Holger Wilker

Band 1

Statistische Hypothesentests in der Praxis

Leitfaden zur Anwendung von Hypothesentests für die Analyse von Unterschieden, Übereinstimmungen, Zugehörigkeiten, Zufälligkeiten und Zusammenhängen

Zweite, vollständig überarbeitete Ausgabe

Holger Wilker

Band 1

Statistische Hypothesentests in der Praxis

Leitfaden zur Anwendung von Hypothesentests für die Analyse von Unterschieden, Übereinstimmungen, Zugehörigkeiten, Zufälligkeiten und Zusammenhängen

Zweite, vollständig überarbeitete Ausgabe

Mit 74 ausgewählten Hypothesentests und ausführlichen Beispielen

Holger Wilker
Wildeshausen, April 2018

Bibliographische Information der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliographie. Detaillierte Daten sind im Internet über http://dnb.ddb.de abrufbar.

ISBN 9783752817706, Zweite Ausgabe · · ·

Herstellung und Verlag: BoD – Books on Demand GmbH, Norderstedt.
Satzsystem: $\LaTeX\ 2_\varepsilon$
Printed in Germany

meiner Frau Marion

meinem Sohn Albert

Vorwort zur zweiten, vollständig überarbeiteten Ausgabe

Der vorliegende Band wurde vollständig und umfassend überarbeitet. Zahlreiche Anregungungen, Vorschläge und Erkenntnisse haben insbesondere die Anwendung statistischer Hypothesentests in den Mittelpunkt dieses Buches rücken lassen. Motivierend hierfür war auch die Unzufriedenheit hinsichtlich eines Dickichts zahlreicher, oft nur Einzelaspekte berücksichtigender, nationaler sowie internationaler Literatur, oft auch von Mathematikern für Mathematikern publiziert, selten gepaart mit erschöpfend behandelten und adaptierfähigen Beispielen.
Im Fokus dieser Überarbeitung werden anhand von vorliegenden Daten hinsichtlich der zu untersuchenden Eigenschaften *Unterschiede*, *Übereinstimmungen*, *Zugehörigkeiten*, *Zufälligkeiten* und *Zusammenhänge*, mathematisch belastbare Aussagen vorab formulierter Hypothesen abgeleitet. Die Reduzierung des ursprünglichen Themenspektrums bei gleichzeitiger Vertiefung der praxisbezogenen Anwendung statistischer Testverfahren machten es erforderlich, den Titel des Buches diesen Inhalten anzupassen.
Dieses Buch befähigt den Anwender, ausgewählte, für eine Vielzahl von Studiengängen, insbesondere im Grundstudium, relevante Hypothesentests gezielt und systematisch anzuwenden. Übersichtlich werden sowohl die Benennungen, die Fragestellungen, die relevanten Voraussetzungen und die zu jedem einzelnen Hypothesentest vollständig berechneten Beispiele dargestellt. Im Anhang sind für jedes behandelte Testverfahren die notwendigen kritischen Werten aufgeführt. Zur Vertiefung der Testverfahren sind für jeden aufgelisteten Hypothesentest die korrespondierenden Literaturhinweise angegeben.
Der Inhalt dieses Buches begleitet sowohl Studenten, Fachhochschul- und Hochschulabsolventen, welche sich einen ergänzenden Überblick zu wesentlichen Testverfahren und deren praktischen Anwendungen verschaffen wollen, als auch die in der Praxis tätigen Berufsgruppen, welche mathematisch belastbare und standardisierte Testentscheidungen zu vorhandenem Datenmaterial treffen müssen.
Vorhandene Redundanzen sind im Sinne einer darstellenden Didaktik ausdrücklich beabsichtigt. In bewährter Weise sind zu jedem behandelten Testverfahren vollständig durchgerechnete Beispiele angegeben, welche die beschriebenen Vorgehensweisen vertiefen und die Grundlage für die Lösung eigener Problemstellungen bilden.
Mein Dank geht insbesondere an meine Frau und meinen Sohn, welche mit erschöpfender Geduld und Rücksicht auch dieses Projekt wieder ermöglichten.

Wildeshausen, April 2018 Holger Wilker

Vorwort

Der vorliegende eigenständige erste Teil zur Optimierung von Produkten und Prozessen soll den Anwender in die Lage versetzen, aus vorliegenden Daten, Meß- und Versuchsergebnissen oder sonstigen Informationen, welche aus ungeplanten Untersuchungen oder ungeplanten Versuchen resultieren, entsprechende statistisch belastbare Aussagen als Grundlage für Erkenntnisse und weitergehende Entscheidungen abzuleiten. Dabei sollen sowohl die in der Praxis so wichtigen mindesterforderlichen Stichprobenumfänge für einzuhaltende Aussagesicherheiten als auch die Konstruktion von Vertrauensintervallen für kleine Stichprobenumfänge behandelt werden.
Dieses Buch soll gerade wegen der Vielzahl von Nichtstatistikern und zur Verfügung stehender Software nicht handlungsanweisenden Fakten- und Verfügungswissen vermitteln, sondern mit dazu beitragen, die Voraussetzungen für ein kompetenzvermittelndes und praxisrelevantes Orientierungswissen auf dem Gebiet der angewandten Statistik zu schaffen.
Es wendet sich gleichfalls an Studenten, Fachhochschul- und Hochschulabsolventen, welche sich einen ergänzenden Überblick zu wesentlichen Verfahren und Anwendungen der angewandten Statistik verschaffen wollen sowie an die in der Praxis tätigen Berufsgruppen, welche vorhandenes Datenmaterial auch geringen Umfangs mit Hilfe statistischer Werkzeuge zu einer gegebenen statistischen Aussagesicherheit analysieren und interpretieren müssen.
Vorhandene Redundanzen sind im Sinne einer zu zeigenden Systematik ausdrücklich beabsichtigt. In bewährter Weise sind zu jedem behandelten Abschnitt vollständig durchgerechnete Beispiele angegeben, welche die beschriebenen Vorgehensweisen verdeutlichen und die Grundlage für die Lösung eigener Problemstellungen bilden sollen.
Mein Dank geht insbesondere an meine Frau und meinen Sohn, welche mit Geduld und Rücksicht tolerierten, daß ich mich dem Familienleben entzog, um an diesem Buch zu arbeiten.

Lauffen, Februar 2006 Holger Wilker

Inhaltsverzeichnis

1. Einführung

Die Beschreibung wesentlicher Begrifflichkeiten, die Skalierung vorhandener Daten (Nominal-, Ordinal-, Intervall- und Intervallskalierung), der Aufbau statistischer Hypothesentests (Formulierung der Hypothesen, Festlegung des verwendeten Stichprobenumfangs, der Irrtumswahrscheinlichkeiten, der Seitigkeit, der Teststatistik sowie die Problematik des multiplen Testens, werden im Abschnitt 2 beschrieben.

Zur Orientierung, welcher Hypothesentest in Abhängigkeit von seiner Ausrichtung (Unterschiede, Konkordanzen, Zugehörigkeiten, Zufälligkeiten, Zusammenhänge), den Fragestellungen, den Voraussetzungen, den Bennenungen und den jeweils behandelten Beispielen, zu verwenden ist, dienen mehrere Übersichten, welche den Anwendern zu einem geeigneten Testverfahren lotsen (Abschnitt 3).

In Abschnitt 4 werden die Hypothesentests, gegliedert nach ihrer Ausrichtung, beschrieben. Dabei wird ein für alle Hypothesentests streng formalisiertes Anwendungsverfahren angewendet: Beschreibung des Problems, Formulierung der Hypothesen, des Stichprobenumfangs und der Testvoraussetzungen, die Darstellung der verwendeten Teststatistik, die zu treffenden Entscheidungen sowie die resultierenden Interpretationen (Fazite). Nach diesem Schema ist jedem Testverfahren ein umfassend behandeltes Beispiel zugeordnet.

Für jedes in diesem Buch vorgestellte Testverfahren stehen im Anhang entsprechende kritische Werte zur Verfügung. Die jedem einzelnen Test zugeordneten Literaturhinweise bieten Möglichkeiten für die weitere Vertiefung des jeweiligen Hypothesentests.

2. Mathematische Grundlagen

2.1. Häufigkeiten

Ein Ereignis sei das Ergebnis eines mit bestimmten Parametern durchgeführten Versuches (Beobachtung, Experiment, etc.). Eine Häufigkeit sei die Anzahl, mit der gleiche Ereignisse, Merkmalsausprägungen, Kategorien auftreten. Das Auszählen von Häufigkeiten kann mit Daten auf allen Skalenniveaus (s. Abschnitt 1.4) vorgenommen werden.

2.1.1. Absolute Häufigkeiten

Die absolute Häufigkeit H eines Ereignisses E sei die Anzahl der Fälle, in denen das Ereignis E eintritt: H(E).

2.1.2. Relative Häufigkeiten

Die relative Häufigkeit h eines Ereignisses E sei die Anzahl der Fälle, in denen das Ereignis E auftritt, bezogen auf den Stichprobenumfang n: $h(E) = H(E)/n$.

2.1.3. Erwartete Häufigkeiten

Die erwartete Häufigkeit ist die in einer Kreuz- oder Kontingenztabelle bei Gültigkeit der Nullhypothese und Kenntnis der Randsummen zu erwartende Häufigkeit. Je größer die Diskrepanz zwischen beobachteten und erwarteten Häufigkeiten, um so eher wird die Nullhypothese abgelehnt.

2.2. Wahrscheinlichkeiten

Relative Häufigkeiten h beschreiben einen vorhandenen Stichprobenzustand. Wahrscheinlichkeiten p hingegen sind, basierend auf den relativen Häufigkeiten h, die zu erwartenden Zustände. Jedem Ereignis E kann eine Zahl p(E) zugeordnet werden, welche als Wahrscheinlichkeit dieses Ereignisses bezeichnet wird.
Ein zufälliges Ereignis kann, muß aber nicht, bei einem mit bestimmten Parametern durchgeführten Versuch eintreten: $0 < p(E) < 1$.
Ein unmögliches Ereignis kann bei einem mit bestimmten Parametern durchgeführten Versuch nicht eintreten: $p(E) = 0$.
Ein sicheres Ereignis tritt bei jedem mit bestimmten Parametern durchgeführten Versuch ein: $p(E) = 1$.

Beispiel – Wahrscheinlichkeiten

Gegeben sei eine Urne mit 21 Kugeln, davon 10 schwarze (s) Kugeln und 11 rote (r) Kugeln. Nach jeder Ziehung einer Kugel wird diese wieder in die Urne gelegt (Ziehen mit Zurücklegen).
Das Ziehen einer roten Kugel ist ein zufälliges Ereignis. Das Ziehen einer schwarzen oder roten Kugel ist ein sicheres Ereignis. Das Ziehen einer grünen Kugel ist ein unmögliches Ereignis.

2.2.1. Beobachtete Wahrscheinlichkeiten

Eine Beobachtung oder ein Ereignis E sei das Resultat einer Stichprobenziehung aus einer Grundgesamtheit. Dann kann jedem Ereignis E eine Zahl $0 \leq p(E) \leq 1$ zugeordnet werden, welche als Wahrscheinlichkeit p(E) eines solchen Ereignisses bezeichnet wird:

$$p(E) = \frac{\text{Anzahl der für ein Ereignis günstigen Fälle}}{\text{Anzahl aller möglichen Fälle}} .$$

Die Wahrscheinlichkeit einer Beobachtung entspricht der Punktwahrscheinlichkeit.

2.2.2. Vermutete Wahrscheinlichkeiten

Die vermutete Wahrscheinlichkeit p sei im Rahmen eines Hypothesentests diejenige Wahrscheinlichkeit, welche, formuliert als Nullhypothese, beizubehalten oder abzulehnen ist.

2.2.3. Theoretische Wahrscheinlichkeiten

Die theoretische Wahrscheinlichkeit P ist die auf lange Sicht erwartete Wahrscheinlichkeit, daß ein bestimmtes Ereignis im Rahmen definierter Randbedingungen auftritt und gilt für voneinander unabhängige, wiederholbare und gleich wahrscheinliche Ereignisse.

2.2.4. Empirische Wahrscheinlichkeiten

Die empirische, aus den vorliegenden Daten gemessene Wahrscheinlichkeit, stellt einen Näherungswert für die tatsächlich gemessene Wahrscheinlichkeit dar. Mit zunehmender Zahl von Stichproben n konvergiert die relative Häufigkeit eines Ereignisses h(E) gegen die empirische Wahrscheinlichkeit:

$$p(E) = \lim_{n \to \infty} \frac{h(E)}{n} .$$

2.2.5. Additionssatz

Aus der Vereinigung von $k = 2$ Ereignissen $E_1 \cup E_2$ resultiert das Eintreten von E_1 oder E_2 bzw. E_1 und E_2. Die Wahrscheinlichkeit p von k Ereignissen, welche sich wechselseitig ausschließen (also nicht gemeinsam auftreten können), ist die Summe der Wahrscheinlichkeiten dieser Ereignisse. Somit folgt für den Additionssatz der Wahrscheinlichkeitsrechnung:

$$p(E_1 \cup E_2 \cup \ldots \cup E_k) = p(E_1) + p(E_2) + \ldots + p(E_k) = \sum_{i=1}^{k} p_i$$

Beispiel – Additionssatz

Die Wahrscheinlichkeit, eine schwarze Kugel zu ziehen, sei: $p(s) = \frac{10}{21}$.
Die Wahrscheinlichkeit, eine rote Kugel zu ziehen, sei: $p(r) = \frac{11}{21}$.
Die Wahrscheinlichkeit, eine schwarze <u>oder</u> rote Kugel zu ziehen, beträgt: $p(s \vee r) = \frac{10}{21} + \frac{11}{21}$.
Eine nachfolgende Ziehung ist von der vorausgegangenen Ziehung unabhängig.

2.2.6. Multiplikationssatz

Aus dem Durchschnitt von $k = 2$ Ereignissen E_1, E_2 resultiert das Eintreten von E_1 und E_2. Die Wahrscheinlichkeit p von k Ereignissen, welche wechselseitig von einander unabhängig sind, ist das Produkt der Wahrscheinlichkeiten dieser Ereignisse. Somit folgt für den Multiplikationssatz der Wahrscheinlichkeitsrechnung:

$$p(E_1 \cap E_2 \cap \ldots \cap E_k) = p(E_1) \cdot p(E_2) \cdot \ldots \cdot p(E_k) = \prod_{i=1}^{k} p_i$$

Beispiel – Multiplikationssatz

Die Wahrscheinlichkeit, eine schwarze Kugel zu ziehen, sei: $p(s) = \frac{10}{21}$.
Die Wahrscheinlichkeit, eine rote Kugel zu ziehen, sei: $p(r) = \frac{11}{21}$.
Die Wahrscheinlichkeit, eine schwarze <u>und</u> anschließend rote Kugel zu ziehen, beträgt: $p(s \wedge r) = \frac{10}{21} \cdot \frac{11}{21}$.
Eine nachfolgende Ziehung ist von der vorausgegangenen Ziehung unabhängig.

2.2.7. Komplementärwahrscheinlichkeiten

Das komplementäre Ereignis $\overline{E}$ ist das Ereignis, welches dann eintritt, wenn das Ereignis E nicht eintritt: $p = p(\overline{E}) = 1 - p(E)$.

Beispiel – Komplementärwahrscheinlichkeit

Die Wahrscheinlichkeit in einer ersten Ziehung keine schwarze Kugel zu erhalten, sei: $p(\overline{s}) = 1 - p(s) = p(r) = \frac{21}{21} - \frac{10}{21} = \frac{11}{21}$.
Die Wahrscheinlichkeit in einer zweiten Ziehung keine schwarze Kugel zu erhalten, sei: $p(\overline{s}) = 1 - p(s) = p(r) = \frac{21}{21} - \frac{10}{21} = \frac{11}{21}$.
Die Wahrscheinlichkeit in einer ersten <u>und</u> in einer zweiten Ziehung keine schwarze Kugel zu erhalten, beträgt:

$$p(\overline{s}) \cdot p(\overline{s}) = \big(1 - p(s)\big) \cdot \big(1 - p(s)\big) = p(r) \cdot p(r) = \frac{11}{21} \cdot \frac{11}{21}.$$

Die Komplementärwahrscheinlichkeit wiederum zu dem Ereignis, in zwei aufeinanderfolgenden Ziehungen keine schwarze Kugel zu erhalten, ist die Wahrscheinlichkeit, mindestens in einer Ziehung eine rote Kugel zu ziehen: $p = 1 - \frac{11}{21} \cdot \frac{11}{21}$.

2.2.8. Bedingte Wahrscheinlichkeiten

Die bedingte Wahrscheinlichkeit $P(E_2 \mid E_1)$ kennzeichnet die Wahrscheinlichkeit eines Ereignisses E_2 unter der Bedingung, daß ein Ereignis E_1 bereits eingetreten ist:

$$P(E_2 \mid E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)}$$

mit $P(E_1) \neq 0$ und entsprechend

$$P(E_1 \mid E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

mit $P(E_2) \neq 0$ sowie dem Multiplikationssatz für das gleichzeitige Eintreffen zweier abhängiger oder unabhängiger Ereignisse E_1, E_2:

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2 \mid E_1) = P(E_2) \cdot P(E_1 \mid E_2) = P(E_2 \cap E_1)$$

bzw. für n abhängige oder unabhängige Ereignisse:

$$P(E_1 \cap E_2 \cap \ldots \cap E_n) = P(E_1) \cdot P(E_2 \mid E_1) \cdot P(E_3 \mid E_1 \cap E_2) \cdot \ldots \cdot P(E_n \mid E_1 \cap E_2 \cap \ldots \cap E_{n-1}).$$

Beispiel – Bedingte Wahrscheinlichkeiten

Gegeben sei eine Urne mit 21 Kugeln, davon 10 schwarze (s) Kugeln und 11 rote (r) Kugeln. Nach Ziehung einer Kugel wird diese nicht in die Urne gelegt (Ziehen ohne Zurücklegen).
Die Wahrscheinlichkeit, eine schwarze (E_1), eine rote (E_2) und eine schwarze Kugel (E_3) zu ziehen, beträgt:

$$\begin{aligned} P(E_1 \cap E_2 \cap E_3) &= P(E_1) \cdot P(E_2 \mid E_1) \cdot P(E_3 \mid E_1 \cap E_2) \\ &= \frac{10}{21} \cdot \frac{11}{20} \cdot \frac{9}{19} \,. \end{aligned}$$

Eine nachfolgende Ziehung ist von der vorausgegangenen Ziehung abhängig.

2.3. Grundgesamtheiten und Stichproben

2.3.1. Grundgesamtheiten

Als begrenzte oder unbegrenzte Grundgesamtheiten (Population, Fertigungslos), welche aus finanziellen, zeitlichen oder sachlichen Gründen nicht vollständig untersuchen werden kann, wird die Menge aller für eine Untersuchung relevanten statistischen Einheiten bezeichnet.

2.3.2. Zufallsstichproben

Informationen über das Vorhandensein und die Verteilung der interessierenden Merkmale in einer Grundgesamtheit sind über eine Stichprobe erhältlich. Für einen Schluß von der Stichprobe auf die Grundgesamtheit sollte die Stichprobe ein möglichst verkleinertes, aber genaues Abbild der Grundgesamtheit sein. Bei einer Zufallsstichprobe erfolgt die Auswahl von n Elementen aus einer Grundgesamtheit N genau dann zufällig, wenn

- die Auswahl eines Elementes die gleiche Wahrscheinlichkeit besitzt,
- die Wahrscheinlichkeit der Auswahl eines Elementes unabhängig von den interessierenden Eigenschaften ist,
- die Wahrscheinlichkeit der Auswahl eines Elementes berechenbar ist.

Die gewonnenen Merkmalsinformationen aus solchen zufälligen Stichproben weisen gegenüber den in der Grundgesamtheit vorhandenen Merkmalen nur die unvermeidlichen zufälligen Abweichungen auf, welche sich bei vielen Wiederholungen ausgleichen.

2.3.3. Abhängige Stichproben

Abhängige Stichproben[1] liegen dann vor, wenn die Wahrscheinlichkeit eines Stichprobenelements mit der Eigenschaft A in einer ersten Stichprobe die Wahrscheinlichkeit eines Stichprobenelements mit der Eigenschaft A in einer zweiten Stichprobe maßgeblich beeinflußt. Dieses ist genau dann der Fall, wenn zwischen den Stichproben kausale Zusammenhänge bestehen. Abhängige Stichproben treten bei wiederholten Messungen am gleichen Untersuchungsobjekt auf. Wenn die Werte der einen Stichprobe die Werte in der anderen Stichprobe beeinflussen, sind die Stichproben voneinander abhängig.

2.3.4. Unabhängige Stichproben

Unabhängige Stichproben liegen dann vor, wenn die Wahrscheinlichkeit eines Stichprobenelements mit der Eigenschaft A in einer ersten Stichprobe die Wahrscheinlichkeit eines Stichprobenelements mit der Eigenschaft A in einer zweiten Stichprobe nicht beeinflußt und genauso groß ist. Dieses ist genau dann der Fall, wenn zwischen den Stichproben keine Zusammenhänge bestehen.
Unabhängige Stichproben treten bei einmaligen Messungen an verschiedenen Untersuchungsobjekten auf. Wenn die Werte der einen Stichprobe keine Informationen über die Werte der anderen Stichprobe enthalten, sind die Stichproben voneinander unabhängig.

2.3.5. Stichprobenfehler

Eine Stichprobe repräsentiert oft nicht genau die Grundgesamtheit, aus welcher diese entnommen wurde. Die Varianz eines Merkmals innerhalb der Stichprobe entspricht häufig nicht der Varianz dieses Merkmals innerhalb der Grundgesamtheit. Dieser Fehler ist bei der Verallgemeinerung der Untersuchungsergebnisse zu berücksichtigen.
Die Verteilung der Stichprobenfehler aller möglichen Stichproben gleicher Größe aus einer Grundgesamtheit bildet die Stichprobenverteilung, die zur Prognose von Stichprobenwerten und auch zur Schätzung der Parameter einer Grundgesamtheit herangezogen wird. Der Stichprobenfehler ist die Streuung der Stichprobenverteilung bzw. die Differenz zwischen den Werten einer Stichprobe und dem entsprechenden wahren Wert in der Grundgesamtheit.
Die aus den Stichproben gewonnen Aussagen können nur für die Stichprobe selbst Gültigkeit beanspruchen. Für die Grundgesamtheit, aus welcher die Stichprobe entnommen wurde, wird die Gültigkeit nur vermutet.
Die durch das Ziehen einer Zufallsstichprobe entstandene Abweichung zwischen dem wahren Wert einer Variablen in der Grundgesamtheit und dem Stichprobenfehler ist um so geringer, je geringer die Varianz einer Verteilung und je größer der Umfang der Stichprobe ist. Aufgrund des Gesetzes der großen Zahlen[2] wird der Stichprobenfehler

[1] In der Literatur auch als gepaarte oder verbundene Stichproben bezeichnet.
[2] Nach dem Gesetz der großen Zahlen nimmt die Differenz zwischen einer Grundgesamtheit und einer dieser entnommenen Stichprobe mit zunehmenden Stichprobenumfang ab: $\lim_{x \to \infty} \overline{X}_n = \mu$.

ab einer gewissen Stichprobengröße so klein, daß eine Vergrößerung des Stichprobenumfangs nicht mehr sinnvoll sein kann.

2.3.6. Zufälligkeiten

Zufällige Ereignisse sind stochastisch unabhängig und gleichverteilt. Liegen Daten in einer zeitlichen Abfolge vor, so deuten Teilfolgen gleicher Werte oder Eigenschaften („Phasen“ , „Runs“ oder „Iterationen“) mit zu geringer Anzahl (zu geringe Durchmischung: AAAABBBB) oder zu großer Anzahl (zu starke Durchmischung: ABABABAB) daraufhin, daß die Abfolge nicht zufällig entstanden ist, obwohl die Häufigkeit dieser Merkmalsalternativen gleich ist. Eine Iteration b ist gekennzeichnet durch die Wiederholung eines bestimmten Zeichens, welchem ein anderes oder kein Zeichen vorangeht oder folgt.
Mit den Hypothesentests (**R01**) und (**R02**) können sowohl dichotome, künstlich dichotome, ordinale als auch kardinale Merkmale in einer Abfolge von Daten auf eine vorhandene Zufälligkeit getestet werden.

Iterationen identischer Merkmale unterschiedlicher Skalen

Die dichotome Datenreihe mit n = 11 Alternativmerkmalen A, B

$$\underbrace{\mathrm{AAA}}_{1}\ \underbrace{\mathrm{BB}}_{2}\ \underbrace{\mathrm{A}}_{3}\ \underbrace{\mathrm{BBB}}_{4}\ \underbrace{\mathrm{AA}}_{5}$$

ist durch b = 4 Iterationen (Phasen) gekennzeichnet.
Die ordinale Datenreihe mit n = 11 Alternativmerkmalen u, v, w, x

$$\underbrace{\mathrm{u}}_{1}\ \underbrace{\mathrm{vvv}}_{2}\ \underbrace{\mathrm{xx}}_{3}\ \underbrace{\mathrm{wwww}}_{4}\ \underbrace{\mathrm{uu}}_{5}$$

ist durch b = 5 Iterationen (Phasen) gekennzeichnet.
Die stetige Datenreihe mit n = 7 kategorialen Merkmalen $\{5\,9\,1\,4\,6\,7\,3\}$ besitzt den Median $\widetilde{x} = 4$. Abweichungen hiervon werden mit $+$ (größer oder gleich dem Median) bzw. – (kleiner als der Median) gekennzeichnet:

$$\underbrace{++}_{1}\ \underbrace{-}_{2}\ \underbrace{+++}_{3}\ \underbrace{-}_{4}$$

Diese künstlich dichotomisierte Datenreihe ist durch b = 4 Iterationen (Phasen) gekennzeichnet.
Die stetige Datenreihe mit n = 13 intervallskalierten Merkmalen

$$2,3 \quad 6,1 \quad 3,3 \quad 5,2 \quad 8,4 \quad 6,2 \quad 2,8 \quad 4,7 \quad 9,2 \quad 8,8 \quad 4,9 \quad 3,3 \quad 1,4$$

besitzt die Vorzeichendifferenzen $(x_{i+1} - x_i)$:

2,3 6,1 3,3 5,2 8,4 6,2 2,8 4,7 9,2 8,8 4,9 3,3 1,4
(–) (+) (–) (–) (+) (+) (–) (–) (+) (+) (+) (+)

2.3.7. Ausreisser

Ein Stichprobenelement (Wert in einer Datenreihe), welches nicht den Erwartungen entspricht bzw. nicht dem Verlauf anderer Werte folgt, wird als Ausreisser bezeichnet. Mögliche Ursachen für Ausreisser können sporadische oder systematische Fehler sein. Ein solcher Meßwert kann statistische Ursachen aufweisen (hohe, jedoch seltene Abweichungen vom wahren Wert). Ausreisser können aber auch auf unbekannte Effekte hindeuten (Erwartungen wären dann zu hinterfragen).
In jedem Fall ist zu entscheiden, wie mit vorhandenen Ausreissern in einem Datensatz zu verfahren ist. Werden Ausreisser von einer weiteren statistischen Analyse ausgeschlossen, so ist dieses zu notieren. Sind auf Basis eines solchen Hypothesentests Entscheidungen von großer Tragweite treffen, so sollte dieser Test einmal mit und ohne Ausreisser durchgeführt werden.

2.4. Skalenniveau

Vorliegende Daten können entsprechend ihres Informationsgehaltes zur Klassifizierung ihrer Merkmale unterschiedlichen Skalenniveaus zugeordnet werden. Die Verdichtung der Information dieser Daten durch Maßzahlen (bspw. Mediane, Mittelwerte, Varianzen, etc.) erfolgt durch eine mathematische Transformation dieser Daten. Zur Vermeidung von Informationsverzerrungen oder Informationsverfälschungen ist das jeweilige Signifikanzniveau zu beachten.

Die gewählte Skala bestimmt die Art und Aussagekraft statistischer Hypothesentests. Ein wesentlicher Unterschied zwischen den Skalenarten besteht im Informationsgehalt. Dabei nimmt dieser entsprechend den nachfolgend aufgeführten Skalen zu. Das Skalenniveau bestimmt die zulässigen mathematischen Operationen, welche auch auf Variablen des jeweils höheren Skalenniveaus durchgeführt werden können, jedoch nicht umgekehrt. Das Skalenniveau gibt keine Auskunft darüber, ob eine Variable diskret (kategorial) oder stetig ist. Nur bei einer Nominal- und Ordinalskalierung ist diese diskret.

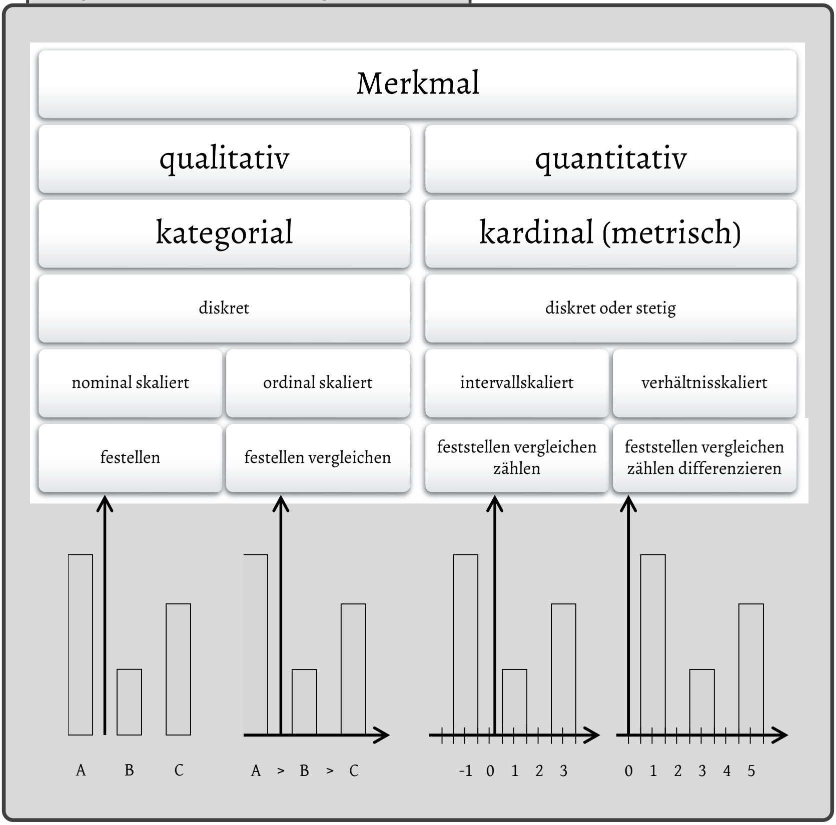
Die Nominalskala und Ordinalskala bilden die kategoriale Skala und die Intervallskala und Verhältnisskala die kardinale Skala. Diese wiederum lassen sich in diskrete und stetige Skalen unterscheiden, wodurch wiederum die Wahl des jeweiligen Hypothesentests bestimmt wird. Für die Anwendung statistischer Methoden ist es jedoch selten relevant, ob eine Intervall- oder eine Verhältnisskala vorliegt.
Nicht immer ist die Wahl des Skalenniveaus für eine Variable eindeutig, insbesondere bei der Nominal- und Ordinalskala[3].

Die Wahl einer Skala erfolgt oft durch Plausibilitätsschluß. Ausgehend von dem Wissen, daß Skalenniveaus hierarchisch angeordnet sind, läßt eine Ordinalskala alle Aus-

[3] Beispiel Schulnoten: Ordinalskala oder Intervallskala ? Mögliche Wahl: Intervallskala. Die Durchschnittsnote entspricht dem arithmetischen Mittelwert bei Annahme von Äquidistanz.

sagen zu, welche über eine Nominalskala möglich sind, gleichfalls eine Intervallskala, welche Eigenschaften einer Ordinal- und Nominalskala impliziert. Vorhandene Daten können daher mindestens einer Nominalskala zugeordnet werden. In einem nächsten Schritt ist zu prüfen, on die nominalen Kategorien dieser Skala in einer Rangfolge angeordnet werden können. Wären diese ordinalskalierten Daten aufgrund gleicher Abstände addierbar oder substrahierbar, so kann eine Intervallskala zugrunde gelegt werden. Falls die Daten miteinander dividiert oder multipliziert werden können, kann eine Verhältnisskala verwendet werden.

Begriffe und metrische Eigenschaften



Obiges Schema zeigt eine Übersicht der Begriffe und Eigenschaften verwendeter Skalen.

Nachstehend sind wesentliche formalen und statistischen Zulässigkeiten der verwendeten Skalenniveaus aufgeführt.

Skalenniveaus: formale und statistische Zulässigkeit

	Nominal -Skala	**Ordinal -Skala**	**Intervall -Skala**	**Verhältnis -Skala**
	kategorial		kardinal (metrisch)	
	diskret		diskret und stetig	
Beispiele	Steuerklasse Geschlecht Familienstand Postleitzahl Beruf Schulart	Schulnoten Plazierungen Zufriedenheit Windstärke Risikoeinstufung Angstzustände	Datum Temperatur (°C) Likertskala IQ Wasserstand Luftdruck	Körpergewicht Temperatur (K) EUR Alter Geburtenrate Frauenanteil
Charakteristika	n. sortierbar	sortierbar	relativer NP	absoluter NP
Kennwerte	Modalwert	Median	arith. Mittel	geom. Mittel
Mathematische Operationen	$= / \neq$	$= / \neq$ $< / >$	$= / \neq$ $< / >$ $- / +$	$= / \neq$ $< / >$ $- / +$ $\times / \div$

2.4.1. Nominalskala

Vorliegende Daten sind dann nominal skaliert, wenn diese unterschieden werden können, aber keine Rangfolge vorhanden ist. Die Skala besteht aus Werten, Zeichen oder Benennungen, zwischen denen keine Ordnungsrelation besteht. Für verschiedene Objekte oder Ausprägungen wird mittels eines Vergleichs lediglich eine Entscheidung über Gleichheit oder Ungleichheit dieser Merkmalsausprägung getroffen. Es liegen keine Abstandsinformationen zwischen den Merkmalsausprägungen vor. Die Werte können nicht nach ihrer Größe bzw. Eigenschaft sortiert werden.
Mit der Nominalskala wird nur eine qualitative Klassifizierung der Objekte vorgenommen, es ergeben sich qualitative Merkmale[4]. Zur Beschreibung der Ergebnisse eignet sich eine Häufigkeitsdarstellung sowie die Angabe des Modalwertes, welcher diejenige Kategorie beschreibt, welche am häufigsten auftritt.

[4] Besteht die nominale Skala nur aus zwei möglichen Werten, so wird diese als dichotome oder binäre Skala bezeichnet (bspw. männlich-weiblich, reaktionär-progressiv.)

Nominalskala

Merkmalsausprägung:	Unterscheidung auf Gleichheit (=, ≠)
Zuordnung der Werte/Zahlen:	Benennung
Mathematische Operationen:	Häufigkeitsvergleiche
Beispiel:	Familienstand:

□ ledig	1
□ verheiratet	2
□ getrennt lebend	3
□ geschieden	4
□ verwitwet	5
□ verpartnert	6
□ entpartnert	7
□ partnerhinterblieben	8

sowie: Kfz-Kennzeichen, Geschlechter von Lebewesen, Farbnamen, Postleitzahlen, Sprachen, Familienstand, Autonummern, Nationalitäten, Sportarten, Berufsbezeichnungen, Häufigkeitsdaten, Produktausfälle, ...

2.4.2. Ordinalskala

Bestehen vorliegende Daten aus Werten oder Zeichen, zwischen denen zwar eine Ordnungsrelation, aber keine Abstände definiert sind, so sind diese ordinalskaliert. Diesen Daten werden Rangzahlen zuordnet. Es ergibt sich ein ranggeordnetes, qualitatives Merkmal, abgestuft in geordnete Kategorien, jedoch besteht noch keine quantitativ messbare Beziehung zwischen diesen, lediglich eine Rangfolge ohne Information der Abstände zwischen diesen. Eine geeignete Beschreibung erfolgt wiederum durch eine Häufigkeitsaufzählung sowie die Angabe des Medians, welcher in der Mitte der nach Größe geordneten Daten liegt.
Eine Sonderform der Ordinalskala ist die Rangskala, bei der jeder Wert nur einmal vergeben wird.

Ordinalskala

Merkmalsausprägung:	Unterscheidung auf Gleichheit (=, $\neq$) Rangordnung ($<$, $>$)
Zuordnung der Werte/Zahlen:	Rangordnung
Mathematische Operationen:	Häufigkeitsvergleiche Rangordnungskorrelationen
Beispiel:	Produktzufriedenheit

☐ sehr zufrieden	1
☐ ziemlich zufrieden	2
☐ zufrieden	3
☐ unzufrieden	4
☐ ziemlich unzufrieden	5
☐ sehr unzufrieden	6

sowie: Benotungen, Handelsklassen, Dienstgrade, Ausprägungen (stark-mittel-schwach), Temperaturen (kalt-lauwarm-warm-heiß), Lautstärken (still-leise-laut), Gehaltsgruppen, Gütekriterien, ...

2.4.3. Intervallskala

Können bei vorliegenden Daten die Differenzen zwischen zwei Kategorien gemessen werden ohne Vorhandensein eines Nullpunktes, dann sind diese intervallskaliert. Die Intervallskala besteht aus Zahlen, zwischen denen gleiche Intervalle bestimmbar sind. Der Nullpunkt wird willkürlich festgelegt, er ist relativ. Es können nur Differenzen, nicht Quotienten, von Messwerten bestimmt werden.
Intervallskalierte Daten, welche auch als metrische Daten bezeichnet werden, können neben dem Modalwert und dem Median auch durch den arithmetischen Mittelwert beschrieben werden, welcher als Quotient aus Summe der Werte und der Anzahl errechnet wird. Die Reihenfolge der Merkmalswerte und die Größe des Abstandes zwischen zwei Werten ist gegeben und sachlich begründet.
Die Intervallskala als metrische Skala gibt Information über den Betrag der Unterschiede zwischen zwei Werten. Die metrischen Skalen werden in diskrete und stetige Merkmale unterteilt.

Intervallskala

Merkmalsausprägung:	Unterscheidung auf Gleichheit (=, $\neq$) Rangordnung ($<, >$) Gleichgroße Intervalle	
Zuordnung der Werte/Zahlen:	Zahlen mit gleicher Differenz ohne natürlichem Nullpunkt	
Mathematische Operationen:	Häufigkeitsvergleiche Rangordnungskorrelationen Intervallrelationen (Addition, Subtraktion)	
Beispiel:	Merkmal	Nullpunkt
	Temperatur in K	Absoluter Nullpunkt
	Preise	Kostenlos
	Masse	keine Masse
	Prozentsatz	0 %

sowie: Temperaturen, Längengrade, Wasserpegel, Zeiten, Intelligenzquotienten, Prüfungsnoten, ...

2.4.4. Verhältnisskala

Können bei vorliegenden Daten die Differenzen zwischen zwei Kategorien gemessen werden bei Vorhandensein eines absoluten Nullpunktes, dann sind diese verhältnisskaliert. Bei verhältnisskalierten Daten, welche auch als metrische Daten bezeichnet werden, können Produkte und Quotienten der Meßwerten gebildet werden. Verhältnisskalen besitzen das höchste Skalenniveau.

Verhältnisskala

Merkmalsausprägung:	Unterscheidung auf Gleichheit ($=, \neq$) Rangordnung ($<, >$) Gleichgroße Intervalle Natürlicher, absoluter Nullpunkt
Zuordnung der Werte/Zahlen:	Null entspricht der Abwesenheit des Merkmals
Mathematische Operationen:	Häufigkeitsvergleiche Rangordnungskorrelationen Intervallrelation (Addition, Subtraktion) Verhältnisrelationen (Multiplikation, Division)
Beispiel:	Angabe der Temperatur in Kelvin Standardisierte IQ-Punkte Ausprägung von Religiösität

sowie: Blutdruck, absolute Temperatur, Einkommen, Lebensalter, Gewicht, Längenmaß, Volumen, Einwohnerzahl, Geschwindigkeit, . . .

2.5. Statistischer Hypothesentest

Mit einem statistischen Hypothesentest kann auf Basis einer zufälligen Stichprobe anhand einer Teststatistik oder Prüfgröße entschieden werden, ob eine Behauptung bezüglich der Verteilungseigenschaften einer Grundgesamtheit beibehalten oder abgelehnt wird.
Wesentliche Aufgabe bei einer vorliegenden Stichprobe ist auch bei einem statistischen Hypothesentest die Bestimmung einer Stichprobenverteilung. Nur wenn diese mit größtmöglicher Aussagewahrscheinlichkeit ermittelt wurde, können weitere statistische Analysen mit entsprechend belastbaren Interpretationen und Folgerungen auch zu plausiblen und konsensfähigen Entscheidungen führen. Vorgehensweisen zur Identifikation der einer Stichprobe zu Grunde liegenden Grundgesamtheit, also die bei höchstem Signifikanzniveau (welches ausschließlich die Stichprobeninformation berücksichtigt) anzunehmende, theoretische Verteilung, sind ausführlich in [hw] beschrieben.
Bei Wahrheit dieser Nullhypothese sollte die interessierende Eigenschaft, beschrieben durch eine festzulegende Prüfgröße (Teststatistik), innerhalb eines festzulegenden Beibehaltungsbereiches liegen und bei zu großer Distanz innerhalb eines festzulegenden Ablehnbereiches liegen. Diese einer Grundgesamtheit zufällig entnommenen

Stichprobenelemente lassen auch zufälligerweise Eigenschaften einer Grundgesamtheit annehmen, welche zu Fehlentscheidungen führen können. Fehlentscheidungen werden durch zugehörige Fehlerwahrscheinlichkeiten charakterisiert. Durch die Wahl einer Fehlerwahrscheinlichkeit ist ein entsprechendes Signifikanzniveaus festgelegt, oberhalb bzw. unterhalb eines solchen eine Hypothese als signifikant bzw. nicht signifikant gilt. Es wird daher auch der Ausdruck „Signifikanztest“ verwendet.
Nachstehend sind die erforderlichen Komponenten eines solchen Signifikanztests aufgeführt.

Komponenten eines Signifikanztests

- Formulierung der Fragestellung (Problem)
- Formulierung der Nullhypothese H_0
- Formulierung der Alternativhypothese H_1
- Festlegung einer Fehlerwahrscheinlichkeit α
- Festlegung einer Fehlerwahrscheinlichkeit β
- Stichprobenentnahme n
- Auswahl eines Hypothesentests
- Wahl der Seitigkeit
- Berechnung der Teststatistik
- Testentscheidung
- Interpretation des Ergebnisses (Fazit)

2.5.1. Grundgesamtheit N

Unter einer Grundgesamtheit (Population) sollen alle diejenigen untersuchbaren Elemente zu verstehen sein, welche ein gemeinsames Merkmal besitzen. Grundgesamtheiten können einen begrenzten oder theoretisch unbegrenzten Umfang aufweisen.

2.5.2. Stichprobe n

Ist aus prinzipiellen, zeitlichen oder finanziellen Gründen eine Grundgesamtheit nicht untersuchbar, kann eine Stichprobe Informationen bzgl. des Auftretens interessierender Merkmale geben. Dabei sind die Stichprobenelemente so auszuwählen, daß alle Stichprobenelemente voneinander unabhängig sind und mit der gleichen Wahrscheinlichkeit ausgewählt werden können. Solche zufällig ausgewählten Stichproben verzerren das Ergebnis nicht, da nur zufällige Abweichungen auftreten, welche sich bei entsprechend häufigen Wiederholungen ausgleichen.
Werden Stichproben nicht zufällig ausgewählt, können systematische Fehler das Ergebnis zusätzlich verfälschen und eine solche Stichprobe wäre als nicht repräsentativ

für die zugrunde liegende Grundgesamtheit zu betrachten. Je kleiner die Stichprobe bei zunehmend heterogener Grundgesamtheit, desto weniger hinreichend ist deren Repräsentation (Ad-hoc-Stichproben).

2.5.3. Nullhypothese H_0

Basis eines statistischen Tests ist die Nullhypothese H_0 als Ausgangshypothese unter der Annahme, daß diese der Realität, also der Wahrheit entspricht. Mit Hilfe einer Nullhypothese wird überprüft, ob das Auftreten eines bestimmten Ereignisses innerhalb einer Stichprobe mit den theoretischen Annahmen, welche in einer Nullhypothese formuliert werden, auch mit dieser korrespondiert, es also zwischen den empirischen Daten und einer zugrunde gelegten theoretischen Verteilungsfunktion kein Unterschied besteht. Die Nullhypothese bewahrt den „Status quo “.
Die Konstruktion des statistischen Hypothesentests ist derart aufgebaut, daß eine Nullhypothese H_0 beibehalten oder abgelehnt wird. Eine Nullhypothese H_0 wird dann beibehalten, wenn eine aus dem Hypothesentest resultierende Irrtumswahrscheinlichkeit größer als ein vor dem Hypothesentest vereinbartes Signifikanzniveau α ist.
Entspricht die Nullhypothese der Wahrheit, so kann diese nur schwer abgelehnt werden. Ist die Nullhypothese jedoch signifikant falsch („Null und nichtig“), so kann dieses auch mit großer Wahrscheinlichkeit erkannt werden. Es wird diejenige Annahme zur Nullhypothese gemacht, welche von besonderem Interesse ist bzw. bei der eine befürchtete Fehlentscheidung zu einem Fehler 1. Art wird.
Die Ablehnung der Nullhypothese führt zu einer Annahme der Arbeitshypothese. Die Beibehaltung der Arbeitshypothese[5] bedeutet nicht deren erwiesene Richtigkeit, sondern nur, daß die Eigenschaft der vorhandenen Stichprobe nicht zu deren Ablehnung ausreicht („Freispruch mangels Beweises“).
Die Nullhypothese wird bei richtiger Entscheidung mit einer statistischen Aussagesicherheit von $P = 1 - \alpha$ angenommen.

2.5.4. Alternativhypothese H_1

Für die Überprüfung der Richtigkeit einer Aussage wird der Nullhypothese eine innovative Alternative, eine Alternativhypothese gegenübergestellt. Die Alternativhypothese H_1, auch als Arbeitshypothese bezeichnet, stellt im Allgemeinen die Negation einer Nullhypothese dar. Diese wird indirekt bestätigt, indem eine Nullhypothese abgelehnt, also als unplausibel verworfen wird.
Eine Alternativhypothese H_1 wird angenommen, wenn eine aus dem Signifikanztest resultierende Irrtumswahrscheinlichkeit kleiner oder gleich einem vor dem Signifikanztest vereinbarten Signifikanzniveau α ist. Es wird diejenige Annahme zur Arbeitshypothese gemacht, welche nachgewiesen oder statistisch gestützt werden soll.
Die Alternativhypothese wird bei Ablehnung der Nullhypothese mit einer statistischen Aussagesicherheit von $(1 - \beta)$ angenommen.

[5] Im Rahmen dieses Buches als Alternativhypothese bezeichnet.

2.5.5. Seitigkeit von Hypothesen

Damit eine Hypothese passend zu vorhandenen Daten formuliert wird, ist vor einem durchzuführenden Hypothesentest festzulegen, ob die Art des Hypothesentests einseitig (gerichtet, zusammengesetzt) oder zweiseitig (ungerichtet, einfach) sein soll.

Seitigkeit eines Signifikanztests – Beispiel

H_0: $\widetilde{\mu}_{\text{vorher}} = \widetilde{\mu}_{\text{nachher}}$: Es gibt keinen Unterschied von interessierenden Eigenschaften vor und nach einer Behandlung.

Bei einem einseitig durchzuführenden Hypothesentest wird die Richtung eines Unterschieds vorgegeben:

H_1: (i) $\widetilde{\mu}_{\text{vorher}} < \widetilde{\mu}_{\text{nachher}}$: Die interessierenden Eigenschaften vor einer Behandlung sind besser als nach einer Behandlung,

(ii) $\widetilde{\mu}_{\text{vorher}} > \widetilde{\mu}_{\text{nachher}}$: Die interessierenden Eigenschaften vor einer Behandlung sind schlechter als nach einer Behandlung.

Bei einem zweiseitig durchzuführenden Hypothesentest bleibt die Richtung eines Unterschieds ungerichtet. Es wird irgendein Unterschied postuliert:

H_1: (iii) $\widetilde{\mu}_{\text{vorher}} \neq \widetilde{\mu}_{\text{nachher}}$: Die interessierenden Eigenschaften vor einer Behandlung sind besser oder schlechter als nach einer Behandlung.

2.5.6. Fehlentscheidung 1. Art α

Die formulierten Hypothesen beschreiben die Verhältnisse in der Grundgesamtheit, hingegen sich das Ergebnis eines Hypothesentests nur auf die Stichprobe bezieht.
Aufgrund eines Stichprobenbefunds besteht die Wahrscheinlichkeit, eine Fehlentscheidung bezüglich der zugehörigen Grundgesamtheit zu treffen.
Das irrtümliche Ablehnen einer Nullhypothese, obwohl diese zutrifft, wird als Fehler 1. Art oder α-Fehler bezeichnet[6], siehe nachfolgende Übersicht.
Die maximale Eintrittswahrscheinlichkeit wird in Form eines Signifikanzniveaus, der Irrtumswahrscheinlichkeit α vorgegeben. Je kleiner diese gewählt wird, desto unwahrscheinlicher wird diese Fehlentscheidung. Mit zunehmend kleinerer Irrtumswahrscheinlichkeit wird die Nullhypothese immer seltener irrtümlicherweise abgelehnt. Jedoch steigt hierdurch auch die Wahrscheinlichkeit eine falsche Nullhypothese nicht abzulehnen. Für $\alpha = 0$ würde die Nullhypothese immer beibehalten werden.
Wird die Nullhypothese abgelehnt, so beträgt das kontrollierte Risiko einer Fehlentscheidung, die Nullhypothese wäre doch richtig, hätte also beibehalten werden müssen, α.

[6] Ein Fehler 1. Art kann auch als „falscher Alarm“ verstanden werden.

2.5.7. Fehlentscheidung 2. Art β

Das irrtümliche Beibehalten einer Nullhypothese, obwohl diese nicht zutrifft, wird als Fehler 2. Art oder β-Fehler bezeichnet[7] und ist von den wahren Eigenschaften einer Grundgesamtheit abhängig. Die Wahrscheinlichkeit einen solchen Fehler zu begehen, beträgt β. Die Wahrscheinlichkeit eine richtige Arbeitshypothese H_1 als solche zu erkennen, beträgt $1 - \beta$. Sinkt die Irrtumswahrscheinlichkeit α, desto geringer ist die Wahrscheinlichkeit, die Nullhypothese irrtümlicherweise abzulehnen, jedoch wird mit zunehmender Wahrscheinlichkeit die Nullhypothese irrtümlicherweise beibehalten.
Der Fehler 2. Art wird umso größer, je kleiner α bei gegebenen Stichprobenumfang gewählt wird. Bei kleinen Stichprobenumfängen und kleinem α können vorhandene Unterschiede kaum nachgewiesen werden. Entsprechende Ergebnisse sind als riskant einzustufen. Mit zunehmend kleinem α und β können vorhandene Unterschiede nur mit sehr großen Stichprobenumfängen nachgewiesen werden.
Wird die Nullhypothese beibehalten, so beträgt das nichtkontrollierte Risiko einer Fehlentscheidung, die Alternativhypothese ist doch richtig, hätte also abgelehnt werden müssen, β , siehe nachfolgende Übersicht.
Für einen Signifikanztest kann keine vorgegebene Irrtumswahrscheinlichkeit β eingehalten werden (diese ergibt sich aus der Irrtumswahrscheinlichkeit α, dem Stichprobenumfang n und dem Unterschied der Aussagen). Etliche Hypothesentests entscheiden bei vorgegebenem α und unbestimmten β zugunsten der Nullhypothese (konservative Tests).

Fehlerarten und Aussagesicherheiten eines Signifikanztests

	Tatsächlicher Sachverhalt	
	H_0 richtig	H_1 richtig
Beibehaltung von H_0	richtige Entscheidung $P(H_0 \mid H_0) = 1 - \alpha$ (Aussagesicherheit)	falsche Entscheidung $P(H_0 \mid H_1) = \beta$ (Fehler 2. Art, β-Fehler)
Annahme von H_1	falsche Entscheidung $P(H_1 \mid H_0) = \alpha$ (Fehler 1. Art, α-Fehler)	richtige Entscheidung $P(H_1 \mid H_1) = 1 - \beta$ (Aussagesicherheit)

2.5.8. Festlegung von α, β und n

Das Signifikanzniveau α ist die vom Anwender vorgegebene, maximale Wahrscheinlichkeit dafür, daß die Nullhypothese abgelehnt wird, obwohl diese wahr ist. Parallel zum Signifikanzniveau α ist auch der Stichprobenumfang n festzulegen, mit welchem

[7] Ein Fehler 2. Art kann auch als „versäumter Alarm“ verstanden werden. (Ein Fehler 3. Art wäre die Wahl einer falschen Methodik oder eines falschen Modells.)