschiedene Tests resultieren also nicht aus einer Veränderung der Elementarmenge, sondern aus einer veränderten Definition der ZV und beruhen damit auf rein inhaltlichen Fragestellungen.

Durch unterschiedliche ZVn und unterschiedliche Tests ändern sich auch Übertretungswahrscheinlichkeiten, sodass Ablehnungs- und Annahmewahrscheinlichkeiten ebenfalls abgepasst werden.

4.3.6.4. Nullverteilung: Zuordnung von W'keiten zur ZV

Bei abhängigen Mengen funktioniert die Gleichung g/m nicht mehr. Zudem interessieren meist nicht die Wahrscheinlichkeiten für ein bestimmtes Ergebnis, sondern die Wahrscheinlichkeit, dass durch Zufall das empirische oder ein extremeres Ereignis eintritt.

Wichtige Begriffe sind in diesem Zusammenhang die Punktwahrscheinlichkeit ($P(X=x_i)$) bzw. analog dazu die Wahrscheinlichkeitsdichtefunktion bei stetigen Zufallsvariablen auf der einen Seite sowie Überschreitungswahrscheinlichkeiten bzw. Wahrscheinlichkeitsverteilungen auf der Seite der modellbezogenen Entscheidung.

4.4. Exakte Tests

In einem Experiment soll geklärt werden, ob ein neues Medikament Ratten den Weg durch ein Labyrinth schneller Lernen lässt. $N = 15$ Ratten werden in Experimentalgruppe ($X+$; $N_1 = 7$) und Kontrollgruppe ($X-$; $N_2 = 8$) aufgeteilt:

	Gelernt Y+	Nicht gelernt Y-	Σ
Experimentalgruppe X+	a = 5	b = 2	$N_1 = 7$
Kontrollgruppe X-	c = 1	d = 7	$N_2 = 8$
Σ	$M_1 = 6$	$M_2 = 9$	$N = 15$

H_0 : Das Medikament erbringt keine Veränderung der Lernleistung:

$$P(Y+|X+) = P(Y+|X-) = P(Y+)$$

H_1 : Veränderung der Lernleistung durch Verabreichung des Medikaments

Das Prinzip des exakten Testens besteht darin, alle möglichen Ausgänge eines Experiments in die Entscheidung für oder gegen die Nullhypothese einzubeziehen.

In diesem Beispiel bezeichne p_{X+} die Wahrscheinlichkeit, dass in der Experimentalgruppe $X+$ mit $N_1 = 7$ genau $a = 5$ Ratten das Labyrinth lernen. Nach der Binomialverteilung ergibt sich:

$$p_{X+} = \binom{N_1}{a} \cdot \pi_{X+}^a \cdot (1 - \pi_{X+})^{N_1-a} \quad (1)$$

wobei π_{X+} die im Allgemeinen unbekannte Wahrscheinlichkeit bezeichnet, dass eine Ratte der Experimentgruppe $X+$ das Labyrinth lernt $P(Y+|X+)$. Genauso berechnet sich die Wahrscheinlichkeit p_{X-} , dass in der Kontrollgruppe $X-$ mit $N_2 = 8$ genau $c = 1$ Ratte das Labyrinth lernt nach:

$$p_{X_-} = \binom{N_2}{c} \cdot \pi_{X_-}^c \cdot (1 - \pi_{X_-})^{N_2 - c} \quad (2)$$

Die Wahrscheinlichkeit, dass beide Ereignisse ($a = 5$ und $c = 1$) auftreten lässt sich ebenfalls angeben, da die beiden Gruppen unabhängig voneinander sind:

$$p_{X_+ \cap X_-} = \binom{N_1}{a} \cdot \pi_{X_+}^a \cdot (1 - \pi_{X_+})^{N_1 - a} \cdot \binom{N_2}{c} \cdot \pi_{X_-}^c \cdot (1 - \pi_{X_-})^{N_2 - c} \quad (3)$$

H_0 postuliert, dass sich die beiden Wahrscheinlichkeiten π_{X_+} und π_{X_-} nicht unterscheiden (das Medikament hat keinen Effekt). Berücksichtigt man diese Annahme und die Beziehungen $N_1 - a = b$, $N_2 - c = d$, $a + c = M_1$ und $b + d = M_2$ lässt sich Formel (3) vereinfachen:

$$p_{X_+ \cap X_-} = \binom{N_1}{a} \cdot \binom{N_2}{c} \cdot \pi^{M_1} \cdot (1 - \pi)^{M_2} \quad (4).$$

Vereinigt man die beiden Stichproben, haben $M_1 = 6$ Ratten das Labyrinth gelernt, $M_2 = 9$ scheiterten. Die Wahrscheinlichkeit p_T ($T = \text{Total}$), genau dieses Ergebnis zu erhalten, ist mit Formel (5) wieder durch die Binomialverteilung gegeben:

$$p_T = \binom{N}{M_1} \cdot \pi_T^{M_1} \cdot (1 - \pi_T)^{M_2} \quad (5)$$

π_T bezeichnet die unbekannte Wahrscheinlichkeit, dass eine Ratte (unabhängig davon, ob sie aus Kontroll- oder Experimentalgruppe kommt), das Labyrinth lernt. Diese Wahrscheinlichkeit ist bei Gültigkeit von H_0 genau dieselbe wie in den beiden Stichproben (s. o.):

$$\pi_{X_+} = \pi_{X_-} = \pi_T$$

Mit p° bezeichnen wir jetzt die Wahrscheinlichkeit, dass M_1 erfolgreiche und M_2 erfolglose Ratten sich *zufällig* so in zwei Stichproben vom Umfang N_1 und N_2 aufteilen, dass sich in der einen Stichprobe X_+ insgesamt a erfolgreiche, in der zweiten Stichprobe $M_1 - a = c$ erfolgreiche Ratten befinden.

Die a bzw. c erfolgreichen Fälle werden dabei unter der Bedingung betrachtet, dass es insgesamt M_1 erfolgreiche Fälle gibt und die Stichproben gerade die Größen N_1 und N_2 haben. Dieses p° berechnet sich als bedingte Wahrscheinlichkeit aus dem Quotienten

$$p^\circ = \frac{p_{X_+ \cap X_-}}{p_T} = \frac{\binom{N_1}{a} \cdot \binom{N_2}{c} \cdot \pi^{M_1} \cdot (1 - \pi)^{M_2}}{\binom{N}{M_1} \cdot \pi_T^{M_1} \cdot (1 - \pi_T)^{M_2}} = \frac{\binom{N_1}{a} \cdot \binom{N_2}{c}}{\binom{N}{M_1}} \quad (6).$$

Wie Formel (6) zeigt, kürzt sich die unbekannte Wahrscheinlichkeit π heraus: die Wahrscheinlichkeit p° (also genau das beobachtete Ergebnis zu erhalten), ist keine Funktion der Wahrscheinlichkeit in der Population mehr, sondern kann unter der Gültigkeit von H_0 aus den Beobachtungswerten berechnet werden.

Merke: Bei Unterschiedsfragestellungen lassen sich unbekannte Wahrscheinlichkeiten π unter der Nullhypothese kürzen.

Berücksichtigt man die Definition

$$\binom{N}{x} = \frac{N!}{x!(N-x)!}$$

lässt sich für p° folgende einfache Formel angeben:

$$p^\circ = \frac{N_1!N_2!M_1!M_2!}{N!a!b!c!d!} \quad (7).$$

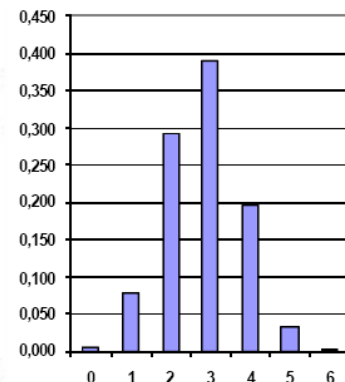
p° wird als Punktwahrscheinlichkeit bezeichnet ($P(X=x_i)$ bzw. „Die Wahrscheinlichkeit, dass genau diese Tafel bei gegebenen Randsummen auftritt“) und berechnet sich für das obige Beispiel zu

$$p^\circ = \frac{7!8!6!9!}{15!5!2!!17!} = 0.03357$$

Jeder anderen Kombination von Zellenbesetzungen, die den Bedingungen der Randsummen gehorcht, lässt sich ebenfalls eine Punktwahrscheinlichkeit p° zuschreiben.

Tabelle 2

a	b	c	d	p°	P_1	P_2
0	7	6	2	.00559	.00559	.00699
1	6	5	3	.07832	.08391	.11888
2	5	4	4	.29371	.37762	.60839
3	4	3	5	.39161	.62238	1.00000
4	3	2	6	.19580	.23077	.31468
5	2	1	7	.03357	.03497	.04056
6	1	0	8	.00140	.00140	.00140



Für einen statistischen Test werden jedoch nicht die Punktwahrscheinlichkeiten benötigt, sondern die Überschreitungswahrscheinlichkeiten („Wahrscheinlichkeit dafür, dass das empirische oder ein extremeres Ereignis auftritt“). Daher werden die entsprechenden Punktwahrscheinlichkeiten bei exakten Tests einfach aufsummiert.

Die Nullhypothese bestimmt dabei die Summationsrichtung (vgl. 4.3.5), wobei in der Tabelle Überschreitungswahrscheinlichkeiten für einseitige Tests in beide Richtungen (P_1) und zweiseitige Tests (P_2) aufgeführt sind.

Für das gegebene Beispiel beträgt die Übertretungswahrscheinlichkeit $P_1 = 0.035$ – es kann also eine Wirkung des Medikaments nachgewiesen werden.

Exkurs: Für zweiseitige Tests werden die Punktwahrscheinlichkeiten von beiden Seiten aufsummiert, bis die Irrtumswahrscheinlichkeit das Signifikanzniveau überschreitet [(0+6)+(1+5)+(2+4)].

Bei exakten Tafeln wird jedoch eine andere Vorgehensweise empfohlen, da hier keine symmetrischen Verteilungen zugrunde liegen müssen. Von daher sollten die Punktwahrscheinlichkeiten der Größe nach sortiert und aufaddiert werden, bis das Signifikanzniveau überschritten wird. Nach dieser Regel gehört $a = 5$ noch zum Ablehnungsbereich der Nullhypothese, nach der symmetrischen Entscheidungsregel hingegen nicht.

4.5. Asymptotische Tests

Exakte Tests sind nur bei diskreten Verteilungen möglich und werden schnell lästig und kompliziert. Beispielsweise müssen bei 100 Pbn 2^{100} verschiedene Fälle (Vierfeldertafeln) unterschieden und berechnet werden.

„Wenn Sie einen solchen Test per Hand rechnen werden sie verrückt oder alt.“

Von daher können exakte Tests über eine Normalverteilungsapproximation vereinfacht werden, wenn genügend Probanden untersucht werden ($N > 20, 25, 30$).

Für die Vierfeldertafel ist der Vierfelder- χ^2 -Test einschlägig. Die Annahme lautet wie bei den exakten Tests (Unabhängigkeit der Versuchsgruppen), wobei als Prüfgröße ein Varianzmaß berechnet wird:

$$\chi^2 = \frac{(b-e)^2}{e}$$

Als Varianzmaß folgt die asymptotische Prüfgröße einer χ^2 -Verteilung.

Exkurs: Verteilungsformen von Prüfgrößen bei mehrmaliger Ziehung aus einer normalverteilten Grundgesamtheit.

$$x_1 + x_2 + x_3 + \dots x_n \rightarrow \Sigma \rightarrow NV \quad (\text{Prüfung von Mittelwerten})$$

$$x_1^2 + x_2^2 + x_3^2 + \dots x_n^2 \rightarrow \Sigma \rightarrow \chi^2 \quad (\text{Prüfung von Varianzen})$$

$$x_1^2 \text{ (also } \chi^2 \text{ mit 1 df): Halbe, breite Normalverteilung}$$

In der Vierfeldertafel gilt unter H_0 , dass die in die Prüfgröße eingehenden Differenzen (b-e) nicht anderes sind als zufällige Streuungen – also die ganz normale Varianz [unter H_0 werden einfach nur quadrierte ZVn aus einer normalverteilten Grundgesamtheit gezogen].

Je größer die Differenzen, desto unwahrscheinlicher ist es, dass es sich wirklich nur um zufällige Schwankungen handelt. Ein großer χ^2 -Wert spricht also gegen die Nullhypothese.

Wichtig ist bei asymptotischen Tests, dass bestimmte Randvoraussetzungen gegeben sind (große Stichproben). Für χ^2 -Tests im speziellen müssen alle erwarteten Häufigkeiten über 5 liegen. Ist diese Voraussetzung nicht erfüllt resultiert ein großer Fehler.

4.6. Die Ausgangslage

Je nach Fragestellung kann man sich Urnen zusammenstellen, um ein Modell zu veranschaulichen. Die Urnen stellen nichts anderes dar als die möglichen Ausprägungen der abhängigen Variablen im Versuch. Anders ausgedrückt: Die AV bestimmt Art und Zusammensetzung der Urne.

Das Ziehen aus diesen Urnen geschieht analog zu den Manipulationen im Experiment und bildet die unabhängigen Variablen nach. Die UV bestimmt also die Art des Ziehens aus der Urne.

Die Urnen unterscheiden sich darin, dass

- endlich viele oder
- unendlich viele

Ereignisse bestimmter Wahrscheinlichkeiten enthalten, die

- qualitativ (nominal)
- oder quantitativ

gestuft sind, wobei sich innerhalb der quantitativen

- diskrete (ordinale) und
- stetige

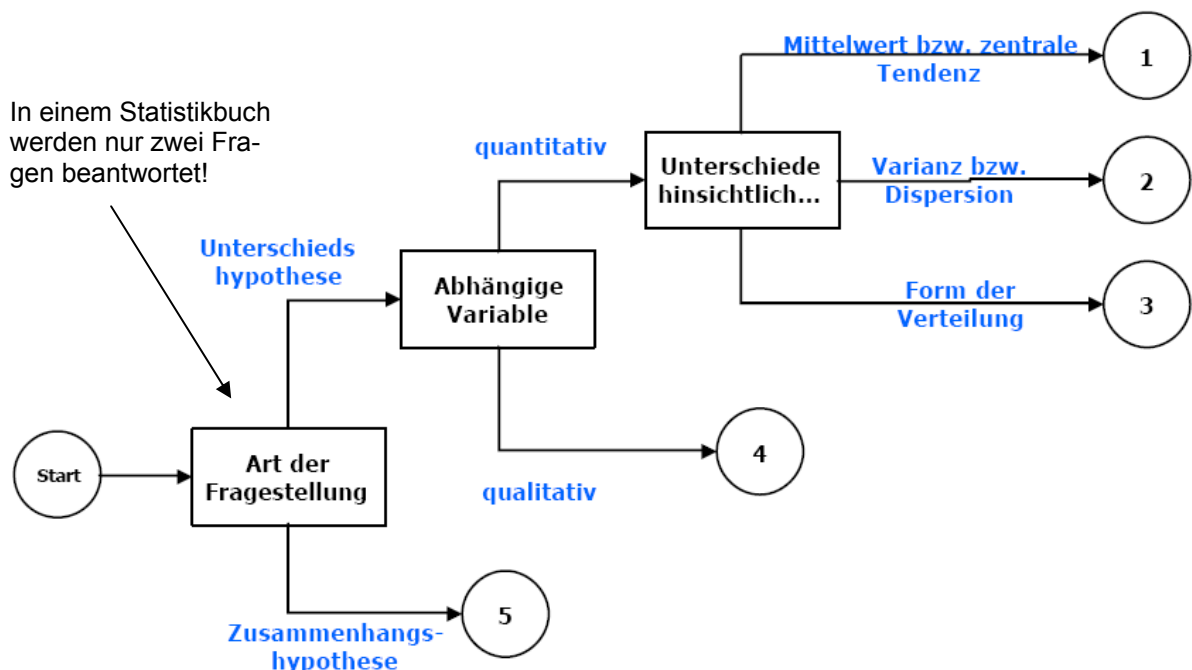
unterscheiden lassen.

Da man diese Zufallsmengen je nach Fragestellung zusammenstellen muss, ergibt sich die Frage: „Welche Fragestellungen gibt es eigentlich?“.

Jede Fragestellung muss nach den folgenden Merkmalen klassifiziert werden:

- Unterschieds- oder Zusammenhangs-Fragestellung?
- Ist die AV eine qualitative oder eine quantitative Größe?
- Wie viele Stichproben liegen vor?
 - o Eine
 - o Zwei
 - o Mehrere
- Welche beschreibende Größe soll untersucht werden?
 - o Lokation
 - o Dispersion
 - o Ganze Verteilung
- Welche Verteilungsform haben die Messwerte?
 - o Normalverteilt
 - o Gleichverteilt
 - o Unbekannte Verteilung

Diese Unterscheidungen ergeben nur 5 unterschiedliche Fragestellungen:



5. Von der Stichprobe zur Population

Im Forschungsprozess wurde zunächst eine allgemeine Frage in untersuchbare Größen umgewandelt und diese wurden anschließend in einer Untersuchung an einer Stichprobe erhoben. Im letzten Schritt müssen die erhaltenen Ergebnisse interpretiert werden, also deren Bedeutung über die Untersuchung hinaus erläutert werden.

Dieser letzte Schritt hängt vor allem von den Eigenschaften und der Größe der Stichprobe ab, die repräsentativ für die Grundgesamtheit sein soll.

5.1. Grundgesamtheiten

Für die Definition einer Grundgesamtheit gibt es mehrere wichtige Faktoren:

- Eine Grundgesamtheit muss abgrenzbar von anderen Grundgesamtheiten sein. Grundgesamtheiten werden daher i.d.R. nach einem ganz bestimmten Muster definiert: „**definitio fit per genus proximum et differentiam specificam**“.

Auf Deutsch: Zu welcher Klasse von Begriffen (Oberklasse) gehört die Grundgesamtheit (genus proximum) und was zeichnet ihn aus (differentia specifica). [„Aus allen lebenden Menschen (genus proximum) nehme nicht nur 5-10-jährige Jungen (differentia specifica)“].

Psychologen machen sich dabei kaum Gedanken, für wen ihre Aussagen gelten sollen, im Gegensatz beispielsweise zu medizinischen Studien, wo meist klar definierte Ein- und Ausschlusskriterien vorliegen.

Ansätze finden sich höchstens im Bereich der Psychotherapieforschung (z.B. werden nur Patienten mit bestimmten Störungen untersucht).

- Einfache vs. unendliche Grundgesamtheiten (Bewohner eines Wohnheims zu einem bestimmten Zeitpunkt vs. alle Menschen)
- Konkrete vs. hypothetische Grundgesamtheiten
- Ein- vs. Mehrdimensionale Grundgesamtheiten: Nicht ein, sondern mehrere Merkmale als differentia specifica (Frauen mit Hochschulabschluss ohne Kinder).

5.2. Stichproben und Repräsentativität

In der Stichprobentheorie werden zwei Arten von Stichproben unterschieden: Zufalls- und Beurteilungsstichproben (random vs. systematic sampling):

- Zufällige Ziehungen gewähren eine Erwartungsgleichheit von Stichprobe und Grundgesamtheit (Repräsentativität). Demnach sind nur echte Zufallsstichproben (random sampling) repräsentativ.
Es existieren verschiedene Arten von Zufallsstichproben: Einfache Zufallsstichproben, Geschichtete Zufallsstichproben (proportionale und optimale Schichtung), Klumpenstichproben, Mehrphasenstichproben (zu mehreren Zeitpunkten gemessen).
- Beurteilungsstichproben (systematic sampling) sind niemals wirklich repräsentativ.

Die Freitagsfrage bei Bundestagswahlen ist beispielsweise keine Zufallsstichprobe (zu teuer), sondern ein konstantes Panel, welches nach den Zahlen des statistischen Jahrbuchs gebildet wird. Die Struktur der Stichprobe entspricht also der Struktur der Grundgesamtheit, es liegt aber keine wirklich repräsentative Stp vor.

Gültigkeitstests ergeben trotzdem Repräsentativität, weil diese nur prüfen, ob richtig viele Männer, 30-jährige etc. in der Stichprobe vorhanden sind, was hier trotz nicht-zufälliger Auswahl natürlich gegeben ist.

5.3. Statistiken und Parameter

Nun stellt sich die Frage, wie das gefundene Ergebnis verallgemeinert werden kann. Die Repräsentativität (= Erwartungswertgleichheit) der Stichprobe wird durch die Stichprobentheorie gewährleistet („Wie wähle ich eine Stp aus?“). Neben der Stichprobentheorie muss jedoch auch ein Regelwerk vorliegen, das beschreibt, was die gefundenen Werte bedeuten: die Parameterschätzung.

Die Parameterschätzung behandelt also, wie die Stichprobenparameter mit den entsprechenden Parametern der Grundgesamtheit zusammenhängen. Zur Bestimmung eines brauchbaren Schätzers für Grundgesamtheitsparameter werden vier Kriterien angeführt:

- **Erwartungstreue** (kein Bias bzw. Schätzer nicht verzerrt)
- **Konsistenz** (mit zunehmendem n wird der Schätzfehler verringert)
- **Effizienz** (relatives Maß im Vergleich zu anderen Schätzern)
- **Suffizienz** (Verwendung aller Informationen aus den Daten)

Man unterscheidet dabei drei Methoden der Parameterschätzung: Momenten-Methode, Prinzip der kleinsten Quadrate und Maximum Likelihood. Häufig kommen alle drei Methoden auf die gleichen Schätzer, aber nicht immer.

5.3.1. Momenten-Methode

Die Momenten-Methode geht davon aus, dass der Stichprobenparameter ein angemessener Schätzer für den entsprechenden Grundgesamtheitsparameter darstellt.

[„Du hast dir so viel Mühe gegeben, eine erwartungstreue Zufallsstichprobe zu ziehen, also wird das schon irgendwie passen.“]

Jedes Moment⁵ einer Stichprobenverteilung wird also als direkt übertragbar auf die Population angenommen. Für die Prüfung der Gültigkeit dieser Annahme existieren Prüfverfahren.

5.3.2. Prinzip der kleinsten Quadrate

Die Annahme dieser Methode ist, dass die Populationsparameter in der Nähe der empirischen Werte liegen sollten (Wo auch sonst?). Also wird der Wert als Schätzer verwendet, der von den empirischen Werten den kleinsten Abstand hat (Minimierungsgedanke).

Ein Beispiel für den Mittelwert:

$$\sum (X_i - \mu)^2 \rightarrow \text{Formel für den Abstand der Messwerte vom Mittelwert.}$$

Für Minimum: Ableiten und Nullsetzen

$$-2 \sum (X_i - \mu) = -2 (\sum X_i - \sum \mu) \stackrel{!}{=} 0$$

⁵ 1. Moment einer Verteilung: \bar{X} , 2. Moment: SD, 3. Moment: Schiefe, 4. Moment: Exzess

$$\mu = \frac{\sum X_i}{n}$$
$$\rightarrow \frac{\sum X_i}{n} = \bar{X} \text{ (Stichprobenmittelwert entspricht Populationsmittelwert)}$$

5.3.3. Maximum Likelihood Methode

Die Maximum Likelihood Methode hat einen ähnlichen Grundgedanken wie das Prinzip der kleinsten Quadrate: Als Schätzer für den Grundgesamtheitsparameter wird der Wert genommen, der die Stichprobenverteilung mit der höchsten Wahrscheinlichkeit erzeugt hat (Maximierungsgedanke).

Hier wird die Wahrscheinlichkeit der Stichprobe und Annahme eines Parameters abgeleitet und nullgesetzt.

6. Modelle der Spieltheorie

Modelle der Spieltheorie⁶ können als Grundlage zur Erklärung komplexer Vorgänge wie etwa auch dem menschlichen Verhalten verwendet werden.

John von Neumann (vgl. 4.1.1: Wie kann ein deterministischer Apparat Zufall erzeugen), definierte ein „Spiel“ im mathematischen Sinne als die „Gesamtheit aller Regeln, die es beschreiben“. Eine wiederholte Anwendung auch von einfachsten Regeln kann dabei zu komplexen Systemen führen.

Eine Regel enthält Angaben über:

- Anzahl der Spieler
- Wer ist an der Reihe?
- Möglichkeiten des Spielers
- Was wissen die anderen Spieler?

Die wichtigste Art von Spielen sind Nullsummenspiele, wobei hier entscheidend ist, ob das Spiel nur einen Durchgang hat oder mehrere (bis unendliche). Darstellen lassen sich solche Spiele in einer Auszahlungsmatrix.

Eine populäre Veranschaulichung der Spieltheorie ist John Conways „Game of Life“, welches komplexe Muster aus Pixeln am Computerbildschirm durch drei einfache Regeln entstehen lässt:

- Jeder Pixel, der genau 3 lebende Nachbarn hat wird in der nächsten Runde neu geboren.
- Jeder Pixel der weniger als 3 lebende Nachbarn hat stirbt an Einsamkeit
- Jeder Pixel der mehr als 3 Nachbarn hat stirbt an Überbevölkerung.

Eine Demonstration findet sich unter <http://psych.hanover.edu/JavaText/Play/life.htm>

Die Spieltheorie kann also auf Basis von einfachen Regeln eine unendliche Vielzahl von empirischen Phänomenen erklären, beispielsweise „Warum gibt es genau diesen Anteil von Neurotikern in der Bevölkerung?“, und sind somit eine große Erweiterung des klassischen methodischen Repertoires.

⁶ Manfred Eigen (Nobelpreis) & Ruthild Winkler (1975). Das Spiel. Naturgesetze steuern den Zufall. München: Piper