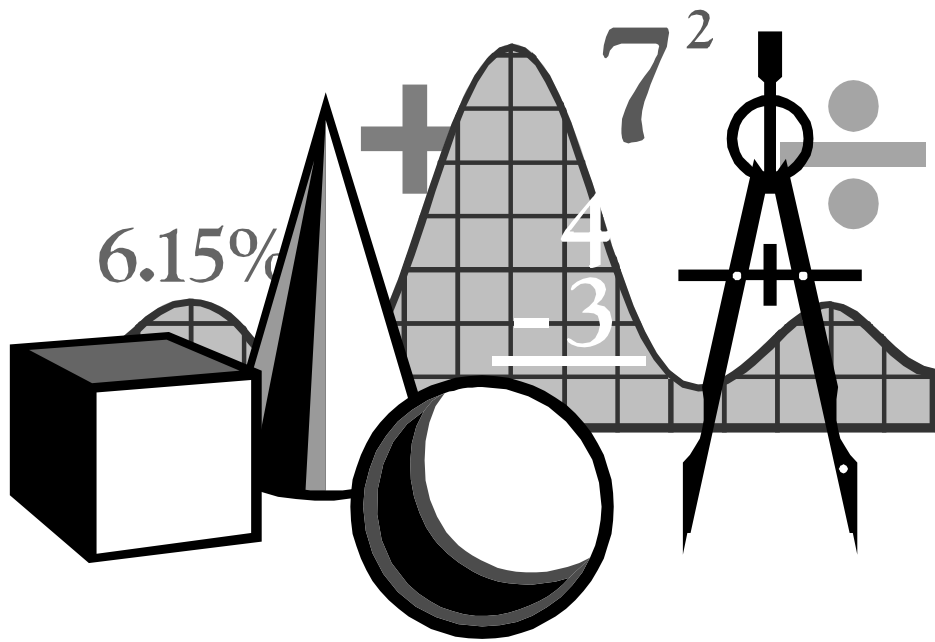


Statistik II



Inferenz-Statistik

Vorwort zur 4. Fassung

Das vorliegende Scriptum soll eine Brücke schlagen zwischen der Vorlesung von Herrn Dr. Huber und der gängigen Fachliteratur. Die einzelnen von ihm vorgestellten inferenzstatistischen Verfahren werden der Reihe nach aufgegriffen, wobei auf nur für den Fachmann verständliche mathematische Ableitungen, wie dies in Lehrbüchern üblich ist, verzichtet wird. Statt dessen wird versucht, die Verfahren auf anschaulichere Begriffe zurückzuführen und mit zahlreichen Abbildungen und Beispielen zu verdeutlichen.

Die Themen Mittelwertvergleiche von zwei Stichproben, Varianzanalysen und Interaktionen werden etwas gründlicher dargestellt, da diese in der Veranstaltung Versuchsplanung bei Herrn Prof. Dr. Hussy behandelt werden und daher im Vordiplom (Fach Methodenlehre) durchaus vereinzelt geprüft werden.

Für das Gelingen der Klausur wünsche ich viel Glück.

Köln, im Mai 2001

Haug Leuschner
Diplom-Psychologe

**Copyright © 1995 – 2001 Haug Leuschner
Email Haug.Leuschner@netcologne.de**

INHALTSVERZEICHNIS

1. Stichprobenverteilungen 6

- 1.1. Der zentrale Grenzwertsatz 6
- 1.2. Die Durchschnittsverteilung 7
- 1.3. Die t-Verteilung 8
- 1.4. Stichprobenverteilungen von Häufigkeiten 9

2. Signifikanztests I 11

- 2.1. Allgemeine Vorgehensweise bei einem Signifikanztest 11
 - 2.1.1. Formulierung der statistischen Hypothesen 11
 - 2.1.2. Festlegung des Signifikanzniveaus 12
 - 2.1.3. Auswahl des Testverfahrens 13
 - 2.1.4. Ziehung der Stichprobe und Ermittlung der Stichprobenkennwerte 15
 - 2.1.5. Berechnung der Prüfgröße 15
 - 2.1.6. Bestimmung der kritischen Grenze 15
 - 2.1.7. Entscheidung über Annahme oder Ablehnung der Nullhypothese 16
- 2.2. Fehlerarten bei statistischen Entscheidungen 17
- 2.3. Zur Bedeutung der Bestätigung oder Ablehnung einer Hypothese 19
- 2.4. Variante des Tests: Die Überschreitungswahrscheinlichkeit P 21
- 2.5. Einseitige Signifikanztests 23
 - 2.5.1. Rechtsseitiger Test 24
 - 2.5.2. Linksseitiger Test 25
 - 2.5.3. Die Überschreitungswahrscheinlichkeit P bei einseitigen Tests 26
 - 2.5.4. Vereinfachung der Annahmebedingung für die H_0 -Hypothese 27
- 2.6. Signifikanztests bei Häufigkeiten (Approximativer Binomialtest) 28

3. Konfidenzintervalle 29

- 3.1. Konfidenzintervall eines Parameters μ mit bekanntem σ 31
- 3.2. Konfidenzintervall eines Parameters μ mit unbekanntem σ 31
- 3.3. Konfidenzintervall eines Parameters π 31
- 3.4. Mathematische Herleitung zum Konfidenzintervall 33
- 3.5. Steigerung der Präzision des Konfidenzintervalls 33

4. Signifikanztests II 34

- 4.1. Prüfung von Korrelationshypothesen 34
- 4.2. Prüfung von Unterschiedshypothesen 35
 - 4.2.1. Vergleich der Mittelwerte zweier abhängiger Stichproben 38
 - 4.2.2. Vergleich der Mittelwerte zweier unabhängiger Stichproben 39
 - 4.2.2.1. z-Test bei bekanntem σ 40
 - 4.2.2.2. WELCH-Test 41
 - 4.2.2.3. Klassischer t-Test 42
 - 4.2.2.4. Prüfung auf Gleichheit zweier Stichprobenvarianzen (F-Test) 44

Varianzanalysen 46

- 5.1. α -Adjustierung 47
- 5.2. Einfaktorielle Varianzanalyse 48
 - 5.2.1. Unabhängige Stichproben 48
 - 5.2.1.1. Verfahrensweise 48
 - 5.2.1.2. Varianzaufklärung und Effektstärke durch Eta-Quadrat 51
 - 5.2.1.3. Multipler Mittelwertvergleich 51
 - 5.2.2. Abhängige Stichproben 52
 - 5.2.2.1. Verfahrensweise 53
- 5.3. Mehrfaktorielle Varianzanalyse 57
 - 5.3.1. Verfahrensweise 58
 - 5.3.1.1. Gesamteffekt 60
 - 5.3.1.2. Haupteffekt Faktor A 60

- 5.3.1.3. Haupteffekt Faktor B 61
- 5.3.1.4. Wechselwirkung AxB 62
- 5.3.2. Tafel der Varianzanalyse 63
- 5.3.3. Varianzaufklärung und Effektstärken 63
- 5.3.4. Interaktion zwischen den Faktoren 67
 - 5.3.4.1. Nullinteraktion 67
 - 5.3.4.2. Interaktionstypen 68
 - 5.3.4.3. Signifikanztests für einfache Haupteffekte 71
- 5.4. Voraussetzungen zu den Varianzanalysen 74
- 5.5. Varianzanalyse und klassischer t-Test 74

6. Nichtparametrische, verteilungsfreie Verfahren 75

- 6.1. Verfahren zur Binomialverteilung 75
 - 6.1.1. Die Binomialverteilung 75
 - 6.1.2. Approximativer Binomialtest 79
 - 6.1.3. Exakter Binomialtest 79
 - 6.1.3.1. Symmetrische Verteilung 79
 - 6.1.3.2. Asymmetrische Verteilungen 81
- 6.2. Verfahren zur Chi-Quadrat-Verteilung 83
 - 6.2.1. χ^2 -Unabhängigkeitstest 83
 - 6.2.2. χ^2 -Anpassungstest 85
 - 6.2.3. χ^2 - und z-Werte 85

7. Klausurvorbereitung 87

- 7.1. Anleitung 87
 - 7.1.1. Stichprobenverteilungen 87
 - 7.1.2. Signifikanztests I 87
 - 7.1.3. Konfidenzintervalle 88
 - 7.1.4. Signifikanztests II 88
 - 7.1.5. Varianzanalysen 89
 - 7.1.6. Nichtparametrische, verteilungsfreie Verfahren 90
- 7.2. Klausuren 91
 - 7.2.1. Klausur 1989 91
 - 7.2.2. Klausur 1990 92
 - 7.2.3. Klausur 1991 93
 - 7.2.4. Klausur 1992 94
 - 7.2.5. Klausur 1993 95
 - 7.2.6. Klausur 1994 96
 - 7.2.7. Klausur 1994 98
- 7.3. Lösungen zu den Fragen 99
 - 7.3.1. Stichprobenverteilungen 99
 - 7.3.2. Signifikanztests I 99
 - 7.3.3. Konfidenzintervalle 100
 - 7.3.4. Signifikanztests II 100
 - 7.3.5. Varianzanalysen 101
 - 7.3.6. Nichtparametrische, verteilungsfreie Verfahren 102
- 7.4. Lösungen zu den Klausuraufgaben 103
 - 7.4.1. Klausur 1989 103
 - 7.4.2. Klausur 1990 105
 - 7.4.3. Klausur 1991 107
 - 7.4.4. Klausur 1992 108
 - 7.4.5. Klausur 1993 109
 - 7.4.6. Klausur 1994 110
 - 7.4.6. Klausur 1996 112

MERKBLATT FÜR SIGNIFIKANZTESTS 115

INDEX 116

VERZEICHNIS DER BEISPIELE ZU DEN VERFAHREN

2. Signifikanztests I

Zweiseitig

Stichprobenmittelwertvergleich mit dem Populationserwartungswert μ (mit σ bekannt) 19

Stichprobenmittelwertvergleich mit Populationserwartungswert μ (mit σ unbekannt) 20

Stichprobenmittelwertvergleich mit Hilfe der Überschreitungswahrscheinlichkeit P 22

Einseitig

Vergleich der Ausprägungshäufigkeit p einer Stichprobe mit dem Populationsparameter π
(Approximativer Binomialtest) 28

3. Konfidenzintervalle

Beispiele für Konfidenzintervalle 32

4. Signifikanztests II

Überprüfung von Korrelationshypthesen 35

Vergleich zweier Mittelwerte von abhängigen Stichproben 38

z-Test zum Mittelwertvergleich zweier unabhängiger Stichproben 40

WELCH-Test zum Mittelwertvergleich zweier unabhängiger Stichproben 42

Klassischer t-Test zum Mittelwertvergleich zweier unabhängiger Stichproben 44

Prüfung auf Gleichheit zweier Stichprobenvarianzen (Voraussetzung klass. t-Test) 45

5. Varianzanalysen

Einfaktorielle Varianzanalyse für unabhängige Stichproben 50

Multipler Mittelwertvergleich 51

Einfaktorielle Varianzanalyse für abhängige Stichproben 56

Zweifaktorielle Varianzanalyse 64

Ermittlung der Interaktion zwischen den Faktoren 72

6. Nichtparametrische, verteilungsfreie Verfahren

Exakter Binomialtest bei symmetrischer Verteilung 82

Exakter Binomialtest bei asymmetrischer Verteilung 82

χ^2 -Unabhängigkeitstest 84

χ^2 -Anpassungstest 85

1. Stichprobenverteilungen

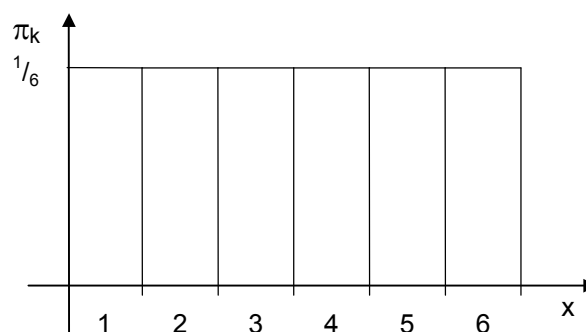
In der experimentellen Forschung werden Stichproben aus einer Grundgesamtheit gezogen und z.B. der Mittelwert eines Merkmals, das untersucht wird, ermittelt, um auf die Gesamtheit zu schließen. Dies wird gemacht, um eine Hypothese über ein Merkmal in einer Gesamtheit zu stützen oder zu widerlegen. Nach welchen Kriterien kann entschieden werden, wie zutreffend der Durchschnittswert der Stichprobe die durchschnittliche Ausprägung des interessierenden Merkmals in der Grundgesamtheit repräsentiert? Oder: Wie brauchbar ist dieser Durchschnittswert \bar{x} als Schätzwert für den Parameter μ ?

Angenommen, man entnimmt mehrere Stichproben aus der Grundgesamtheit und vergleicht diese, so wird man feststellen, dass diese aufgrund zahlreicher Einflüsse voneinander abweichen. Als einen genaueren Schätzwert für μ könnte man den Durchschnitt dieser Durchschnitte der Stichproben bilden in der Hoffnung, einen besseren Wert für μ zu erhalten. Dies ist die empirische Vorgehensweise.

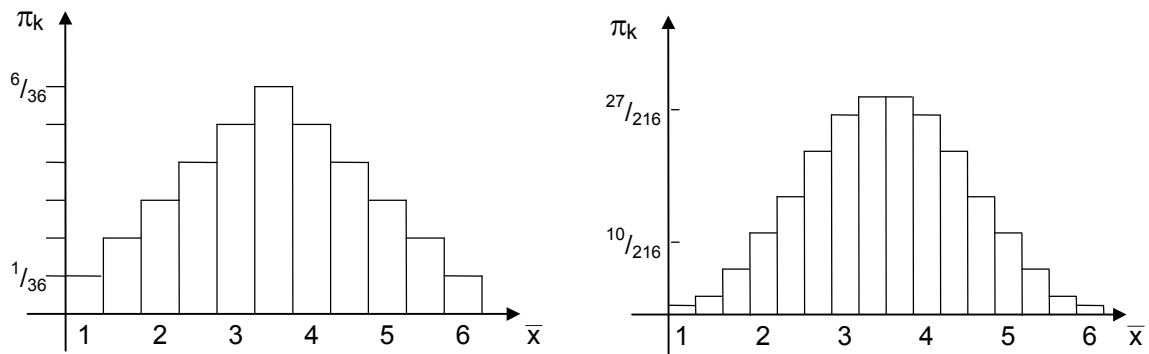
Theoretisch wird eine Stichprobenverteilung aufgestellt. Eine Stichprobenverteilung ist die Wahrscheinlichkeitsverteilung eines statistischen Kennwertes (Maß der zentralen Tendenz: Arithmetisches Mittel, Median; Streuungsmaß: Standardabweichung, Varianz, durchschnittliche Abweichung etc.) von allen möglichen Stichproben (mit dem Umfang n), die aus einer Grundgesamtheit gezogen werden können. Sie gibt damit an, wie wahrscheinlich es ist, dass ein Stichprobenkennwert in bestimmter Weise vom zugehörigen Parameter der Grundgesamtheit abweicht.

1.1. Der zentrale Grenzwertsatz

Wir wollen zur Stichprobenverteilung folgendes Gedankenbeispiel verfolgen: Eine hypothetische Grundgesamtheit möge aus allen Ereignissen des Würfeln, d. h. den Zahlen 1 bis 6, bestehen. Die Wahrscheinlichkeitsverteilung ist in diesem Fall eine Rechteckverteilung:



Nun betrachten wir dieselbe Grundgesamtheit, entnehmen aber zufällig jeweils zwei bzw. drei Augenzahlen und betrachten in den folgenden Abbildungen links die Wahrscheinlichkeitsverteilung der Durchschnitte von je zwei Augenzahlen und rechts die Wahrscheinlichkeitsverteilung der Durchschnitte von je drei Augenzahlen (beides Stichprobenverteilungen, genauer: Durchschnittsverteilungen):



Wie wir sehen, hat die Form der Verteilung der Mittelwerte für verschiedene Anzahlen von Augenzahlen mit der Rechteckform der Grundgesamtheit nichts mehr zu tun. Mit steigenden Anzahlen n der Augenzahlen nähert sich die Durchschnittsverteilung zunehmend der Normalverteilung an. Entsprechendes gilt für jede beliebige Stichprobenverteilung einer Grundgesamtheit. Dies ist die Aussage des zentralen Grenzwertsatzes, der wie folgt formuliert wird:

Die Verteilung von Mittelwerten aus Stichproben des Umfanges n , die sämtlich derselben Grundgesamtheit entnommen wurden, geht mit wachsendem Stichprobenumfang in eine Normalverteilung über.

Für praktische Zwecke können wir davon ausgehen, dass die Durchschnittsverteilung für beliebige Verteilungsformen des Merkmals in der Grundgesamtheit bereits dann hinreichend normalverteilt ist, wenn $n \geq 30$ ist.

1.2. Die Durchschnittsverteilung

Eine Durchschnittsverteilung ist eine Stichprobenverteilung des Mittelwertes, d.h. die Wahrscheinlichkeitsverteilung der an der Gesamtheit aller möglichen Zufallsstichproben (des Umfanges n) erhaltenen Mittelwerte. Sie ist nach dem zentralen Grenzwertsatz bei hinreichend großen Stichprobenumfängen normalverteilt, auch wenn das zugrundeliegende Merkmal nicht normalverteilt ist. Bei den bisherigen Normalverteilungen haben wir die Wahrscheinlichkeitsverteilung einer normalverteilten Zufallsvariable für ein stetiges Merkmal betrachtet, jetzt betrachten wir die Wahrscheinlichkeitsverteilung einer normalverteilten Zufallsvariable, deren numerische Werte Durchschnitte von gleich großen Stichproben repräsentieren:

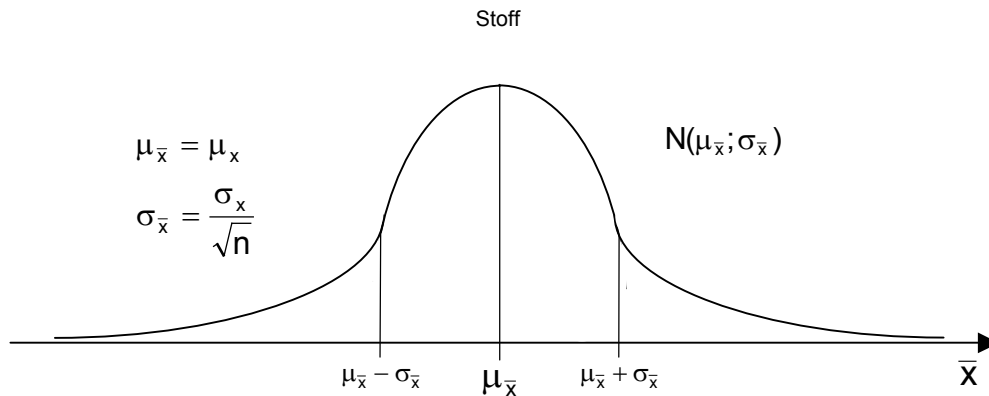
$$\bar{X}: \bar{x}_1, \bar{x}_2, \bar{x}_3, \dots$$

Die Normalverteilung ist eindeutig bestimmt durch die Angabe zweier Parameter: dem Erwartungswert μ und der Standardabweichung σ .

Beispiel: Die Körpergröße einer Grundgesamtheit sei $N(170,20)$ -verteilt. Dann haben die Körpergrößen dieser Grundgesamtheit einen Durchschnitt von $\mu=170\text{cm}$ und streuen um diesen Durchschnitt mit $\sigma=20\text{cm}$.

Die Parameter einer normalverteilten Durchschnittsverteilung sind wie folgt definiert:

$$\mu_{\bar{x}} = \mu_x \quad \sigma_{\bar{x}} = \sqrt{\frac{\sum (\bar{x}_i - \mu)^2}{n-1}} = \frac{\sigma_x}{\sqrt{n}}$$



Der Erwartungswert $\mu_{\bar{x}}$ der Durchschnittsverteilung entspricht dem Erwartungswert μ_x der Wahrscheinlichkeitsverteilung der Messwerte. Die Standardabweichung $\sigma_{\bar{x}}$ der Durchschnittsverteilung gibt die durchschnittliche Abweichung der Durchschnitte \bar{x}_i vom wahren Wert μ aufgrund vielfältiger zufälliger Einflüsse an und wird daher auch Standardfehler, Zufallsstreuung oder Fehlerstreuung genannt. Sie kann ermittelt werden, indem die Standardabweichung σ_x der Wahrscheinlichkeitsverteilung der Messwerte durch die Wurzel des Stichprobenumfangs n dividiert wird. Dadurch ergibt es sich, dass der Standardfehler $\sigma_{\bar{x}}$ mit wachsenden Stichprobenumfang n immer kleiner und damit die Durchschnittsverteilung immer schmalgipfelig wird gegenüber der ihr zugrunde liegenden Normalverteilung der Ausgangsvariable.

Beispiel: Die Durchschnittsverteilung zu obiger $N(170;20)$ -Verteilung von Körpergrößen ist bei einem Stichprobenumfang von 100 Personen eine $N(170;20/\sqrt{100})=N(170;2)$ -Verteilung. D. h. während die Körpergrößen durchschnittlich um 20cm von μ_x abweichen, weichen die Durchschnitte der Körpergrößen von Stichproben mit einem Umfang von 100 Personen nur noch um 2cm von $\mu_{\bar{x}}$ ab.

Um die Wahrscheinlichkeit zu ermitteln, dass ein Stichprobendurchschnitt z.B. kleiner als eine Grenze a ist, muss zunächst – wie in Statistik I besprochen – zur ϕ -Flächenberechnung eine z -Transformation durchgeführt werden:

$$z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \Rightarrow z = \frac{\bar{X} - \mu_x}{\frac{\sigma_x}{\sqrt{n}}}$$

Nun kann die gesuchte Wahrscheinlichkeit über die ϕ -Werte in der Tabelle der Standardnormalverteilung ermittelt werden.

Aufgabe: Wie groß ist die Wahrscheinlichkeit, dass aus obiger Grundgesamtheit zufällig eine Stichprobe von 100 Personen gezogen wird, deren durchschnittliche Körpergröße kleiner als 175cm ist?

Lösung: $W\{\bar{X} < 175\} = \phi(z(175)) = \phi(2,5) = 0,994$. Die gesuchte Wahrscheinlichkeit beträgt also 99,4%.

1.3. Die t-Verteilung

Eine Schwierigkeit in der Berechnung obiger z -Formel liegt in den meisten Fällen darin, dass die Parameter μ_x und σ_x nicht bekannt sind. Der Parameter μ_x wird dadurch gewonnen, dass er der Hypothese, also der zu prüfenden Vermutung des Forschers über μ_x , gleichgesetzt wird. Wenn die statistischen Verfahren, die wir noch kennenlernen werden, ein positives Ergebnis ergeben, wird der vom Forscher angenommene Wert von μ_x als bestätigt angesehen.

Anders sieht es mit der Streuung σ_x der Messwerte der Grundgesamtheit aus. Die ist in den meisten Fällen genau so wenig bekannt wie μ_x . In diesem Fall greift man auf die Standardabweichung s der Stichprobe als Schätzwert für σ_x zurück und berechnet statt einem z-Wert einen t-Wert:

$$t = \frac{\bar{x} - \mu_x}{\frac{s}{\sqrt{n}}}$$

Bei diesem Verfahren ersetzt die t-Verteilung die Standardnormalverteilung, wobei die t-Verteilung berücksichtigt, dass die Standardabweichung s einer Stichprobe genauso zufällig vom wahren Wert σ_x abweichen kann wie der Durchschnitt \bar{x} einer Stichprobe vom wahren Wert μ_x . Diese Abweichung ist vom Stichprobenumfang abhängig, denn je kleiner eine Stichprobe ist, desto mehr wird ihre Standardabweichung s von der wahren Streuung σ_x abweichen können. Zu jedem möglichen Stichprobenumfang gibt es daher eine eigene t-Verteilung, die sich in ihrem Freiheitsgrad df (degree of freedom) bestimmt:

$$df = n - 1$$

Mit zunehmenden Stichprobenumfang n und damit zunehmenden Freiheitsgrad df nähert sich die t-Verteilung der Standardnormalverteilung an. Ab einer Stichprobengröße von ca. 100 Versuchspersonen sind die t-Verteilungen der Standardnormalverteilung derart ähnlich, dass in den meisten Tabellen nur die Werte bis zu $df=100$ enthalten sind. Die Tabelle der t-Werte, die wir benutzen, befindet sich direkt neben der Tabelle der Standardabweichung bei HUBER. Zumeist werden t-Werte zu vier bestimmten Wahrscheinlichkeit gesucht, so dass diese Tabelle auch nur vier Wahrscheinlichkeiten berücksichtigt. Ab einem Stichprobenumfang von über 100 Probanden entspricht die t-Verteilung weitgehend der Standardnormalverteilung, so dass dann die Werte für $df=\infty$ benutzt werden, welche z-Werte der Standardnormalverteilung darstellen. Bei einem nicht aufgeführten df -Wert wird der nächstkleinere gewählt. Die Werte werden wie folgt angegeben:

$$t_{\% ; df}$$

Beispiele:

$t_{0,99;9} = 2,82$ t-Wert zu einer t-Verteilung mit 10 Probanden und einer Wahrscheinlichkeit von 99%

$t_{0,975;19} = 2,09$ t-Wert zu einer t-Verteilung mit 20 Probanden und der Wahrscheinlichkeit von 97,5%

1.4. Stichprobenverteilungen von Häufigkeiten

Die bisherigen Ausführungen bezogen sich auf *stetige, quantitative*, d.h. Messbare, und *metrische* Merkmale. Sie lassen sich auf *diskrete* und *qualitative* Merkmale, von denen nur Häufigkeiten bekannt sind, erweitern, wenn diese *dichotom* angegeben werden können. Diese Häufigkeiten werden relativ angegeben, die in der Stichprobe mit p (relative Häufigkeit) und in der hypothetischen Grundgesamtheit mit π (Wahrscheinlichkeit) bezeichnet werden. Auch von diesem Stichprobenkennwert p kann eine Stichprobenverteilung aufgestellt werden.

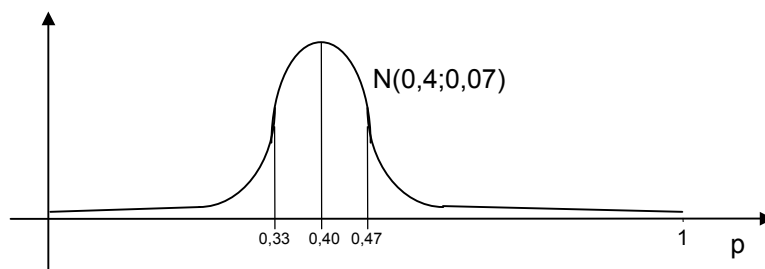
Beispiel: Es werden unendlich viele gleich große Stichproben von Personen gezogen und ihr Familienstand festgestellt: ledig, verheiratet, verwitwet und geschieden. Von jeder Stichprobe wird der relative Anteil p der Ledigen und der relative Anteil $q=1-p$ der übrigen Personen bestimmt (Dichotomisierung der Daten, d.h. auf zwei Ausprägungen reduziert). Wie sieht die Stichprobenverteilung der Häufigkeiten der Ledigen aus? Wie hoch ist der Anteil der Ledigen in der Grundgesamtheit, d.h. wie hoch ist der Durchschnittswert μ der Häufigkeiten p der Ledigen von allen Stichproben? In welchem Bereich σ streuen diese Stichprobenhäufigkeiten der Ledigen um diesen Durchschnittswert μ ?

Die gesuchte Stichprobenverteilung ist eine Binomialverteilung, die im letzten Kapitel näher behandelt wird. Mit wachsendem Stichprobenumfang n geht auch diese in eine Normalverteilung über. Der Erwartungswert μ dieser Stichprobenverteilung ist der Durchschnitt der Stichprobenhäufigkeiten p der Ausprägung eines qualitativen Merkmals von unendlich vielen Stichproben, der Standardfehler σ die durchschnittliche Abweichung dieser Häufigkeiten von μ . Diese Parameter sind wie folgt bestimmt:

$$\mu = \pi \quad \sigma = \sqrt{\frac{\pi(1-\pi)}{n}}$$

Der Erwartungswert μ entspricht dem wahren Wert π der der Stichprobenverteilung zugrunde liegenden Grundgesamtheit, also der wahren relativen Häufigkeit der betrachteten Ausprägung eines qualitativen Merkmals. Der Standardfehler σ berechnet sich aus dem Wert π und dem Stichprobenumfang n aller gleich großen Stichproben der Stichprobenverteilung.

Beispiel: Der Anteil der Ledigen in einer Bevölkerung beträgt $\pi=40\%$. Dann kann erwartet werden, dass der Durchschnittswert μ des Anteils der Ledigen von unendlich vielen Stichproben aus dieser Bevölkerung ebenfalls $\mu=\pi=40\%$ beträgt und dass die Anteile der Ledigen all dieser Stichproben bei einem Stichprobenumfang von $n=50$ Personen um durchschnittlich $\sigma=7\%$ von $\mu=40\%$ abweichen ($\mu-\sigma=33\%$ bzw. $\mu+\sigma=47\%$):



$$\mu = \pi = 0,4$$

$$\sigma = \sqrt{\frac{0,4 \cdot 0,6}{50}} = 0,07$$

Die Stichprobenverteilung der Häufigkeit einer Ausprägung eines dichotomen Merkmals, die eine Binomialverteilung ist, darf erst dann durch die oben dargestellte Normalverteilung approximiert (angenähert) werden, wenn der Stichprobenumfang n genügend groß ist. Dies wird geprüft durch:

$$n > \frac{9}{\pi \cdot (1-\pi)}$$

Beispiel: Setzt man in obige Formel 0,4 für π ein, so erhält man $n > 37,5$. Die Stichprobengröße muss also mindestens 38 Personen betragen. Im obigen Beispiel ist mit 50 Personen die Voraussetzung für die Approximierung der Stichprobenverteilung durch die Normalverteilung gegeben. Grundsätzlich ist das bei Stichprobenverteilungen von Häufigkeiten immer zu prüfen.

2. Signifikanztests I

Signifikanztests sind die am häufigsten angewandten statistischen Verfahren in der experimentellen psychologischen Forschung und auch gleichzeitig die problematischsten, weil bei der Prüfung der Voraussetzungen, der Durchführung und der Interpretation der Ergebnisse gravierende Fehler gemacht werden können. Signifikanztests (statistische Tests, Prüfverfahren) gestatten es, Hypothesen des Forschers zu überprüfen, indem statistische Kennwerte miteinander verglichen und daraufhin geprüft werden, ob diese sich *signifikant* (überzufällig) voneinander unterscheiden. Unterschiede können sich ergeben, nicht weil ein signifikanter Unterschied besteht, sondern weil der Unterschied sich aufgrund vielfältiger Einflüsse *zufällig* ergibt. Eine solche Prüfung ist unerlässlich, wenn man von dem gewonnenen empirischen Wert mehr als eine Beschreibung des in der Stichprobe festgestellten Sachverhaltes erwartet, wenn man die gefundenen Ergebnisse verallgemeinern möchte, um eine vorher aufgestellte Hypothese ablehnen oder beibehalten zu können.

2.1. Allgemeine Vorgehensweise bei einem Signifikanztest

1. Formulierung der statistischen Hypothesen
2. Festlegung des Signifikanzniveaus
3. Auswahl des Testverfahrens
4. Ziehung der Stichprobe und Ermittlung der Stichprobenkennwerte
5. Berechnung der Prüfgröße
6. Bestimmung entweder (a) der kritischen Grenze oder (b) der Überschreitungswahrscheinlichkeit
7. Entscheidung über Annahme oder Ablehnung der Nullhypothese

Siehe hierzu und im folgenden auf das Merkblatt auf S. 105!

2.1.1. Formulierung der statistischen Hypothesen

Die zu prüfende statistische Hypothese nennt man *Nullhypothese* und bezeichnet sie mit H_0 . Die Nullhypothese drückt immer die Vermutung aus, dass sich ein Stichprobenkennwert nicht „echt“, sondern nur zufällig vom entsprechenden Parameter der Grundgesamtheit unterscheidet. Genau besagt sie folgendes: *Die untersuchte Stichprobe entstammt aus einer Grundgesamtheit mit der in der Nullhypothese formulierten Aussage*. Wenn eine Nullhypothese durch das Prüfverfahren bestätigt werden konnte, sagt man, dass die Nullhypothese *verifiziert* (d.h. bestätigt) werden konnte und damit beibehalten werden kann, da die Abweichung des Stichprobenkennwertes vom Erwartungswert μ der Stichprobenverteilung auf Zufallseinflüsse zurückgeführt werden kann. Die der Nullhypothese entgegengesetzte Hypothese nennt man *Alternativhypothese* und bezeichnet sie mit H_1 . Die Alternativhypothese drückt immer die Vermutung aus, dass sich ein Stichprobenkennwert sich *signifikant*, d. h. überzufällig vom entsprechenden Parameter der Grundgesamtheit unterscheidet. Genau besagt sie folgendes: *Die untersuchte Stichprobe entstammt nicht aus einer Grundgesamtheit mit der in der Nullhypothese formulierten Aussage*. Kann dies bestätigt werden, ist eine Abweichung des Stichprobenkennwertes vom Erwartungswert μ der Stichprobenverteilung nicht auf Zufallseinflüsse zurückführbar, womit die Alternativhypothese angenommen werden muss. Die Nullhypothese ist dann *falsifiziert* (d.h. widerlegt) und kann nicht beibehalten werden.

Von den statistischen Hypothesen zu unterscheiden ist die inhaltliche Hypothese, die der Forscher über seinen Untersuchungsgegenstand aufgestellt hat. Diese ist in die statistischen Hypothesen H_0 und H_1 umzusetzen. Da Experimente meistens mit dem Ziel durchgeführt werden, Unterschiede zu belegen, ist die inhaltliche Hypothese häufig als die Alternativhypothese H_1 zu formulieren. Wenn aber Gleichheit vermutet wird, muss die inhaltliche Hypothese als die Nullhypothese H_0 formuliert werden.

Beispiel: Ein Forscher stellt die Hypothese auf, dass der Durchschnitt des IQ der Studierenden nicht dem mittleren IQ der Bevölkerung entspricht. Der IQ der Bevölkerung ist normalverteilt mit einem Erwartungswert von $\mu=100$ und einer Standardabweichung von $\sigma=15$. Da ein Unterschied postuliert wird, ist die inhaltliche Hypothese als die Alternativhypothese auszudrücken: $H_1: \mu \neq 100$, und die Nullhypothese ist entsprechend: $H_0: \mu = 100$. Die Nullhypothese besagt hier, dass die Stichprobe, die gezogen wird, aus einer Grundgesamtheit mit dem Erwartungswert $\mu = 100$ stammt, d.h. aus einer Population mit einem mittleren IQ von 100. Bei positivem Ausgang des Testes wird die Nullhypothese, dass die Grundgesamtheit, aus der die Stichprobe gezogen wurde, einen mittleren IQ von 100 hat, als bestätigt angenommen und damit die inhaltliche Hypothese verworfen. Die Alternativhypothese besagt in diesem Beispiel, dass die Stichprobe nicht aus einer Grundgesamtheit mit dem Erwartungswert $\mu = 100$ stammt, d.h. bei einem negativen Ergebnis des Testes wird angenommen, dass die Population, aus der die Stichprobe entnommen wurde, keinen mittleren IQ-Wert von 100 besitzt und damit die inhaltliche Hypothese bestätigt. Auch wenn der Forscher die gegenteilige Vermutung hegt, dass der Durchschnitt des IQ der Studierenden dem Durchschnitt des IQ der Bevölkerung entspricht, werden die gleichen statistischen Hypothesen aufgestellt. Allerdings steckt in diesem Fall die inhaltliche Hypothese in der Nullhypothese, da diese immer die Gleichheit, die Alternativhypothese immer die Ungleichheit vermutet.

2.1.2. Festlegung des Signifikanzniveaus

Bei der Entscheidung über die Gültigkeit der Hypothese können wir unterschiedlich strenge Maßstäbe anlegen. Das hängt davon ab, wie groß unsere Bereitschaft ist, eine *Fehlentscheidung* zu treffen. Sind wir bereit, das Risiko zu übernehmen, im Mittel bei 5 von 100 Stichproben ein Fehlurteil zu fällen, dann entscheiden wir uns für eine *Signifikanzniveau (Irrtumswahrscheinlichkeit) von $\alpha=0,05$ (5%)* bzw. eine *statistische Sicherheit (Sicherheitwahrscheinlichkeit) von $1-\alpha=0,95$ (95%)*. In diesem Fall genügen bereits relativ kleine Abweichungen des Stichprobenkennwertes vom Erwartungswert μ der Stichprobenverteilung, um die Nullhypothese zurückzuweisen. Aussagen, denen ein Signifikanzniveau von 5% zugrunde liegt, nennt man *signifikant*. Sie liegen auf dem „*5-Prozent-Niveau der Verlässlichkeit*“.

Wir können jedoch auch höhere Ansprüche an die Zuverlässigkeit unserer Entscheidung stellen. In diesem Fall werden wir die Hypothese nur zurückweisen oder beibehalten, wenn wir sicher sein können, mit dieser Entscheidung durchschnittlich in einem von 100 Fällen einen Fehler zu begehen. Dazu legen wir eine *Signifikanzniveau von $\alpha=0,01$ (1%)* bzw. *statistische Sicherheit von $1-\alpha=0,99$ (99%)* fest. Um eine Hypothese auf diesem Niveau zurückzuweisen, bedarf es bereits größerer Abweichungen des Stichprobenkennwertes vom Erwartungswert μ der Stichprobenverteilung. Aussagen, denen ein Signifikanzniveau von 1% zugrunde liegt, nennt man *sehr signifikant*. Sie liegen auf dem „*1-Prozent-Niveau der Verlässlichkeit*“.

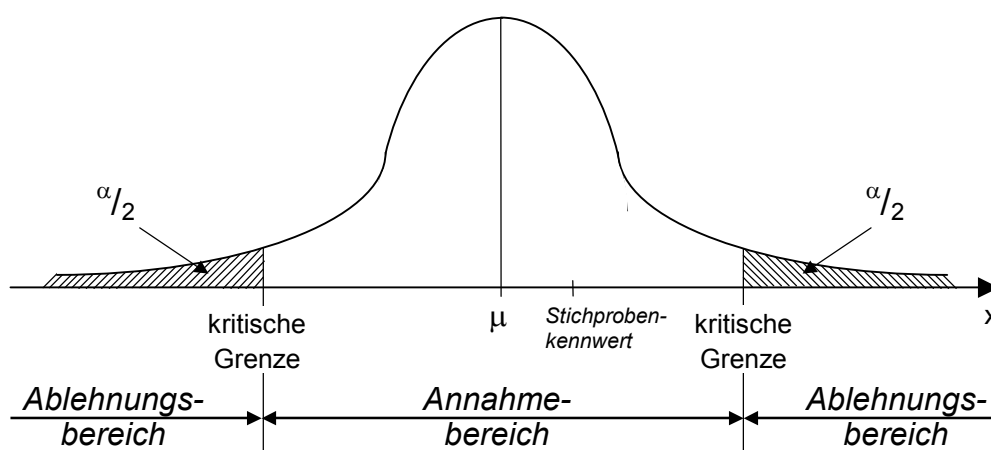
Schließlich gibt es Probleme, bei denen der Untersuchende eine Fehlentscheidung mit möglichst großer Sicherheit vermeiden will. In einem solchen Fall wird die Hypothese erst dann zurückgewiesen, wenn der Test mit einem *Signifikanzniveau* von

weniger als 1% fehlschlägt. Dies ist meistens $\alpha=0,001$ (0,1%). Hier gelingt es nur bei sehr großen Abweichungen, die Nullhypothese zurückzuweisen. Diese gelten dann als *hochsignifikant* oder als *äußerst signifikant*. Davon betroffene Aussagen gelten als auf dem *Signifikanzniveau von 1 Promille* bzw. auf dem „1-Promille-Niveau der *Verlässlichkeit*“ gesichert.

Beispiel: Für die Untersuchung des IQ der Studierenden setzen wir ein Signifikanzniveau von 1% fest.

2.1.3. Auswahl des Testverfahrens

Die Testverfahren funktionieren alle nach dem gleichen Prinzip: Zu einem statistischen Test gehören fest vorgegebene statistische Hypothesen, die eben durch dieses Verfahren geprüft werden. In Abhängigkeit davon, ob es sich um Hypothesen über z.B. bestimmte Durchschnitte, Häufigkeiten oder Korrelationen handelt, in Abhängigkeit von der Stichprobengröße und letztendlich natürlich auch in Abhängigkeit vom Skalenniveau der Testdaten wird jeweils eine bestimmte Stichprobenverteilung zugrunde gelegt: Dabei kann es sich um eine z-, t-, F-, Binomial- oder auch χ^2 -Verteilung handeln. In jedem Fall handelt es sich um eine Stichprobenverteilung (bzw. Wahrscheinlichkeitsverteilung von Stichprobenkennwerten), die es ermöglicht, mit Hilfe von Tabellenwerken die Wahrscheinlichkeit dafür zu bestimmen, dass ein Stichprobenkennwert in einem bestimmten Maße zufällig vom Parameter μ der Stichprobenverteilung abweicht. Umgekehrt – und was hier die ausschlaggebende Rolle spielt – lässt sich auch bestimmen, wie weit ein Stichprobenkennwert mindestens vom Erwartungswert μ entfernt liegen muss, um bei einer bestimmten Irrtumswahrscheinlichkeit α die Nullhypothese zu verwerfen. Es lassen sich also kritische Grenzen ermitteln, die ein Stichprobenkennwert über- bzw. unterschreiten muss, um die Nullhypothese ablehnen zu müssen. Entsprechend lässt sich der Bereich möglicher Stichprobenkennwerte in einen Annahmebereich und einen Ablehnungsbereich für die Nullhypothese aufteilen:



In obiger Abbildung sehen wir allgemein eine Stichprobenverteilung für die Hypothese, dass der Parameter μ einer Grundgesamtheit einen bestimmten Wert annimmt: Der Durchschnitt des IQ der Studierenden ist 100, der Anteil der Frauen an der Bevölkerung ist 50%, der Zusammenhang zwischen Einkommen und Schuhgröße ist gleich Null. Die Stichprobenverteilung zeigt an, dass die Wahrscheinlichkeit, dass

eine Stichprobe einen bestimmten Kennwert hat, mit steigender Entfernung vom Erwartungswert μ abnimmt. Die Stichprobenverteilung geht dabei rechts und links ins Unendliche, d.h. für jede zufällige Abweichung des Stichprobenkennwertes vom Erwartungswert μ gibt es eine zugehörige Wahrscheinlichkeit. Wann ist diese Abweichung nicht mehr zufällig? Auf diese Frage gibt es keine „wahre“ Antwort, die frei von Irrtümern ist, denn rein theoretisch gibt es für jede Abweichung eine Wahrscheinlichkeit, und sei sie noch so minimal. Um überhaupt eine Antwort zu finden, wird das Signifikanzniveau α gesetzt. Von der Stichprobenverteilung wird jeweils rechts und links die Hälfte des Signifikanzniveaus α „abgeschnitten“, d.h. es werden kritische Grenzen so gesetzt, dass diese vom Stichprobenkennwert in einer minimalen Anzahl von Fällen (5%, 1% bzw. 0,01%) links unter- bzw. rechts überschritten werden. Damit wird zwar allerdings der Irrtum zugelassen, dass die Nullhypothese auch dann abgelehnt wird, wenn die Stichprobe nur zufällig die kritischen Grenzen hinter sich lässt. Aber mit einer statistischen Sicherheit von 95%, 99% bzw. 99,9% wird diese Überschreitung signifikant, d.h. nicht zufällig sein. Die zur Verfügung stehenden statistischen Prüfverfahren beruhen alle auf Stichprobenverteilungen, die standardisiert sind: die z-Verteilung (Standardnormalverteilung), die t-Verteilung, F-Verteilung und χ^2 -Verteilung. Damit lässt sich der Stichprobenkennwert nicht sofort mit den kritischen Grenzen vergleichen, die sich zu einem Signifikanzniveau aus den zugehörigen Tabellen ablesen lassen, sondern muss zum Vergleich zuerst standardisiert werden. Der standardisierte Wert wird Prüfgröße genannt und die Berechnung der Prüfgröße ist ein wesentlicher Bestandteil eines Signifikanztests.

Zur Auswahl eines Verfahrens geht man folgendermaßen vor: Entsprechend der inhaltlichen Hypothese sucht man sich ein Verfahren aus, die eine bestimmte statistische Hypothese prüft. Die statistische Hypothese des Verfahrens muss zur inhaltlichen Hypothese passen. Dann muss geprüft werden, unter welchen Umständen die Anwendung des Verfahrens erlaubt ist. Dies umfasst:

- Skalenniveau der Messwerte der Stichprobe
- Stichprobengröße
- weitere verfahrensspezifische Voraussetzungen

In Haus- bzw. Klausuraufgaben wird an dieser Stelle die Prüfgröße zu dem ausgewählten Verfahren niedergeschrieben.

Beispiel: Für unsere Untersuchung verwenden wir ein Verfahren zum Vergleich eines Stichprobenmittelwertes mit einem Populationsparameter. Es ist bei HUBER (STAT -9-) unter Punkt 3.7 „Spezielle Signifikanztests“ angegeben. Zu vier möglichen statistischen Hypothesen sind acht verschiedene Verfahren angegeben, indem die jeweilige Berechnungsformel für die Prüfgröße angegeben ist. Das Verfahren, das wir brauchen, ist der erste Punkt der Aufzählung. Die allgemeine Hypothese $H_0: \mu = \mu_0$ entspricht unserer oben aufgestellten Hypothese $H_0: \mu = 100$, d.h. μ_0 steht stellvertretend für den vermuteten Wert 100 des Erwartungswertes μ . Die angegebene Prüfgröße beruht, je nachdem, ob die Standardabweichung σ der zu untersuchenden Grundgesamtheit (a) bekannt ist oder (b) nicht bekannt ist, auf der Standardnormalverteilung (Formel 1a) oder auf der t-Verteilung (Formel 1b). In unserem Beispiel ist die Standardabweichung mit $\sigma = 15$ bekannt, wir benutzen also die Formel (1a) zur Berechnung der Prüfgröße. (Voraussetzung für dieses Verfahren ist, dass der Stichprobenumfang mindestens 30 Probanden beträgt, wenn die zugrundeliegenden Messwerte nicht normalverteilt sind, da nach dem zentralen Grenzwertsatz Gesamtheiten von Stichprobendurchschnitten erst ab 30 Personen normalverteilt sind. Für Haus- oder Klausuraufgaben ist diese Unterscheidung irrelevant.)

2.1.4. Ziehung der Stichprobe und Ermittlung der Stichprobenkennwerte

Die Stichprobe wird erhoben, die Untersuchung durchgeführt, die Messwerte erfasst und die Daten mit Hilfe der deskriptiven Statistik ausgewertet, um den entsprechenden Stichprobenkennwert zu ermitteln. In Haus- und Klausuraufgaben werden an dieser Stelle die Werte, die zur Berechnung der Prüfgröße benötigt werden, bestimmt und niedergeschrieben.

Beispiel: Wir ziehen eine Zufallsstichprobe mit 36 Studenten aus allen Universitäten Deutschlands und verwenden einen wissenschaftlich anerkannten Intelligenztest zur Erfassung des Intelligenzquotienten jedes einzelnen Studenten. Damit sind die Voraussetzungen für das ausgewählte Verfahren erfüllt. Die statistische Auswertung der Messdaten ergeben folgende für die Berechnung der Prüfgröße erforderlichen Werte: $n=36$, $\bar{x}=106$, $\mu_0=100$ (der Populationsparameter unserer inhaltlichen Hypothese) und $\sigma=15$ (als bekannt vorausgesetzt).

2.1.5. Berechnung der Prüfgröße

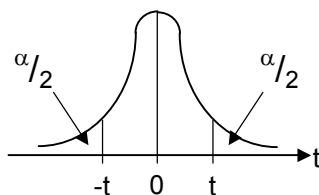
Die Prüfgröße wird nun berechnet, um dann vergleichen zu können, ob die kritischen Grenzen von ihr über- bzw. unterschritten werden.

Beispiel:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{106 - 100}{\frac{15}{\sqrt{36}}} = 2,4$$

2.1.6. Bestimmung der kritischen Grenze

In Abhängigkeit vom Signifikanzniveau α und der dem Verfahren zugrunde liegenden Stichprobenverteilung werden nun die kritischen Grenzen bestimmt. Sowohl für z- als auch für t-Verteilungen wird hierzu die Übersicht zu den t-Werten bei HUBER verwendet. Die t-Werte sind zeilenweise nach den verschiedenen Freiheitsgraden df der t-Verteilung und spaltenweise nach der Wahrscheinlichkeit (wie ϕ), dass solch ein t-Wert oder ein kleinerer auftritt, geordnet. Der Freiheitsgrad einer t-Verteilung berechnet sich mit $df=n-1$, also dem um eins verminderten Stichprobenumfang. Die z-Werte sind in dieser Übersicht ebenfalls angegeben, und zwar in der Zeile „ ∞ “, weil die t-Verteilung bei einem unendlich großen Stichprobenumfang der Standardnormalverteilung entspricht. Aber unter welcher Wahrscheinlichkeit müssen wir nachschauen für ein bestimmtes Signifikanzniveau? Betrachten wir dazu folgende Abbildung:



Die Gesamtfläche unter einer Verteilung beträgt 1, da sie die Gesamtwahrscheinlichkeit, dass irgendein Wert auftritt, angibt. Das Signifikanzniveau α teilt sich in zwei Hälften auf und schneidet die Verteilung auf beiden Seiten ab. Um die rechte, zur linken spiegelbildliche kritische Grenze zu erfahren, müssen wir also unter dem Wert $1-\alpha/2$ nachschauen.

Damit ergeben sich folgende Wahrscheinlichkeiten, unter der eine kritische Grenze gesucht werden muss zu den verschiedenen Signifikanzniveaus:

$$\begin{aligned} 5\%: & 1 - \frac{0,05}{2} = 0,975 \\ 1\%: & 1 - \frac{0,01}{2} = 0,995 \end{aligned}$$

Allgemein werden die kritischen Grenzen für die z- und die t-Verteilung mit folgender Symbolik angegeben: $z_{1-\alpha/2}$ bzw. $t_{1-\alpha/2;df}$.

Folgende kritische Grenzen gelten dann also bei einem Prüfverfahren, der die Standardnormalverteilung zugrunde liegt:

$$5\%: z_{1-\alpha/2} = z_{0,975} = 1,96$$

$$1\%: z_{1-\alpha/2} = z_{0,995} = 2,58$$

Bei den kritischen Grenzen eines Prüfverfahrens, der die t-Verteilung zugrunde liegt, ist zusätzlich der Freiheitsgrad df zu berücksichtigen. Bei einer Stichprobe von 20 Personen muss hier unter df=19 in der dem Signifikanzniveau zugehörigen Spalte nachgeschaut werden (ist df nicht aufgeführt, das nächstkleinere df wählen!):

$$5\%: t_{1-\alpha/2;df} = t_{0,975;19} = 2,09$$

$$1\%: t_{1-\alpha/2;df} = t_{0,995;19} = 2,86$$

Beispiel: In unserer Untersuchung wurde ein Signifikanzniveau von 1% festgesetzt. Dem Verfahren liegt die Standardnormalverteilung zugrunde, was an der Formel der Prüfgröße abgelesen werden kann, da hier ein z-Wert berechnet wird. Die zugehörige kritische Grenze lautet: $z_{0,995} = 2,58$.

2.1.7. Entscheidung über Annahme oder Ablehnung der Nullhypothese

Nachdem Prüfgröße und kritische Grenze ermittelt sind, kann entschieden werden, ob der Stichprobenkennwert im Annahmebereich oder im Ablehnungsbereich für die Nullhypothese liegt. Dazu muss man wissen, dass die gefundene kritische Grenze auf der rechten Seite der Stichprobenverteilung liegt und dass sich die linke Grenze dazu spiegelbildlich auf der linken Seite befindet, also nur noch mit einem negativen Vorzeichen versehen werden muss. Die Prüfgröße muss dann zur Annahme der Nullhypothese zwischen diesen beiden Grenzen liegen. Man kann es sich aber auch einfacher machen: Nimmt man den Betrag einer Prüfgröße, lässt also ein evtl. negatives Vorzeichen weg, wird eine negative Prüfgröße auf die rechte Seite gespiegelt, und man muss zur Annahme der Nullhypothese nur noch zusehen, ob die Prüfgröße kleiner als die positive kritische Grenze ist:

Standardnormalverteilung	t-Verteilung
$ z < z_{1-\alpha/2}$	$ t < t_{1-\alpha/2;df}$

Im anderen Fall muss die Nullhypothese verworfen werden, wodurch die Alternativhypothese angenommen wird.

Im letzten Schritt wird das Ergebnis auf die der gesamten Untersuchung zugrunde liegenden inhaltlichen Hypothese verallgemeinert. Je nachdem, ob sie entsprechend ihrer Aussage als H_0 - oder als H_1 -Hypothese formuliert wurde, ist sie aufgrund der Annahme bzw. Ablehnung der Nullhypothese nun *verifiziert* oder *falsifiziert*.

Beispiel: Als Prüfgröße hatten wir oben einen Wert von 2,4 berechnet und als kritische Grenze einen Wert von 2,58 ermittelt. Die Prüfung ergibt: $|2,4| < 2,58$. Damit ist die Bedingung zur Annahme der

Nullhypothese gegeben. Unsere inhaltliche Hypothese steckte allerdings in der Alternativhypothese, womit mit der Ablehnung der Alternativhypothese auch unsere inhaltliche verworfen ist: Studenten haben keinen von dem der Bevölkerung verschiedenen IQ.

2.2. Fehlerarten bei statistischen Entscheidungen

Nachdem die Nullhypothese und die Alternativhypothese formuliert sind, kann die Untersuchung, auf Grund derer die Tragfähigkeit der inhaltlichen Hypothese ermittelt werden soll, durchgeführt werden. Die Entscheidung hierüber ist jedoch erschwert, da sich das Ergebnis der Untersuchung nur auf die Stichprobe bezieht, die in der Untersuchung einbezogen wurde, während die inhaltliche Hypothese die Verhältnisse in der Population beschreibt.

Die inferenzstatistische Hypothesenprüfung bezieht sich auf Hypothesen, die für eine Population, der die untersuchte Stichprobe entnommen ist, gültig sein sollen.

Damit ist nicht auszuschließen, dass das Ergebnis der Untersuchung auf Grund der Stichprobenauswahl zufällig die Alternativhypothese bestätigt, wenngleich „in Wahrheit“, d.h. bezogen auf die gesamte Population, die Nullhypothese zutrifft. Umgekehrt können stichprobenspezifische Zufälle für die Annahme der Nullhypothese sprechen, während in der Population die Alternativhypothese richtig ist.

Die Entscheidung, vor der der Forscher steht, lässt sich schematisch wie in folgender Tabelle darstellen:

		In der Population gilt die	
		H_0	H_1
Entscheidung	H_0	richtige Entscheidung	<i>β-Fehler</i>
zugunsten der	H_1	<i>α-Fehler</i>	richtige Entscheidung

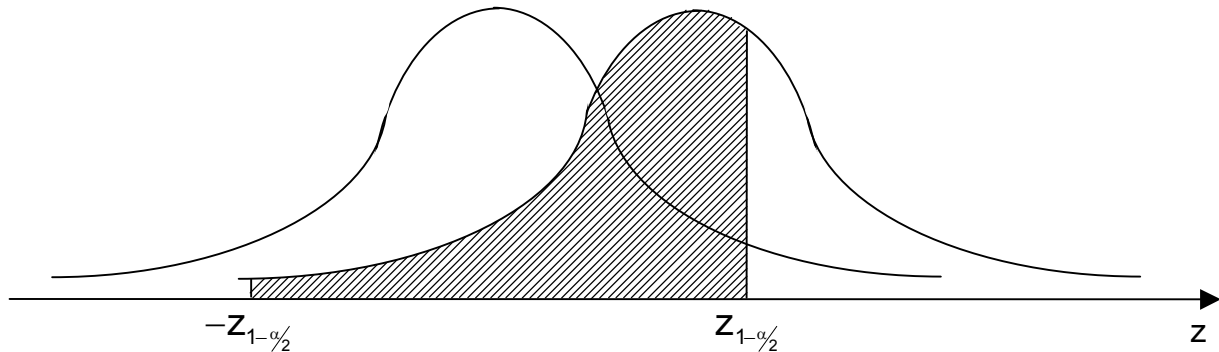
Neben den beiden richtigen Entscheidungen, bei denen auf Grund der Stichprobenergebnisse die Populationsverhältnisse korrekt erschlossen werden, können zwei fehlerhafte Entscheidungen getroffen werden. Entweder kann eine an sich richtige Nullhypothese auf Grund der Stichprobenergebnisse zugunsten der Alternativhypothese verworfen werden (*α -Fehler* oder *Fehler erster Art*), oder es wird die Nullhypothese akzeptiert, obwohl die Alternativhypothese richtig ist (*β -Fehler* oder *Fehler zweiter Art*).

In der Prüfstatistik bezeichnet man eine fälschliche Entscheidung zugunsten von H_1 als α -Fehler (Fehler 1. Art) und eine fälschliche Entscheidung zugunsten von H_0 als β -Fehler (Fehler 2. Art).

Der α -Fehler ist durch das Setzen des Signifikanzniveaus α bestimmt. Der Forscher geht bewusst das Risiko ein, dass er in 1% bzw. 5% der Fälle, in der er eine Stich-

probe zieht, eine falsche Entscheidung treffen wird, nämlich die an sich richtige Nullhypothese zu verwerfen.

Der β -Fehler lässt sich grafisch folgendermaßen veranschaulichen:



Links sehen wir die Normalverteilungskurve der Nullhypothese. Diese wird angenommen, wenn die Prüfgröße innerhalb der zugehörigen und eingezeichneten kritischen Grenzen liegt. Im Falle des β -Fehlers ist die Nullhypothese aber falsch, die Alternativhypothese ist richtig, die zugehörige Normalverteilungskurve ist rechts dargestellt. Diese ist maßgeblich für die Wahrscheinlichkeit des Zustandekommens einer Prüfgröße. Die Wahrscheinlichkeit, dass eine Prüfgröße zwischen die vom Forscher aufgestellten kritischen Grenzen fällt, und damit seine falsche Hypothese bestätigt wird, ist durch die schraffierte Fläche dargelegt. Die Größe des β -Fehler lässt sich nur berechnen, wenn sowohl die falsche als auch die wahre Hypothese bekannt ist und dies ist nur in konstruierten Beispielen der Fall. Allerdings kann man aufgrund obiger Abbildung folgende Einflussgrößen des β -Fehler aufstellen:

- Unterschied zwischen hypothetischem und tatsächlichem Mittelwert:
Je mehr die beiden Kurven in obiger Abbildung auseinander liegen, desto kleiner ist die schraffierte Fläche: je größer der Unterschied, desto kleiner β .
- Fehlerwahrscheinlichkeit α :
Je mehr die z-Grenzen beieinander liegen, desto kleiner ist die dargestellte Fläche: je größer α , desto kleiner β .

Die folgenden zwei Faktoren beeinflussen die Streuung einer Stichprobenverteilung: $\sigma_{\bar{x}} = \sigma_x / \sqrt{n}$. Je kleiner die Streuung $\sigma_{\bar{x}}$, desto schmaler die Kurven, desto kleiner die schraffierte Fläche.

- Stichprobenumfang n: je größer n, desto kleiner β
- der Streuung σ der Messwerte: je kleiner σ , desto kleiner β
- (verwendetes Testverfahren)

In jeder statistischen Prüfung ist darüber zu entscheiden, ob man lieber einen Fehler erster Art oder zweiter Art in Kauf nehmen möchte. Dabei wird man sich an den Konsequenzen orientieren, die mögliche Fehlentscheidungen haben können: Die Freigabe eines toxischen, aber statistisch als nicht toxisch „erwiesenen“ Medikamentes oder die Einführung eines neuen Lehrplanes an allen Schulen Deutschlands aufgrund einer nicht wirksamen, statistisch aber als wirksam bestätigte Lehrmethode

oder die Anerkennung einer unwirksamen, aber statistisch als wirksam anerkannte Therapiemethode von den Krankenkassen sind mögliche Konsequenzen einer Fehlentscheidung, wo sich jeder Forscher klar werden muss, was eine Fehlentscheidung in seinem Fall bedeutet. Die folgenden Alternativen verdienen daher Beachtung:

- Ist es sehr folgenschwer, eine bestimmte Nullhypothese fälschlicherweise abzulehnen, werden wir dieses Risiko möglichst klein halten, also ein hohes Signifikanzniveau wählen.
- Ist es dagegen folgenschwer, eine Nullhypothese fälschlich anzunehmen, empfiehlt es sich, mit geringerer Verlässlichkeit zu prüfen.

2.3. Zur Bedeutung der Bestätigung oder Ablehnung einer Hypothese

Bei weitem der größte Teil aller Experimente und empirischen Untersuchungen in den Sozialwissenschaften ist auf die Zurückweisung der Nullhypothese gerichtet bzw. – positiv ausgedrückt – auf den Nachweis verlässlicher (signifikanter) Unterschiede. Dabei bezieht sich die statistisch prüfbare Nullhypothese stets auf die Parameter einer Grundgesamtheit, wobei zur Prüfung jedoch Stichprobenbefunde verwendet werden. Die Nullhypothese kann daher durch Stichprobenergebnisse günstigenfalls widerlegt, keinesfalls jedoch „bewiesen“ werden. Vielmehr bleibt die Nullhypothese – auch wenn sie durch empirische Befunde nicht zurückgewiesen werden kann – nur eine *haltbare* Hypothese, vielleicht eine von vielen und einander widersprechenden haltbaren Hypothesen. Über die *Richtigkeit* von Hypothesen ist damit noch nichts ausgesagt.

„Eine Hypothese anzunehmen heißt lediglich, diese Hypothese der anderen vorzuziehen. Es bedeutet nicht, dass wir diese Hypothese für unbedingt richtig halten. ... Wenn eine Hypothese abgelehnt wird, so heißt dies nur, dass nach der vereinbarten Vorschrift die andere Hypothese vorzuziehen ist, und schließt nicht die Aussage ein, dass die Hypothese falsch ist.“ (E. WEBER, 1972, S. 172)

Beispiel eines Signifikanztests: Vergleich eines Stichprobenmittelwertes mit dem Populationserwartungswert μ (mit σ bekannt)

Angaben: Der Durchschnitt des IQ der Studenten soll nicht identisch mit dem Durchschnitt des IQ der Bevölkerung sein. Der IQ der Bevölkerung ist $N(100;15)$ -verteilt. Es wird eine Stichprobe von 36 Personen erhoben. Der Durchschnitt des IQ dieser Stichprobe beträgt 106. Es soll auf dem 1%-Niveau auf Signifikanz geprüft werden.

1. *Statistische Hypothesen:* $H_0: \mu = 100$ $H_1: \mu \neq 100$

2. *Festlegung des Signifikanzniveaus:* $\alpha = 0,01$

3. *Auswahl des Verfahrens:*
$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

4. *Stichprobenwerte:* $\bar{x} = 106$
 $\mu_0 = 100$

Stoff

$$\begin{aligned}\sigma &= 15 \\ n &= 36\end{aligned}$$

5. Berechnung der Prüfgröße:

$$z = \frac{106 - 100}{\frac{15}{\sqrt{36}}} = 2,4$$

6. Ermittlung der kritischen Grenze:

$$z_{1-\frac{\alpha}{2}} = z_{1-0,05/2} = z_{0,975} = 2,58$$

7. Entscheidung:

$$|2,4| < 2,58$$

⇒ Nullhypothese wird beibehalten

Damit ist auf dem Signifikanzniveau von 1% die Hypothese verworfen, dass Studenten einen anderen IQ besitzen als die Normalbevölkerung.

Beispiel eines Signifikanztests: Vergleich eines Stichprobenmittelwertes mit dem Populationserwartungswert μ (mit σ unbekannt)

Angaben: Der Durchschnitt der Schuhgröße der Studenten soll identisch sein mit dem Durchschnitt der Schuhgröße der Bevölkerung. Die mittlere Schuhgröße der Bevölkerung beträgt nach Angaben des deutschen Verbandes der Schuhhersteller Größe 42. Die Streuung der Schuhgrößen in der Bevölkerung sei aber leider nicht bekannt. Es wird eine Stichprobe von 35 Personen erhoben. Als Durchschnitt der Schuhgröße dieser Stichprobe wurde Größe 40,7, als Standardabweichung die Größe 2,7 ermittelt. Es soll auf dem 5%-Niveau auf Signifikanz geprüft werden.

1. Statistische Hypothesen:

$$H_0: \mu = 42 \quad H_1: \mu \neq 42$$

2. Festlegung des Signifikanzniveaus:

$$\alpha = 0,05$$

3. Auswahl des Verfahrens:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}; \quad df = n-1$$

4. Stichprobenwerte:

$$\begin{aligned}\bar{x} &= 40,7 \\ \mu_0 &= 42 \\ s &= 2,7 \\ n &= 35\end{aligned}$$

5. Berechnung der Prüfgröße:

$$t = \frac{40,7 - 42}{\frac{2,7}{\sqrt{35}}} = -2,85$$

6. Ermittlung der kritischen Grenze:

$$t_{1-\frac{\alpha}{2};df} = t_{1-0,05/2;35-1} = t_{0,975;34} = 2,03$$

7. Entscheidung:

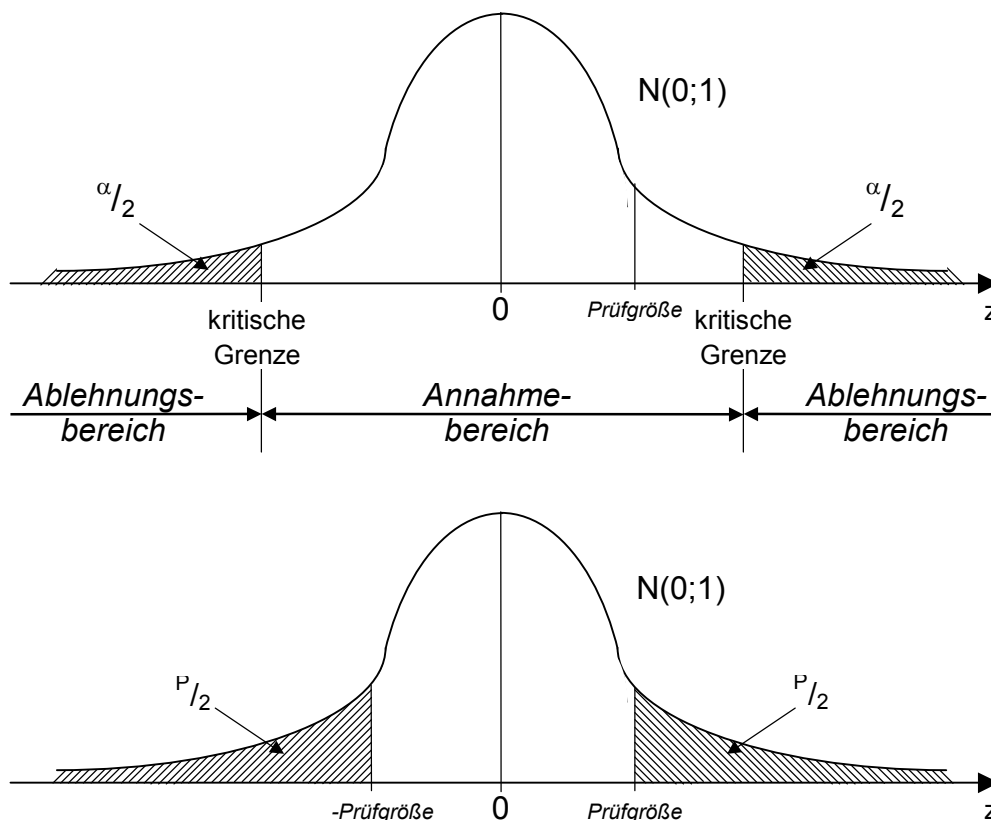
$$|-2,85| \geq 2,03$$

⇒ Nullhypothese wird abgelehnt

Damit ist mit einer Irrtumswahrscheinlichkeit von 5% die Hypothese widerlegt, dass Studenten die gleiche Schuhgröße wie die Normalbevölkerung besitzen.

2.4. Variante des Tests: Die Überschreitungswahrscheinlichkeit P

Bei der Durchführung eines Signifikanztests haben wir das Problem, dass ein Signifikanzniveau gesetzt wird, welches eine Wahrscheinlichkeit darstellt, und eine Prüfgröße berechnet wird, die ein z- oder t-Wert ist. Beides lässt sich nicht ohne weiteres vergleichen. Da die z- bzw. t-Werte und die Wahrscheinlichkeiten zu ihnen eine eindeutige Beziehung zueinander haben, die sich in den Tabellen äußert, wird dieses Problem dadurch gelöst, dass zum Signifikanzniveau die entsprechende kritische Grenze ermittelt wird, die sich dann mit der Prüfgröße vergleichen lässt. Man kann auch umgekehrt vorgehen, indem man zur Prüfgröße eine Wahrscheinlichkeit bestimmt, die Überschreitungswahrscheinlichkeit P, und diese dann mit dem Signifikanzniveau, der Irrtumswahrscheinlichkeit α , vergleicht. Die Überschreitungswahrscheinlichkeit ist die Wahrscheinlichkeit, dass der Stichprobenkennwert einer Zufallsstichprobe mehr vom Erwartungswert μ als der im Experiment ermittelte Stichprobenkennwert abweicht. Folgende Abbildungen verdeutlichen diesen Sachverhalt an der Standardnormalverteilung:



In der oberen Abbildung ist das bisherige Prüfverfahren mit der Irrtumswahrscheinlichkeit α , in der unteren die neue Variante mit der Überschreitungswahrscheinlichkeit P dargestellt. Bisher hatten wir zu der Irrtumswahrscheinlichkeit α die entsprechenden kritischen Grenzen ermittelt und mit der Prüfgröße verglichen. In der neuen Variante ermitteln wir zur Prüfgröße die Überschreitungswahrscheinlichkeit P und vergleichen diese mit der Irrtumswahrscheinlichkeit α . Wie die obige Abbildung erkennen lässt, liegt die Prüfgröße innerhalb der kritischen Grenzen, wodurch die Nullhypothese bestätigt wird. Entsprechend können wir der unteren Abbildung entnehmen, dass die Nullhypothese dann bestätigt wird, wenn die Überschreitungswahrscheinlichkeit P größer als das Signifikanzniveau α wird.

Aber wie wird die Überschreitungswahrscheinlichkeit P berechnet? Die Überschreitungswahrscheinlichkeit P besteht, wie es in der unteren Abbildung deutlich wird, aus zwei gleich großen, zur Null spiegelbildlichen Flächen. Die rechte Fläche lässt sich berechnen, indem wir von der Gesamtwahrscheinlichkeit 1 den ϕ -Wert des Absolutbetrags der Prüfgröße abziehen, also die Fläche unter der Kurve links von dem positiven Prüfwert, der sich in der Abbildung rechts befindet. Nun brauchen wir den so erhaltenen Wert nur noch zu verdoppeln, da die linke Hälfte der Überschreitungswahrscheinlichkeit P genauso groß ist wie die rechte Hälfte:

$$P = 2 \cdot (1 - \phi(|z|))$$

Der ϕ -Wert zum positiven z-Wert einer Prüfgröße lässt sich bei HUBER in der Tabelle zur Standardnormalverteilung ablesen.

Beispiel: In der Untersuchung über den IQ der Studenten wird die Überschreitungswahrscheinlichkeit P wie folgt berechnet: $P = 2 \cdot (1 - \phi(|2,4|)) = 2 \cdot (1 - 0,992) = 0,016$. Der Vergleich mit dem Signifikanzniveau von $\alpha=0,01$ ergibt, dass die Überschreitungswahrscheinlichkeit P größer ist: $0,016 > 0,01$. Damit ist die Nullhypothese bestätigt.

Analog könnte man auch zur t-Verteilung eine Überschreitungswahrscheinlichkeit P für eine t-Prüfgröße berechnen. Allerdings ist das unüblich und die geläufigen Tabellenwerke lassen oft das Ablesen der Wahrscheinlichkeit zu einem t-Wert nur eingeschränkt zu (auf die Werte für die Signifikanzniveaus eingeschränkt, da die Tabellen nur zum Ablesen der kritischen Grenzen dienen).

Die Überschreitungswahrscheinlichkeit P hat den Vorteil, dass man direkt erkennen kann, auf welchem Niveau der Verlässlichkeit eine Aussage signifikant ist. Oft wird in der Forschung kein Signifikanzniveau zugrunde gelegt, sondern es wird die Überschreitungswahrscheinlichkeit P ausgerechnet, um zu sehen, auf welchem Signifikanzniveau eine Hypothese bestätigt wird. So hat es sich in Fachzeitschriften und Forschungsberichten eingebürgert, bei der Darstellung empirischer Ergebnisse die drei Signifikanzniveaus durch Abkürzungen und Symbole wie folgt zu kennzeichnen:

Überschreitungswahrscheinlichkeit	> 0,05	≤ 0,05	≤ 0,01	≤ 0,001
Bedeutung	nicht signifikant	signifikant	sehr signifikant	hochsignifikant
Abkürzung	n.s.	s.	s.s.	h.s.
Symbol		*	**	***

Auch bei der EDV-gestützten Auswertung werden die Ergebnisse oft durch Angabe der Überschreitungswahrscheinlichkeit P vom Rechner ausgegeben.

**Beispiel zur Variante des Signifikanztests:
Vergleich eines Stichprobenmittelwertes mit dem Populationsparameter μ mit Hilfe der Überschreitungswahrscheinlichkeit P**

Das folgende Beispiel zum IQ der Studenten ist bis auf Punkt 6 und 7 mit der bisherigen Vorgehensweise identisch (s. Musterbeispiel auf S. 19).

1. *Statistische Hypothesen:*

$$H_0: \mu = 100 \quad H_1: \mu \neq 100$$

2. Festlegung des Signifikanzniveaus: $\alpha = 0,01$
3. Auswahl des Verfahrens: $z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$
4. Stichprobenwerte:
 $\bar{x} = 106$
 $\mu_0 = 100$
 $\sigma = 15$
 $n = 36$
5. Berechnung der Prüfgröße: $z = \frac{106 - 100}{\frac{15}{\sqrt{36}}} = 2,4$
6. Überschreitungswahrscheinlichkeit P : $P = 2 \cdot (1 - \Phi(|2,4|)) = 0,016$
7. Entscheidung: $0,016 > 0,01$
 \Rightarrow Nullhypothese wird beibehalten

Damit ist mit einer Sicherheitswahrscheinlichkeit von 99% die Hypothese verworfen, dass Studenten einen anderen IQ als die Normalbevölkerung besitzen.

2.5. Einseitige Signifikanztests

Mit der bisher erläuterten Vorgehensweise zur Prüfung der Signifikanz eines Stichprobenergebnisses können wir nur unspezifische bzw. ungerichtete Hypothesen testen. Das sind Hypothesen, die nur die Gleichheit im Falle der Nullhypothese bzw. die Ungleichheit im Falle der Alternativhypothese postulieren: Studenten haben einen Durchschnitts-IQ von 100 oder nicht, der Anteil der Frauen an hoch dotierten Posten beträgt 50% oder nicht etc. Um ungerichtete Hypothesen zu prüfen, setzen wir – wie besprochen – ein Signifikanzniveau, das auf beiden Seiten einer Stichprobenverteilung liegt, wodurch auf beiden Seiten kritische Grenzen bestimmt sind, deren Unter- bzw. Überschreitung durch die Prüfgröße die Nullhypothese verwerfen, also die Gleichheit ablehnen und die Ungleichheit bestätigen lässt. Umgekehrt lässt die Nichtüberschreitung der kritischen Grenzen durch die Prüfgröße die Nullhypothese beibehalten, also die Gleichheit annehmen und die Ungleichheit verwerfen. Da das Signifikanzniveau auf beiden Seiten der Stichprobenverteilung liegt, nennt man ein solches Verfahren auch einen zweiseitigen Signifikanztest.

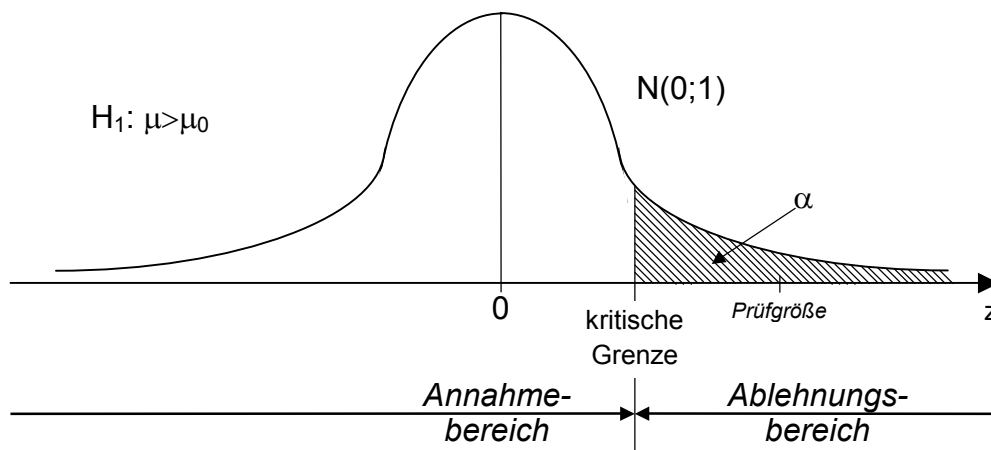
Oft möchte der Forscher aber spezifische bzw. gerichtete Hypothesen prüfen: Ob eine Therapiemethode besser ist als eine andere, ob sich der Anteil der Singles an der deutschen Bevölkerung in Großstädten erhöht hat gegenüber der letzten 10 Jahre oder ob sich der IQ einer Person ab 20 Jahren mit seinem Lebensalter stetig verringert (das konnte übrigens in dieser Allgemeinheit widerlegt werden). Gerichtete Hypothesen fordern die signifikante Abweichung eines Stichprobenkennwertes vom Erwartungswert μ der Stichprobenverteilung in eine bestimmte Richtung, nämlich entweder nach oben oder nach unten, in Richtung höher oder niedriger, als Verbesserung oder Verschlechterung. Es handelt sich also um eine Spezifizierung bzw. Richtungsgebung der Alternativhypothese H_1 : Soll der Stichprobenkennwert die kriti-

sche Grenze auf der linken oder auf der rechten Seite der Stichprobenverteilung überschreiten? Für diesen Zweck wird die Alternativhypothese (und nur die Alternativhypothese!) nun entsprechend der gerichteten inhaltlichen Hypothese spezifisch umformuliert. In dem uns bekannten Beispiel, dass der durchschnittliche IQ der Studenten höher sei als der mittlere Wert der Bevölkerung, schreibt man $H_1: \mu > 100$ (rechtsseitige Fragestellung), und für den Fall, dass der mittlere IQ unter dem durchschnittlichen Wert der Bevölkerung liegen soll, schreibt man $H_1: \mu < 100$ (linksseitige Fragestellung). Die Nullhypothese behält ihr Postulat der Gleichheit und verändert bei manchen Autoren ihre Schreibweise nicht, andere Autoren und auch Herr HUBER berücksichtigen in der Nullhypothese die gegenteilige Richtung der Alternativhypothese. Allgemein gibt man für inhaltliche, gerichtete Hypothesen, die sich auf Durchschnitte von metrischen Messgrößen beziehen, die statistischen Hypothesen in folgender Form wieder:

Fragestellung	Nullhypothese	Alternativhypothese	Bedeutung der H_1 -Hypothese
rechtsseitig	$H_0: \mu \leq \mu_0$	$H_1: \mu > \mu_0$	Stichprobenkennwert weicht signifikant nach oben ab
linksseitig	$H_0: \mu \geq \mu_0$	$H_1: \mu < \mu_0$	Stichprobenkennwert weicht signifikant nach unten ab

Dementsprechend braucht man nur noch auf einer Seite das Signifikanzniveau anzulegen und man unterscheidet links- und rechtsseitige Signifikanztests. Allgemein spricht man von einseitigen Tests.

2.5.1. Rechtsseitiger Test



Beim rechtsseitigen Test postuliert die Alternativhypothese die signifikante Abweichung des Stichprobenkennwertes nach oben. Entsprechend befindet sich das Signifikanzniveau α in seiner Gesamtheit auf der rechten Seite und nicht nur zur Hälfte. Damit verändert sich auch die Berechnung der Wahrscheinlichkeit, unter der die kritische Grenze in der t-Wert-Tabelle im Anhang des Skriptes von Herrn HUBER gesucht wird. Diesmal erhält man die Wahrscheinlichkeit, indem von der Gesamtwahrscheinlichkeit 1 den ganzen Wert des Signifikanzniveaus abzieht: $1-\alpha$. Dieser Wert entspricht der Sicherheitswahrscheinlichkeit, beim 5%-Niveau also 0,95 und beim 1%-Niveau 0,99.

Im Falle einer z-Prüfgröße erhalten wir für einseitige Tests also zwei neue kritische z-Grenzen, die in der Zeile „∞“ der t-Wert-Tabelle abgelesen werden können:

$$\begin{aligned} 5\%: z_{1-\alpha} &= z_{1-0,05} = z_{0,95} = 1,64 \\ 1\%: z_{1-\alpha} &= z_{1-0,01} = z_{0,99} = 2,33 \end{aligned}$$

Im Falle einer t-Prüfgröße müssen wir dementsprechend bei einseitigen Tests unter der Spalte „0,95“ im Falle eines 5%-Niveaus und unter der Spalte „0,99“ im Falle eines 1%-Niveaus nachschauen. Auch hier müssen wir zusätzlich den Freiheitsgrad $df=n-1$ berücksichtigen, indem wir in der entsprechenden Zeile die kritische t-Grenze nachschauen. Die kritischen Grenzen im Fall eines Stichprobenumfangs von $n=20$ lauten:

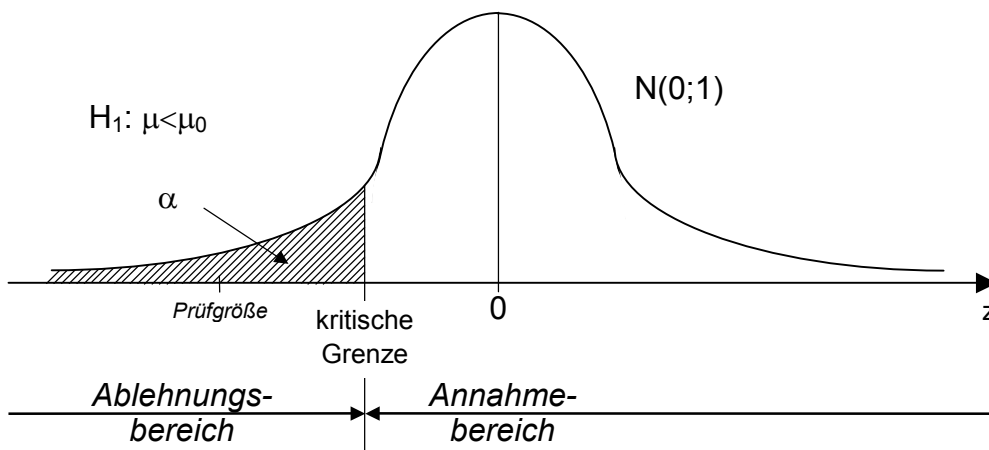
$$\begin{aligned} 5\%: t_{1-\alpha;df} &= t_{1-0,05;20-1} = t_{0,95;19} = 1,73 \\ 1\%: t_{1-\alpha;df} &= t_{1-0,01;20-1} = t_{0,99;19} = 2,54 \end{aligned}$$

Vergleicht man die kritischen Grenzen vom einseitigen Test mit denen vom zweiseitigen, so fällt auf, dass die vom einseitigen Test stets kleiner sind. Das verwundert kaum, da doch das Signifikanzniveau nicht mehr halbiert wird, sondern in seiner Gesamtheit die rechte Seite der Stichprobenverteilung abschneidet. Daraus folgt, dass die kritischen Grenzen beim einseitigen Test näher zur Innenseite der Stichprobenverteilung liegen, also kleiner sind als die kritischen Grenzen beim zweiseitigen Test. Dies hat zur Konsequenz, dass die Wahrscheinlichkeit, dass die H_0 -Hypothese abgelehnt wird, beim einseitigen Test größer ist als beim zweiseitigen Test. Beim einseitigen Test wird eine H_0 -Hypothese eher verworfen als beim zweiseitigen Test.

Zur Annahme der Nullhypothese muss die Prüfgröße kleiner sein als die kritische Grenze, im anderen Falle ist die Alternativhypothese bestätigt:

Bedingung zur Annahme der Nullhypothese bei einem rechtsseitigen Test	
Standardnormalverteilung	t-Verteilung
$z < z_{1-\alpha}$	$t < t_{1-\alpha;df}$

2.5.2. Linksseitiger Test



2.5.4. Vereinfachung der Annahmebedingung für die H_0 -Hypothese

Die Durchführung eines einseitigen Prüfverfahrens mutet etwas kompliziert an, wenn sie über die Variante mit den kritischen Grenzen berechnet wird: die kritischen Grenzen sind mal positiv, mal negativ zu nehmen, und beim Vergleich muss die Prüfgröße mal größer, mal kleiner als die kritische Grenze sein. Daher seien ein paar Regeln genannt, deren Beachtung die Anwendung eines einseitigen Verfahrens erleichtern:

- Ist in der inhaltlichen Hypothese die Rede von „größer“, „besser“, „höher“, „weiter“ etc., so lautet die Alternativhypothese $H_1: \mu > \mu_0$. Ist hingegen die Rede von „kleiner“, „schlechter“, „niedriger“, „näher“ etc. so lautet sie: $H_1: \mu < \mu_0$. *Die Richtung der inhaltlichen Hypothese schlägt sich also im Vergleichszeichen der Alternativhypothese nieder.*
- Enthält dann die Alternativhypothese ein „>“-Zeichen, so ist die einseitige kritische Grenze $(1-\alpha!)$ positiv, enthält sie ein „<“-Zeichen, so ist sie negativ. *Die Richtung des Vergleichszeichens in der H_1 verweist also auf das Vorzeichen.*
- Der Vergleich des Prüfwertes mit der kritischen Grenze muss zur Annahme der H_0 -Hypothese dasselbe Vergleichszeichen ergeben, wie sie in ihr formuliert wurde. Ergibt der Vergleich das Vergleichszeichen der H_1 -Hypothese, so ist diese angenommen. *Es wird die Hypothese beibehalten, deren Vergleichszeichen bestätigt wird.*

In manchen Lehrbüchern wird vorgeschlagen, die Annahmebedingung beim einseitigen Test genauso durchzuführen wie beim zweiseitigen Test:

Annahmebedingung für die Nullhypothese bei ein- und zweiseitigen Tests
Prüfgröße < positive kritische Grenze

Das scheint auf den ersten Blick zu funktionieren, da durch die Verwendung des Absolutbetrags der Prüfgröße und der positiven kritischen Grenze ein möglicher linksseitiger Test gespiegelt, also zum rechtsseitigen Test transformiert wird und die Annahmebedingung für den rechtsseitigen Test verwendet wird. Der Haken an der Sache ist, dass in zwei von acht möglichen Vergleichssituation diese Art der Prüfung fehlschlägt:

1. Ist der Test rechtsseitig ($H_1: \mu > \mu_0$) und damit die kritische Grenze positiv, die Prüfgröße aber negativ, kann es bei genügend großem Absolutbetrag der Prüfgröße passieren, dass die Nullhypothese fälschlicherweise abgelehnt wird.

Beispiel: Die Prüfgröße lautet $z = -2,9$ und die kritische Grenze $z_{1-\alpha} = 2,54$. Da $-2,9$ kleiner als $2,54$ ist, liegt $-2,9$ im Annahmebereich für die Nullhypothese. Die Prüfung nach obigem Vorschlag ergibt aber: $|-2,9| > 2,54$, womit die Nullhypothese fälschlicherweise abgelehnt wird.

2. Das gleiche gilt für den umgekehrten Fall: Ist der Test linksseitig ($H_1: \mu < \mu_0$) und damit die kritische Grenze negativ, die Prüfgröße aber positiv, kann es bei einer genügend großen Prüfgröße ebenfalls passieren, dass die Nullhypothese fälschlicherweise abgelehnt wird.

Beispiel: Die Prüfgröße lautet $z = 2,9$ und die kritische Grenze $z_{1-\alpha} = -2,54$. Da $2,9$ größer als $-2,54$ ist, liegt $2,9$ im Annahmebereich für die Nullhypothese. Die Prüfung nach obigem Vorschlag ergibt aber: $|2,9| > 2,54$, womit die Nullhypothese fälschlicherweise abgelehnt wird.

Dieser Vorschlag wird häufig damit begründet, dass die übliche Prüfung mit Berücksichtigung der Vor- und Vergleichszeichen zu einer etwas komplizierten und für den Anfänger verwirrenden Formulierung der Annahmebedingungen führt. Dabei wird aber nicht auf die beiden Möglichkeiten einer Fehlentscheidung hingewiesen. Es empfiehlt sich, die einseitigen Tests mit den Regeln durchzuführen, die ich angegeben habe und ja doch recht einfach zu merken sind. Im Falle eines „Blackouts“ in der Klausur mag man sich daran erinnern, dass die Prüfung in 75% der möglichen Fälle mit der Annahmebedingung für den zweiseitigen Test klappt. Am besten kommt man weg, wenn die Variante „Überschreitungswahrscheinlichkeit“ benutzt wird und in der Klausur- oder Hausaufgabe die andere Variante nicht verlangt wird.

2.6. Signifikanztests bei Häufigkeiten (Approximativer Binomialtest)

Bei dichotomen Daten kann die signifikante Abweichung der Häufigkeit p einer Ausprägung des untersuchten Merkmals einer Stichprobe von einem Populationsparameter π geprüft werden. Es wird das Verfahren (2) (HUBER, STAT -9-) angewendet. Dem Verfahren zugrunde liegt die Standardnormalverteilung, wie sie im Kapitel »Stichprobenverteilungen von Häufigkeiten« auf S. 9 vorgestellt wurde. Für diese Standardnormalverteilung gilt:

$$\mu = \pi \quad \sigma = \sqrt{\frac{\pi(1-\pi)}{n}}$$

Wenn man diese Berechnungsformeln für die Stichprobenverteilungsparameter μ und σ , statt dem Stichprobendurchschnitt \bar{x} die Stichprobenhäufigkeit p der Ausprägung des gemessenen Merkmals und anstelle μ_0 das Symbol π_0 in die z-Formel zur Berechnung der Prüfgrößen von Durchschnitten (Formel (1a), HUBER, STAT -9-) einsetzt, so ergibt sich:

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 \cdot (1 - \pi_0)}{n}}}$$

Diese Prüfgröße ist genauso handzuhaben wie die bisher besprochenen Prüfgrößen. Allerdings ist vor der Anwendung die Voraussetzung zu prüfen, dass die Stichprobenverteilung von Stichprobenhäufigkeiten in Abhängigkeit vom Stichprobenumfang n durch eine Standardnormalverteilung repräsentiert werden darf (s. S. 10):

$$n > \frac{9}{\pi \cdot (1 - \pi)}$$

Beispiel eines einseitigen Signifikanztests: Vergleich der Ausprägungshäufigkeit p einer Stichprobe mit dem Populationsparameter π (Approximativer Binomialtest)

Angaben: Die inhaltliche Hypothese lautet, dass in einer bestimmten Berufsgruppe Frauen unterrepräsentiert (< 50%) seien. Dies soll auf einem 5%-Niveau geprüft werden. Eine entsprechende Zufallsstichprobe mit 50 Personen ergab einen Anteil von 20 Frauen.

1. *Statistische Hypothesen:* $H_0: \pi \geq 0,5$ $H_1: \pi < 0,5$

2. *Festlegung des Signifikanzniveaus:* $\alpha = 0,05$

3. *Auswahl des Verfahrens:*
$$z = \frac{p - \pi_0}{\sqrt{\frac{\pi_0 \cdot (1 - \pi_0)}{n}}}$$

Voraussetzung:
$$n > \frac{9}{\pi_0 \cdot (1 - \pi_0)} = \frac{9}{0,5 \cdot (1 - 0,5)} = 36$$

4. *Stichprobenwerte:* $n = 50$ (Voraussetzung $n > 36$ o.k.)
 $f = 20 \Rightarrow p = \frac{f}{n} = \frac{20}{50} = 0,4$
 $\pi_0 = 0,5$

5. *Berechnung der Prüfgröße:*
$$z = \frac{0,4 - 0,5}{\sqrt{\frac{0,5 \cdot 0,5}{50}}} = -1,41$$

Variante „Kritische Grenze“:

6. *Ermittlung der kritischen Grenze:* $-z_{1-\alpha} = -z_{1-0,05} = -z_{0,95} = -1,64$

7. *Entscheidung:* $-1,41 > -1,64$

\Rightarrow Nullhypothese wird beibehalten

Variante „Überschreitungswahrscheinlichkeit“:

6. *Überschreitungswahrscheinlichkeit:* $P = 1 - \Phi(|-1,41|) = 1 - 0,919 = 0,081$

7. *Entscheidung:* $0,081 > 0,05$

\Rightarrow Nullhypothese wird beibehalten

Damit ist mit einer statistischen Sicherheit von 95% die Hypothese widerlegt, dass Frauen in einer bestimmten Berufsgruppe unterrepräsentiert seien.

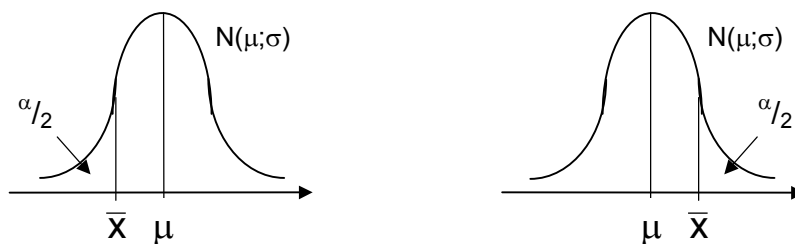
3. Konfidenzintervalle

Nachdem wir Signifikanztests zum Vergleich eines Stichprobenkennwertes (bisher Durchschnitte und Häufigkeiten) mit dem zugehörigen Populationsparameter kennengelernt haben, stellt sich die Frage, wie der Parameter einer Grundgesamtheit bestimmt werden kann. Leider sind die Parameter zumeist unbekannt. In diesem Fall müssen wir sie durch die Kennwerte aus der Stichprobe bestimmen. Wir betrachten hierzu die Kennwerte der Stichprobe aber als (mehr oder minder gute) *Schätzwerte* für die entsprechenden Parameter der Grundgesamtheit und müssen davon ausgehen, dass die Kennwerte von Zufallsstichprobe zu Zufallsstichprobe schwanken.

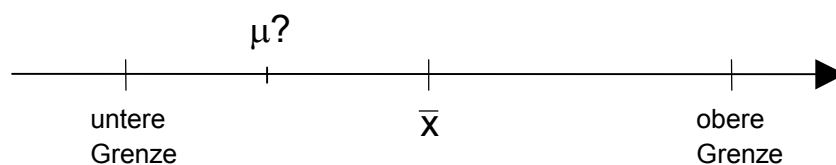
Stichprobenkennwerte als Schätzwerte für Populationsparameter sind als Werte einer Zufallsvariablen anzusehen, deren Verteilung bekannt sein muss, wenn wir die

Brauchbarkeit des konkreten Schätzwertes richtig bewerten wollen. Diese Verteilung, die wir im ersten Kapitel Stichprobenverteilung genannt haben, ist uns jedoch nur teilweise bekannt. Wir kennen – zumindest, wenn wir als Stichprobenkennwert erneut vorerst nur das arithmetische Mittel betrachten – ihre Verteilungsform, denn diese ist nach dem zentralen Grenzwertsatzes normalverteilt ($n > 30$). Auch ihre Streuung (Standardfehler) kennen wir bereits, die wir mit σ/\sqrt{n} bzw., wenn σ unbekannt, mit s/\sqrt{n} angeben können (s. S. 8 und 9). Unbekannt ist aber noch immer der zu schätzende Parameter μ . Nach wie vor ist damit die Frage offen, was wir über diesen unbekannt Parameter wissen, wenn wir nur das Ergebnis einer Stichprobenuntersuchung kennen.

Wir wollen folgendes Gedankenexperiment machen. Wir wissen, dass ein normalverteilter Stichprobenkennwert mit einer bestimmten Wahrscheinlichkeit in einem bestimmten Abstand vom Parameter μ entfernt liegt. Mit einer Wahrscheinlichkeit von 95% weicht er um 1,96 Standardabweichungen ab und mit einer Wahrscheinlichkeit von 99% um 2,58 Standardabweichungen. Das sind unsere kritischen Grenzen. Nehmen wir nun den ungünstigsten Fall an, dass unser Stichprobenkennwert tatsächlich um ganze 1,96 bzw. 2,58 Standardabweichungen vom Erwartungswert μ liegt. Dann kann er das unterhalb oder oberhalb des Parameters tun:



Wenn der Stichprobenkennwert im Extremfall mit einer Wahrscheinlichkeit von 95% um 1,96 bzw. mit einer Wahrscheinlichkeit von 99% um 2,58 Standardabweichungen vom Parameter μ abweicht, so kann man zwangsläufig daraus schließen, dass umgekehrt der gesuchte Parameter μ ebenfalls um die angegebenen Standardabweichungen von dem ermittelten Stichprobenkennwert entfernt liegen muss. Damit haben wir eine Bewertung dafür gefunden, wie brauchbar der Stichprobenkennwert als Schätzwert für den unbekannt Parameter ist. Wir können angeben, in welchem Bereich der Parameter μ um den Kennwert zu einer bestimmten Wahrscheinlichkeit vermutet werden kann. Dieser Bereich wird Konfidenzintervall genannt und die Wahrscheinlichkeit, dass der Parameter in diesem Konfidenzintervall liegt, die Sicherheitswahrscheinlichkeit $1-\alpha$ (Konfidenzniveau, Konfidenzkoeffizient):



Das Konfidenzintervall berechnet sich je nach Fall etwas unterschiedlich. Im Prinzip liegt jedoch immer eine Normal- bzw. t-Verteilung zugrunde. Für die untere Grenze des Konfidenzintervalls wird vom Stichprobenkennwert ein Abstand entsprechend

der kritischen Grenze zu einer Sicherheitswahrscheinlichkeit abgezogen, für die obere Grenze dieser Abstand auf den Stichprobenkennwert aufaddiert. Dieser Abstand berechnet sich, indem die kritische Grenze zu einem Konfidenzniveau mit der Streuung der Stichprobenverteilung, die ja bekannt ist, multipliziert wird, da die kritischen Grenzen z-Werte sind und diese dadurch gewonnen werden, indem Abstände durch Streuungen dividiert werden. Der Vorgang der z-Transformation wird also einfach umgekehrt. Für drei verschiedene Fälle werden wir Konfidenzintervalle berechnen:

3.1. Konfidenzintervall eines Parameters μ mit bekanntem σ

Voraussetzung: σ bekannt

Berechnung:
$$\mu = \bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Eine weitere Voraussetzung hierfür ist eigentlich, dass die Stichprobe mehr als 30 Personen umfasst, wenn die zugrundeliegenden Messwerte nicht normalverteilt sind, da nach dem zentralen Grenzwertsatz Stichprobendurchschnitte sich erst ab dieser Größe hinreichend normalverteilen. Bei nicht normalverteilten Messwerten und Stichproben unter 30 Personen verhält sich die Durchschnittsverteilung t-verteilt, so dass in dieser Formel dann nicht die z-, sondern die entsprechende t-Grenze eingesetzt wird. Allerdings dürfte der Fall, dass ein Parameter anhand einer Stichprobe mit $n < 30$ geschätzt wird, in der Praxis selten vorkommen, da bei diesem Umfang nur sehr ungenaue Parameterschätzungen möglich sind. Für Haus- und Klausuraufgaben ist dieser Umstand irrelevant.

3.2. Konfidenzintervall eines Parameters μ mit unbekanntem σ

Berechnung:
$$\mu = \bar{x} \pm t_{1-\frac{\alpha}{2}; n-1} \cdot \frac{s}{\sqrt{n}}$$

Bei unbekannter Streuung σ wird sie durch die Standardabweichung s der Stichprobe geschätzt, wodurch die Durchschnittsverteilung sich dann t-verteilt verhält.

3.3. Konfidenzintervall eines Parameters π

Voraussetzung:
$$n > \frac{9}{p \cdot (1-p)}$$

Berechnung:
$$\pi = p \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$$

Da der Parameter π unbekannt ist, wird er in den uns bereits bekannten Formeln durch den Stichprobenkennwert p geschätzt. Der Kennwert p ist die relative Häufigkeit einer Ausprägung eines Merkmals: $p = f/n$. Die Stichprobenverteilung von relativen Häufigkeit nähert sich erst mit obiger Voraussetzung hinsichtlich des Stichprobenumfangs der Normalverteilung an.

Beispiele für Konfidenzintervalle

Aufgabe: Bei einer Person wird der IQ mit 110 bestimmt. Wie groß ist der wahre IQ dieser Person bei einer statistischen Sicherheit von 95%? Die Streuung der IQ-Werte einer Person ist bekannt mit $\sigma=5$.

$$\begin{aligned}\mu &= \bar{x} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} & 1-\alpha &= 0,95 \\ &= 110 \pm z_{0,975} \cdot \frac{5}{\sqrt{1}} & \alpha &= 0,05 \\ &= 110 \pm 1,96 \cdot 5 & 1-\alpha/2 &= 0,975 \\ &= 110 \pm 9,8 \\ 100,2 &\leq \mu \leq 119,8\end{aligned}$$

Lösung: Mit einer Sicherheitswahrscheinlichkeit von 95% liegt der wahre IQ dieser Person zwischen 100,2 und 119,8.

Aufgabe: Die durchschnittliche Körpergröße eines neu entdeckten Stammes im Amazonasgebiet soll auf einem Konfidenzniveau von 99 % ermittelt werden. Eine Stichprobe von 25 Eingeborenen ergibt eine mittlere Körpergröße von 160cm mit einer Standardabweichung von 12cm.

$$\begin{aligned}\mu &= \bar{x} \pm t_{1-\alpha/2; n-1} \cdot \frac{s}{\sqrt{n}} & 1-\alpha &= 0,99 \\ &= 160 \pm t_{0,995; 24} \cdot \frac{12}{\sqrt{25}} & \alpha &= 0,01 \\ &= 160 \pm 2,80 \cdot 2,4 & 1-\alpha/2 &= 0,995 \\ &= 160 \pm 6,7 \\ 153,3 &\leq \mu \leq 166,7\end{aligned}$$

Lösung: Mit einer statistischen Sicherheit von 99% liegt die durchschnittliche Körpergröße des neu entdeckten Stammes zwischen 153,3cm und 166,7cm.

Aufgabe: Ein Therapeut hat bei einer Stichprobe von 100 Klienten eine Erfolgsquote von 20%. Mit welcher Erfolgsquote kann er mit einem Konfidenzkoeffizienten von 0,95 in Zukunft rechnen?

$$\text{Voraussetzung } n > \frac{9}{p \cdot (1-p)} = \frac{9}{0,2 \cdot 0,8} = 56,25 \text{ ist mit 100 Personen erfüllt.}$$

$$\begin{aligned}
 \pi &= p \pm z_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} \\
 &= 0,2 \pm z_{0,975} \cdot \sqrt{\frac{0,2 \cdot 0,8}{100}} \\
 &= 0,2 \pm 1,96 \cdot 0,04 \\
 &= 0,2 \pm 0,08 \\
 &0,12 \leq \pi \leq 0,28
 \end{aligned}$$

$1-\alpha = 0,95$
 $\alpha = 0,05$
 $1-\alpha/2 = 0,975$

Lösung: Auf einem Konfidenzniveau von 95% kann der Therapeut in Zukunft mit einem Behandlungserfolg zwischen 12% und 28% rechnen.

3.4. Mathematische Herleitung zum Konfidenzintervall

$$\begin{aligned}
 1-\alpha &= W \left\{ \bar{x} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right\} && | -\bar{x} \\
 &= W \left\{ -z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu - \bar{x} \leq z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right\} && | \div \frac{\sigma}{\sqrt{n}} \\
 &= W \left\{ -z_{1-\alpha/2} \leq \frac{\mu - \bar{x}}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2} \right\} && | \cdot (-1) \\
 &= W \left\{ z_{1-\alpha/2} \geq \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \geq -z_{1-\alpha/2} \right\} && | \text{umdrehen} \\
 &= W \left\{ -z_{1-\alpha/2} \leq z(\bar{x}) \leq z_{1-\alpha/2} \right\}
 \end{aligned}$$

3.5. Steigerung der Präzision des Konfidenzintervalls

Die Breite eines Konfidenzintervalls ist die Differenz zwischen der unteren und der oberen Grenze eines Konfidenzintervalls und entspricht der doppelten Breite des Abstandes, der in der Berechnung des Intervalls zum Stichprobendurchschnitt addiert bzw. subtrahiert wird:

$$\left(\bar{x} + z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right) - \left(\bar{x} - z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right) = 2 \cdot \left(z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

Auf die Frage, wie die Präzision des Konfidenzintervalls erhöht werden kann, womit gemeint ist, inwieweit die Breite des Konfidenzintervalls, indem μ vermutet wird, verengt werden kann, können in Betrachtung des rechten Teils des obigen Ausdrucks folgende Antworten gegeben werden:

- Je größer die Irrtumswahrscheinlichkeit α ist, desto kleiner wird die z-Grenze im Ausdruck und um so schmaler und damit auch präziser das Konfidenzintervall.
 ⇒ *Je unsicherer eine Aussage, desto präziser ist die Aussage.*
- Je größer der Stichprobenumfang n ist, desto kleiner wird der Ausdruck σ/\sqrt{n} in der Formel und um so schmaler und präziser die Breite des Konfidenzintervalls.
 ⇒ *Je größer die Stichprobe, um so präziser ist die Aussage.*
- Je kleiner die Streuung σ des Merkmals bzw. der Messwerte der Stichprobe, desto schmaler wird das Konfidenzintervall. σ hat verfahrensabhängig einen Einfluss.
 ⇒ *Je kleiner die Streuung σ , desto präziser ist die Aussage.*

4. Signifikanztests II

Im folgenden werden wir weitere Verfahren zur Prüfung auf Signifikanz kennenlernen, die sich zu denen von Kapitel Signifikanztests I dadurch unterscheiden, dass andere Kennwerte als der Durchschnitt oder Anteilwert einer Stichprobe auf Signifikanz geprüft werden. Es handeln sich um die restlichen Verfahren der Übersicht der speziellen Signifikanztests bei HUBER, STAT -9-.

4.1. Prüfung von Korrelationshypothesen

Der Korrelationskoeffizient r einer Stichprobe wird bei Ausweitung auf die Grundgesamtheit ρ (ρ) genannt. Die Stichprobenverteilungen von unendlich vielen Korrelationskoeffizienten hat einen Erwartungswert von $\mu=\rho=0$, d.h. es wird angenommen, dass kein Zusammenhang besteht. Entsprechend werden die statistischen Hypothesen wie folgt formuliert:

Zweiseitige Tests		Bedeutung	
$H_0: \rho=0$		kein sign. Zus.hang	
$H_1: \rho \neq 0$		signifikanter Zus.hang	
Einseitige Tests			
<i>rechtsseitig</i>	<i>Bedeutung</i>	<i>linksseitig</i>	<i>Bedeutung</i>
$H_0: \rho \leq 0$	kein sign. Zus.hang	$H_0: \rho \geq 0$	kein sign. Zus.hang
$H_1: \rho > 0$	sign. pos. Zus.hang	$H_1: \rho < 0$	sign. neg. Zus.hang

Anmerkung: Dem interessierten Leser sei geraten, sich die Sachverhalte an einer Zeichnung der Stichprobenverteilung von Korrelationskoeffizienten mit $\mu=\rho=0$ zu verdeutlichen.

Die Prüfgröße dieses Verfahrens ist ein t-Wert und wird für den Produkt-Moment-Korrelationskoeffizient r von Pearson berechnet. Im Zähler wird von r im Prinzip wie üblich der Erwartungswert von $\mu=\rho=0$ abgezogen, im Nenner steht der Standardfehler der Stichprobenverteilung von Korrelationskoeffizienten. Besondere Beachtung verdient, dass der Freiheitsgrad df zur t-Grenze sich jetzt durch den um zwei verminderten Stichprobenumfang berechnet:

$$t = \frac{r}{\sqrt{\frac{(1-r^2)}{(n-2)}}}; \quad df = n-2$$

Beispiel eines einseitigen Signifikanztests: Überprüfung von Korrelationshypothesen

Angaben: Bei Kindern soll ihre Taschengeldhöhe mit ihrem Konsum an Süßigkeiten positiv zusammenhängen. Eine Stichprobe an 36 Kindern ergab einen Korrelationskoeffizienten zwischen Taschengeld und verzehrten Süßwarenkalorien in Höhe von $r=0,4$. Die Prüfung erfolgt mit einer Irrtumswahrscheinlichkeit von 5%.

1. *Statistische Hypothesen:* $H_0: \rho \leq 0 \quad H_1: \rho > 0$
2. *Festlegung des Signifikanzniveaus:* $\alpha = 0,05$
3. *Auswahl des Verfahrens:*
$$t = \frac{r}{\frac{\sqrt{(1-r^2)}}{\sqrt{(n-2)}}}; df = n-2$$
4. *Stichprobenwerte:* $r = 0,4$
 $n = 36$
5. *Berechnung der Prüfgröße:*
$$t = \frac{0,4}{\frac{\sqrt{(1-0,4^2)}}{\sqrt{(36-2)}}} = 2,54$$
6. *Ermittlung der kritischen Grenze:* $t_{1-\alpha;df} = t_{1-0,05;36-2} = t_{0,95;34} = 1,69$
7. *Entscheidung:* $2,54 > 1,69$
 \Rightarrow Nullhypothese wird verworfen

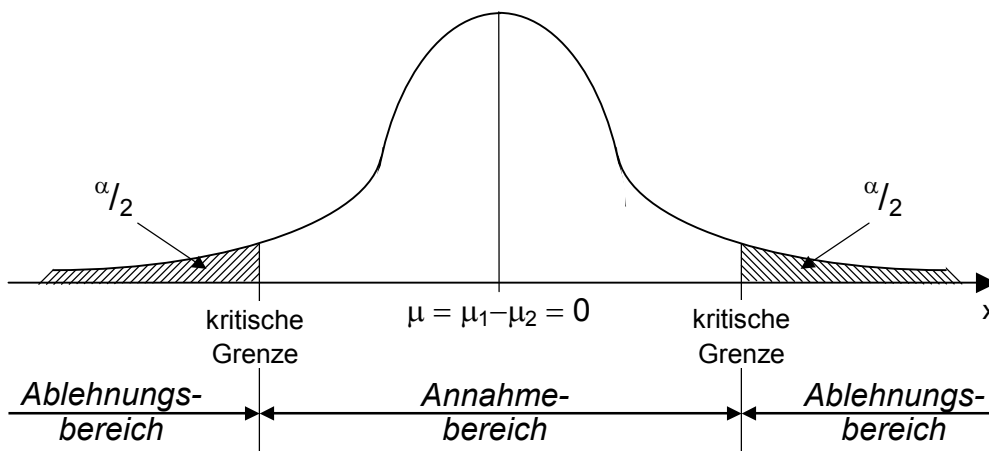
Damit ist auf dem 5%-Signifikanzniveau die Hypothese bestätigt, dass es bei Kindern einen pos. Zusammenhang zwischen Taschengeldhöhe u. Süßigkeitenkonsum gibt.

4.2. Prüfung von Unterschiedshypothesen

Die für das psychologischen Experiment wichtigsten Signifikanztests dienen dem Vergleich zweier Stichprobenmittelwerte. Das Wesentliche eines Experimentes liegt darin, dass der Versuchsleiter systematisch einen Faktor (UV=unabhängige VariableEinflussgröße , Bedingung) variiert, um seinen Effekt (Wirkung) auf die abhängige Variable (AV=abhängige Variable, das untersuchte Merkmal) zu ermitteln, während er versucht, alle anderen möglichen Faktoren konstant zu halten, um möglichen Störeinflüssen vorzubeugen. Der kleinstmögliche Versuchsplan sieht die Variation eines Faktors in zwei Abstufungen vor: In einer Experimentalgruppe wird eine Zufallsstichprobe dem Einfluss des Faktors unterworfen, in einer Kontrollgruppe wird eine andere Zufallsstichprobe nicht der vermuteten (oder einer anderen Wirkung) des Faktors ausgesetzt. In beiden Fällen wird das zu untersuchende Merkmal gemessen. Nun ist es für den Forscher interessant, ob die statistische Auswertung der Versuchsdaten einen signifikanten Effekt der unabhängigen Variable auf die abhän-

gige Variable bestätigen kann, ob also die Einflussgröße tatsächlich die vermutete Wirkung auf das untersuchte Merkmal ausübt, was sich in unterschiedlichen Mittelwerten der beiden Versuchsgruppen ausdrückt. Der Mittelwert des untersuchten Merkmals der Experimentalgruppe, in dem der Faktor wirkt, wird also mit dem Mittelwert der Kontrollgruppe, in dem der Faktor keinen (oder anderen) Einfluss hat, verglichen.

Hierzu wird die Differenz der Mittelwerte der beiden Stichproben gebildet. Ist die Differenz Null, so liegt kein Unterschied vor. Ist die Differenz positiv oder negativ, so muss ein Unterschied vorliegen. Dieser Unterschied kann zufällig auftreten, daher drängt sich die Frage auf, wann diese Differenz signifikant ist. Um diese Frage zu klären, wird auch hier ein Signifikanztest durchgeführt. Grundlage dieses inferenzstatistischen Prüfverfahrens ist die normalverteilte Stichprobenverteilung der *Differenz* der beiden Stichprobenmittelwerte:



Der Erwartungswert μ dieser Stichprobenverteilung entspricht der Vermutung der H_0 , dass die Differenz der Stichprobenmittelwerte statistisch gleich Null ist: μ_1 und μ_2 sind gleichgroß. μ_1 ist der Erwartungswert für die Stichprobenverteilung der Grundgesamtheit der Stichproben, die dem Einfluss des vermuteten Faktors unterliegen (Experimentalgruppen). μ_2 ist der Erwartungswert der Stichprobenverteilung für die Grundgesamtheit der Stichproben, wenn der Faktor keinen (oder einen anderen) Einfluss auf sie ausübt (Kontrollgruppen). Entsprechend lauten die statistischen Hypothesen:

Zweiseitige Tests	Bedeutung
$H_0: \mu_1 = \mu_2 \Rightarrow H_0: \mu_1 - \mu_2 = 0$	kein signifikanter Unterschied zwischen den Mittelwerten; kein Effekt des Faktors
$H_1: \mu_1 \neq \mu_2 \Rightarrow H_1: \mu_1 - \mu_2 \neq 0$	signifikanter <i>Unterschied</i> zwischen den Mittelwerten
Rechtsseitige Tests	Bedeutung
$H_0: \mu_1 \leq \mu_2 \Rightarrow H_0: \mu_1 - \mu_2 \leq 0$	kein signifikanter Unterschied zwischen den Mittelwerten; kein Effekt des Faktors
$H_1: \mu_1 > \mu_2 \Rightarrow H_1: \mu_1 - \mu_2 > 0$	\bar{X}_1 signifikant <i>größer</i> als \bar{X}_2

Linksseitige Tests	Bedeutung
$H_0: \mu_1 \geq \mu_2 \Rightarrow H_0: \mu_1 - \mu_2 \geq 0$	kein signifikanter Unterschied zwischen den Mittelwerten; kein Effekt d. Faktors
$H_1: \mu_1 < \mu_2 \Rightarrow H_1: \mu_1 - \mu_2 < 0$	\bar{X}_1 signifikant <i>kleiner</i> als \bar{X}_2

Je nachdem, ob *unabhängige* oder *abhängige Stichproben* vorliegen, werden unterschiedliche Tests durchgeführt.

- Unabhängige oder unverbundene Stichproben liegen vor, wenn in beiden Stichproben *nicht* dieselben Versuchspersonen teilnehmen. Der Forscher spricht in diesem Fall von einem Versuchsplan *ohne Messwiederholung* (interindividuelle Bedingungsvariation).

Beispiel: Ein Wissenschaftler möchte die Auswirkung eines fiebersenkenden Medikamentes untersuchen. Dazu bildet er zwei Stichproben: Eine Experimentalgruppe aus fiebrigen Personen, die das Medikament erhält und eine Kontrollgruppe aus ebenfalls fiebrigen Personen, die statt des Medikamentes ein Placebo erhält. Nach vermuteten Eintritt der Wirkung des Medikamentes wird die Temperatur gemessen und eine inferenzstatistische Auswertung durchgeführt.

- Abhängige oder verbundene Stichproben liegen vor, wenn die Messwerte paarweise einander zugeordnet sind (aus jeder Stichprobe jeweils einer). Dies ist der Fall, wenn in beiden Stichproben *dieselben* Versuchspersonen teilnehmen. Der Forscher spricht dann von einem Versuchsplan *mit Messwiederholung* (intraindividuelle Bedingungsvariation). Auch *parallelisierte* Stichproben gelten als abhängig, da durch die Parallelisierung (zur Kontrolle von Störeinflüssen) die Messwerte von jeweils zwei Personen aufgrund gleicher oder ähnlicher Merkmalsausbildungen der Personen paarweise einander zugewiesen sind.

Beispiel: Derselbe Forscher könnte die Untersuchung auch mit einer Stichprobe realisieren: Einmal misst er die Temperatur vor Verabreichung des Medikamentes, einmal nach vermutetem Eintritt der Wirkung des Medikamentes. Da zwei Reihen an Versuchsdaten ermittelt werden, kann man theoretisch von zwei Stichproben sprechen.

Auch das zuerst genannte Beispiel gilt als abhängig, wenn die beiden Stichproben hinsichtlich eines oder mehrerer Störfaktoren parallelisiert werden. Solch eine störende Einflussgröße könnte sein, dass von vorne herein in der Experimentalgruppe nur leicht überhöhte Temperaturen und in der Kontrollgruppe äußerst fiebrige Patienten vorkommen. Die Mittelwerte der beiden Gruppen unterscheiden sich dann auch ohne medikamentöse Einwirkung und ein scheinbar signifikanter Temperaturabfall ist auch dann feststellbar, wenn das Medikament keine Wirkung ausübt. Um diesen Störfall auszuschließen, ist eine Parallelisierung der Versuchspersonen hinsichtlich ihrer Körpertemperatur ratsam. Dazu werden aus einer Zufallsstichproben aufgrund eines Vortestes, in dem das Fieber gemessen wird, die beiden Versuchsgruppen gebildet. Die beiden Personen mit den höchsten Temperaturen werden auf die Experimental- und die Kontrollgruppe verteilt, danach die beiden Personen mit den zweithöchsten Temperaturen usw. Dadurch ist sichergestellt, dass die beiden Versuchsgruppen gleich hohe Mittelwerte hinsichtlich der Körpertemperatur vor dem Versuch aufweisen.

4.2.1. Vergleich der Mittelwerte zweier abhängiger Stichproben

Von jeder Versuchsperson wird die Differenz der ersten Messung und der zweiten Messung seines Merkmals *einzel*n gebildet. Zu jeder Person wird also eine eigene Kategorie von Messwert gebildet: die Differenz zwischen seinem ersten und seinem zweiten Messwert:

$$d_i = x_{i1} - x_{i2}$$

Die Prüfgröße zur Stichprobenverteilung von Mittelwertdifferenzen ist ein t-Wert. Der Freiheitsgrad df zur t-Grenze wird durch den um eins verminderten Stichprobenumfang n berechnet:

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}; \quad df = n-1$$

Im Zähler wird vom Durchschnitt der Differenzen d_i im Prinzip wie üblich der Erwartungswert $\mu = \mu_1 - \mu_2 = 0$ abgezogen, im Nenner steht der Standardfehler der Stichprobenverteilung von Messwertpaardifferenzen. Dieser wird – wie bei t-Werten der Fall – durch den Quotienten aus der Standardabweichung s_d der Messwertpaardifferenzen d_i und der Wurzel des Stichprobenumfangs n gebildet. Durchschnitt und Standardabweichung der Messwertpaardifferenzen sind:

$$\bar{d} = \frac{\sum d_i}{n} \quad s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}}$$

Voraussetzung dieses Verfahrens ist es, dass der Stichprobenumfang $n > 30$ ist. Bei kleineren Stichproben muss die Voraussetzung erfüllt sein, dass sich die Differenzen in der Grundgesamtheit normalverteilen. Dies kann mit einem besonderen Verfahren geprüft werden, was nicht Gegenstand dieses Kurses ist, da dieser Test auf Voraussetzungsverletzungen relativ robust reagiert. In Aufgaben gibt es daher keine Beschränkungen zu beachten.

Beispiel eines einseitigen Signifikanztests: Vergleich zweier Mittelwerte von abhängigen Stichproben

Angaben: Schlechte Schüler sollen mit einem neu entwickelten Förderunterricht bessere Leistungen erbringen. In einer Zufallsstichprobe wird die Leistung der Schüler sowohl vor der Durchführung des Förderunterrichts als auch danach durch Zeugnisnoten erhoben. (Signifikanzniveau: 5%)

1. *Statistische Hypothesen:* $H_0: \mu_1 - \mu_2 \leq 0$ $H_1: \mu_1 - \mu_2 > 0$

2. *Festlegung des Signifikanzniveaus:* $\alpha = 0,05$

3. *Auswahl des Verfahrens:* $t = \frac{\bar{d}}{s_d / \sqrt{n}}; \quad df = n-1$

Voraussetzung:

- Stichproben sind abhängig

4. Stichprobenwerte:

x_{i1}	x_{i2}	$d_i = x_{i1} - x_{i2}$	$d_i - \bar{d}$	$(d_i - \bar{d})^2$
5	3	2	0	0
4	2	2	0	0
6	3	3	1	1
4	3	1	-1	1
3	1	2	0	0
		10		2

$$n=5 \quad \bar{d} = \frac{\sum d_i}{n} = \frac{10}{5} = 2 \quad s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{2}{4}} = 0,707$$

5. Berechnung der Prüfgröße:

$$t = \frac{2}{\frac{0,707}{\sqrt{5}}} = 6,33$$

6. Ermittlung der kritischen Grenze:

$$t_{1-\alpha; df} = t_{1-0,05; 5-1} = t_{0,95; 4} = 2,13$$

Entscheidung:

$$6,3 > 2,13 \Rightarrow H_1 \text{ wird bestätigt}$$

Damit ist auf dem 5%-Niveau der Verlässlichkeit die Hypothese angenommen, dass der neu entwickelte Förderunterricht die Leistung bei schlechten Schülern verbessert.

4.2.2. Vergleich der Mittelwerte zweier unabhängiger Stichproben

Das Verfahren zum Mittelwertvergleich von abhängigen Stichproben beruht auf einer Stichprobenverteilung der Mittelwerte \bar{d} der Messwertdifferenzen d_i der jeweiligen Einzelpersonen. Diese Stichprobenverteilung ist also eine Wahrscheinlichkeitsverteilung von Durchschnitten von Messwertpaardifferenzen. Dies kann gemacht werden, da die beiden Stichproben entweder aus denselben oder aus einander zugewiesenen Personen bestehen. Bei unabhängigen Stichproben tritt das Problem auf, dass die beiden Stichproben nicht aus den gleichen oder ähnlichen Personen bestehen, so dass hier keine Differenz der Messwerte als eigenständige Messgröße, deren Mittelwert stichprobenverteilt betrachtet werden kann, gebildet werden kann. Wir brauchen also ein anderes Verfahren. Wir betrachten nun nicht mehr die Verteilung der Stichprobendurchschnitte von Messwertpaardifferenzen, sondern die Verteilung von Stichprobenmittelwertdifferenzen, wie es bereits eingangs auf S. 35 erläutert wurde. Der Erwartungswert $\mu_{\bar{x}_1 - \bar{x}_2}$ dieser Stichprobenverteilung ist mit Null bekannt, jedoch wie wird die Zufallsstreuung, der Standardfehler $\sigma_{\bar{x}_1 - \bar{x}_2}$ der Stichprobenverteilung der Mittelwertdifferenzen bestimmt, um eine Prüfgröße aufstellen zu können? Je nachdem, ob die Streuungen der Grundgesamtheiten der Messwerte beider Stichproben bekannt sind, nicht bekannt sind oder zwar nicht bekannt, aber als gleich hoch eingeschätzt werden können, kann man drei Fälle unterscheiden, woraus sich drei verschiedene Prüfgrößen und Verfahren ergeben:

Ausgangslage	Verfahren
σ_1, σ_2 bekannt	z-Test zum Mittelwertvergleich
σ_1, σ_2 unbekannt	WELCH-Test
σ_1, σ_2 unbekannt, aber $\sigma_1 = \sigma_2$	klassischer t-Test

4.2.2.1. z-Test bei bekanntem σ

Wenn die Standardabweichungen σ_1 und σ_2 der Normalverteilungen des gemessenen Merkmals in den beiden Stichproben, also von den Grundgesamtheiten der beiden Stichproben, bekannt sind, lassen sich Durchschnittsverteilungen mit folgenden Standardfehlern und Varianzen aufstellen:

$$\sigma_{\bar{x}_1} = \frac{\sigma_1}{\sqrt{n_1}} \Leftrightarrow \sigma_{\bar{x}_1}^2 = \frac{\sigma_1^2}{n_1}$$

$$\sigma_{\bar{x}_2} = \frac{\sigma_2}{\sqrt{n_2}} \Leftrightarrow \sigma_{\bar{x}_2}^2 = \frac{\sigma_2^2}{n_2}$$

Nun wird eine neue Zufallsvariable gebildet: die Differenz der Durchschnitte von zwei Stichproben. Mit Satz 9 (HUBER, STAT -7-) gilt als die gesuchte Zufallsstreuung der Mittelwertdifferenzenverteilung:

$$\sigma_{x-y}^2 = \sigma_x^2 + \sigma_y^2 \Rightarrow \sigma_{\bar{x}_1 - \bar{x}_2}^2 = \sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2 \Leftrightarrow \sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2}$$

Da der gesuchte Standardfehler $\sigma_{\bar{x}_1 - \bar{x}_2}$ der Verteilung von Stichprobenmittelwertdifferenzen jetzt bekannt ist, können wir eine z-Prüfgröße ableiten:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\sigma_{\bar{x}_1}^2 + \sigma_{\bar{x}_2}^2}} \Leftrightarrow z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Beispiel eines zweiseitigen Signifikanztests zur Prüfung von Unterschiedshypothesen: z-Test zum Mittelwertvergleich zweier unabhängiger Stichproben (σ bekannt)

Angaben: Die inhaltliche Hypothese besagt, dass zwei bestimmte Personen A und B sich in ihrem wahren IQ unterscheiden. Es wird mit jeder Person ein standardisierter Intelligenztests durchgeführt. Es ergibt sich ein IQ von 115 bzw. 125. Die Standardabweichung σ der IQ-Testergebnisse von einer Person ist mit 5 bekannt. Ist der Unterschied zwischen den beiden Testergebnissen auf dem 5%-Niveau signifikant?

1. *Statistische Hypothesen:* $H_0: \mu_1 - \mu_2 = 0$ $H_1: \mu_1 - \mu_2 \neq 0$

2. *Festlegung des Signifikanzniveaus:* $\alpha = 0,05$

3. *Auswahl des Verfahrens:*

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Voraussetzungen:

- Stichproben sind unabhängig
- Varianzen sind bekannt

4. Stichprobenwerte:

$$\begin{aligned}\bar{x}_1 &= 115, & \bar{x}_2 &= 125 \\ \sigma_1 &= 5, & \sigma_2 &= 5 \\ n_1 &= 1, & n_2 &= 1\end{aligned}$$

5. Berechnung der Prüfgröße:

$$z = \frac{115 - 125}{\sqrt{\frac{5^2}{1} + \frac{5^2}{1}}} = -1,41$$

Variante „Kritische Grenze“:

6. Ermittlung der kritischen Grenze:

$$z_{1-\alpha/2} = z_{1-0,05/2} = z_{0,975} = 1,96$$

7. Entscheidung:

$$|-1,41| < 1,96 \Rightarrow \text{Nullhypothese bestätigt}$$

Variante „Überschreitungswahrscheinlichkeit“:

6. Überschreitungswahrscheinlichkeit:

$$P = 2 \cdot (1 - \Phi(|-1,41|)) = 2 \cdot (1 - 0,919) = 0,162$$

7. Entscheidung:

$$0,162 > 0,05 \Rightarrow \text{Nullhypothese bestätigt}$$

Damit ist mit einer Irrtumswahrscheinlichkeit von 5% die Hypothese widerlegt, dass sich die Personen A und B mit den Testergebnissen von 115 bzw. 125 in ihrem wahren IQ unterscheiden.

4.2.2.2. WELCH-Test

Der WELCH-Test wird für den Fall benutzt, dass die Standardabweichungen σ_1 und σ_2 nicht bekannt sind. Der Test verwendet als Prüfgröße die t-Variante des oben erläuterten z-Prüfwerts:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Allerdings gestaltet sich die Berechnung des Freiheitsgrades df zur kritischen Grenze etwas kompliziert, falls die beiden Stichprobengrößen nicht größer als 50 sind:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1 - 1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2 - 1)}}$$

Sind die Stichprobenumfänge $n_1, n_2 \geq 50$, entspricht die t-Verteilung annähernd der Standardnormalverteilung, so dass die entsprechenden z-Grenzen (t-Werte mit $df = \infty$) benutzt werden können.

Beispiel eines zweiseitigen Signifikanztests zur Prüfung von Unterschiedshypothesen: WELCH-Test zum Mittelwertvergleich zweier unabhängiger Stichproben (σ unbekannt)

Angaben: Die inhaltliche Hypothese lautet, dass zwei unterschiedliche Lehrmethoden auch einen unterschiedlichen Erfolg bewirken. Es wurden zwei Klassen gesucht und diese einen Zeitraum lang nach den beiden Methoden unterrichtet. Zum Abschluss wurde ein Leistungstest durchgeführt. Die Klasse, die nach Methode A unterrichtet wurde, bestand aus 10 Schülern und erbrachte eine durchschnittliche Leistung von 51 Punkten bei einer Standardabweichung von 12 Punkten. Die andere Klasse mit 30 Schülern wurde nach Methode B unterrichtet und erbrachte eine durchschnittliche Leistung von 59 Punkten bei einer Standardabweichung von 7 Punkten. Es soll auf dem 5%-Niveau auf Signifikanz geprüft werden.

1. *Statistische Hypothesen:* $H_0: \mu_1 - \mu_2 = 0$ $H_1: \mu_1 - \mu_2 \neq 0$

2. *Festlegung des Signifikanzniveaus:* $\alpha = 0,05$

3. *Auswahl des Verfahrens:*
$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}; \quad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1 - 1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2 - 1)}}$$

Voraussetzungen:

- Stichproben sind unabhängig
- Varianzen sind unbekannt

4. *Stichprobenwerte:*

$$\begin{array}{ll} \bar{x}_1 = 51 & \bar{x}_2 = 59 \\ s_1 = 12 & s_2 = 7 \\ n_1 = 10 & n_2 = 30 \end{array}$$

$\frac{s_1^2}{n_1} = 12^2 / 10 = 14,4$
$\frac{s_2^2}{n_2} = 7^2 / 30 = 1,63$

5. *Berechnung der Prüfgröße:*

$$t = \frac{51 - 59}{\sqrt{14,4 + 1,63}} = -2$$

6. *Ermittlung der kritischen Grenze:*

$$df = \frac{(14,4 + 1,63)^2}{\frac{(14,4)^2}{(10 - 1)} + \frac{(1,63)^2}{(30 - 1)}} = 11,06$$

$$t_{1-\frac{\alpha}{2}; df} = t_{0,975; 11} = 2,20$$

7. *Entscheidung:*

$$|-2| < 2,20 \Rightarrow \text{Nullhypothese bestätigt}$$

Damit ist mit einer statistischen Sicherheit von 95% die Hypothese widerlegt, dass die beiden Lehrmethoden einen unterschiedlichen Erfolg bewirken.

4.2.2.3. Klassischer t-Test

Der klassische t-Test geht genauso wie der WELCH-Test von der oben vorgestellten z-Formel auf S. 40 aus, schätzt aber die Streuungen σ_1 und σ_2 nicht durch die Stan-

Standardabweichungen s_1 und s_2 der Stichproben, sondern geht davon aus, dass die Streuungen σ_1 und σ_2 gleich groß sind, was als Varianzhomogenität bezeichnet wird. Unter dieser Voraussetzung kann man für den Standardfehler der Stichprobenverteilung der Mittelwertdifferenzen ableiten:

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\sigma^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

σ^2 ist aber weiterhin unbekannt und wird daher durch die Größe s_1^2 geschätzt, die den Mittelwert der in den Stichproben gefundenen Varianzen s_1^2 und s_2^2 darstellt:

$$s_1^2 = \frac{(s_1^2 \cdot (n_1 - 1)) + (s_2^2 \cdot (n_2 - 1))}{(n_1 + n_2 - 2)} \quad \left[= \frac{SAQ_1 + SAQ_2}{(n_1 - 1) + (n_2 - 1)} \right]$$

Damit ergibt sich folgende Prüfgröße für den klassischen t-Test:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{\bar{x}_1 - \bar{x}_2}{s_1 \cdot \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}; \quad df = n_1 + n_2 - 2$$

Die Voraussetzungen zu diesem Verfahren sind vielfältig:

1. Die Durchführung eines t-Testes setzt voraus, dass die beiden Stichproben voneinander unabhängig sind, d.h. dass beide Stichproben nicht parallelisiert sind und auch nicht aus denselben Versuchspersonen bestehen.
2. Die Grundgesamtheiten müssen varianzhomogen sein, d.h. dass die Varianzen σ_1^2 und σ_2^2 in den – beide Stichproben zugrundeliegenden – Populationen gleich groß sind. Daher muss eine *Prüfung zur Gleichheit von Varianzen* stattfinden, die wir weiter unten behandeln werden und die durchgeführt werden muss, wenn in Aufgaben der klass. t-Test verlangt wird.
3. Bei kleineren Stichproben müssen sich die Grundgesamtheiten, aus denen die Stichproben entnommen wurden, normalverteilen. Sind die Verteilungsformen der den Stichproben zugrundeliegenden Grundgesamtheiten unbekannt, so sollte die Normalverteilungsannahme mit einem besonderen dafür zuständigen Testverfahren überprüft werden, welcher nicht Gegenstand dieses Kurses ist.

Sind die Voraussetzungen für den klassischen t-Test nicht erfüllt, wird der WELCH-Test benutzt. Sind die Voraussetzungen erfüllt, ist es selten, dass der WELCH-Test und der klassische t-Test zu unterschiedlichen Ergebnissen kommen. In der Vergangenheit wurde dabei der t-Test vorgezogen, da er eine größere Teststärke besitzt. Allerdings müssen für den klass. t-Test die Voraussetzungen erfüllt sein, sonst ist sein Ergebnis ungültig. Da bei der Prüfung der Voraussetzungen ebenfalls Tests eingesetzt werden, die ihrerseits Fehlerwahrscheinlichkeiten (α und β) in sich bergen, kann man sich hinsichtlich der Erfüllung der Voraussetzungen für den klassischen t-Test nie ganz sicher sein. In neuerer Zeit wird nun empfohlen, grundsätzlich den WELCH-Test durchzuführen, da dieser hinsichtlich seiner Voraussetzungen unproblematischer ist. In Aufgaben ohne nähere Angabe wird aufgrund dessen immer der WELCH-Test durchgeführt.

**Beispiel eines zweiseitigen Signifikanztests zur Prüfung von
Unterschiedshypothesen: Klassischer t-Test zum Mittelwertvergleich zweier
unabhängiger Stichproben (σ unbekannt, aber $\sigma_1 = \sigma_2$)**

Angaben: Beispiel von S. 42, jetzt den klass. t-Test statt dem WELCH-Test.

1. *Statistische Hypothesen:* $H_0: \mu_1 - \mu_2 = 0$ $H_1: \mu_1 - \mu_2 \neq 0$
2. *Festlegung des Signifikanzniveaus:* $\alpha = 0,05$
3. *Auswahl des Verfahrens:*
$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}; \quad df = n_1 + n_2 - 2$$
- $$s_1^2 = \frac{(s_1^2 \cdot (n_1 - 1) + s_2^2 \cdot (n_2 - 1))}{(n_1 + n_2 - 2)}$$

Voraussetzungen:

- Stichproben sind unabhängig
- Stichproben sind varianzhomogen

4. *Stichprobenwerte:* $\bar{x}_1 = 51$ $\bar{x}_2 = 59$
 $s_1 = 12$ $s_2 = 7$
 $n_1 = 10$ $n_2 = 30$
5. *Berechnung der Prüfgröße:*
$$s_1^2 = \frac{(12^2 \cdot 9 + 7^2 \cdot 29)}{(10 + 30 - 2)} = 71,5$$
- $$t = \frac{51 - 59}{\sqrt{71,5 \cdot \left(\frac{1}{10} + \frac{1}{30}\right)}} = -2,59$$
6. *Ermittlung der kritischen Grenze:* $t_{1-\frac{\alpha}{2}; df} = t_{1-0,05/2; 10+30-2} = t_{0,975; 38} = 2,02$
7. *Entscheidung:* $|-2,59| > 2,02 \Rightarrow H_0$ wird abgelehnt!

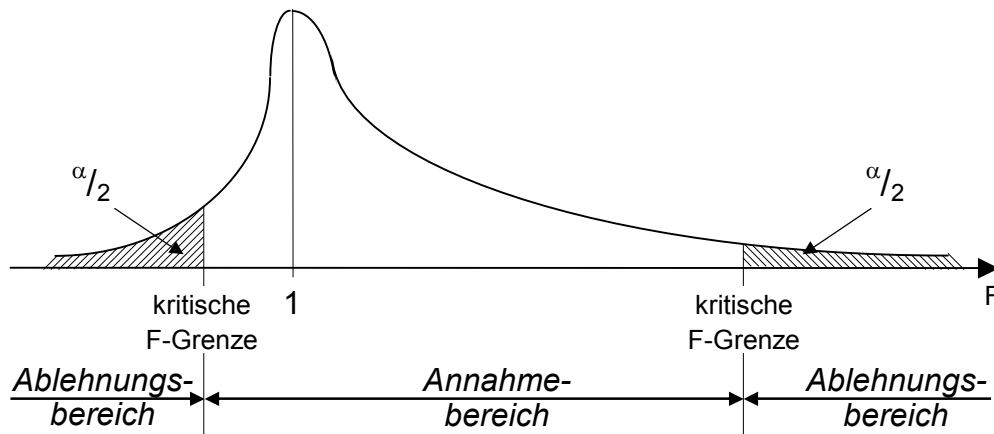
Damit ist mit einer statistischen Sicherheit von 95% die Hypothese bestätigt, dass die beiden Lehrmethoden einen unterschiedlichen Erfolg bewirken!

4.2.2.4. Prüfung auf Gleichheit zweier Stichprobenvarianzen (F-Test)

Die Nullhypothese lautet, dass die Varianzen der den Stichproben zugrundeliegenden Populationen gleich sind (Varianzhomogenität), die Alternativhypothese entsprechend umgekehrt: $H_0: \sigma_1^2 = \sigma_2^2$ und $H_1: \sigma_1^2 \neq \sigma_2^2$. Grundlage des Prüfverfahrens ist eine Prüfgröße, die den Quotienten aus den Varianzen der beiden Stichproben bildet, welcher F-Wert genannt wird:

$$F = \frac{s_1^2}{s_2^2}$$

Die F-Werte von unendlich vielen Stichprobenpaaren sind F-verteilt. Die F-Verteilung ist dabei asymmetrisch, wie aus folgender Abbildung ersichtlich ist:



Der Erwartungswert μ der F-Verteilung entspricht der Vermutung der Nullhypothese, dass die Varianzen gleich sind, so dass der Quotient der F-Prüfgröße und damit der Erwartungswert zu Eins wird. Eine große Abweichung von der Eins über die kritischen Grenzen hinweg führt zur Ablehnung der Nullhypothese. Da die Verteilung asymmetrisch ist, liegen die kritischen F-Grenzen nicht spiegelbildlich, was bedingt, dass die kritischen F-Grenzen, die in der F-Wert-Tabelle bei HUBER abgelesen werden, nur für die rechte Seite der Verteilung gelten. Dies hat zur Konsequenz, dass bei der Bildung der F-Prüfgröße folgende Regel eingehalten werden muss, damit die Prüfgröße größer als Eins wird und auf die rechte Seite der Verteilung zu liegen kommt:

$$s_1^2 > s_2^2$$

Die F-Verteilung hat zwei Freiheitsgrade, den Zählerfreiheitsgrad df_1 und den Nennerfreiheitsgrad df_2 :

$$df_1 = n_1 - 1 \quad df_2 = n_2 - 1$$

Das Signifikanzniveau α wird grundsätzlich auf 10% festgelegt, damit der β -Fehler möglichst klein wird. Eine falsche H_0 soll falsch bleiben, es ist lieber der α -Fehler in Kauf zu nehmen, dass irrtümlicherweise angenommen wird, dass die Varianzen nicht gleich groß sind, als einen t-Test durchzuführen, wo man nicht hinreichend genug sicher ist, ob dessen Voraussetzung gegeben ist. Beim Ablesen der kritischen F-Grenze aus der Tabelle bei HUBER ist daher zu beachten, dass die oberen zwei Hälften zu benutzen sind: $1 - \alpha/2 = 1 - 0,10/2 = 0,95$. Die linke Hälfte gilt für df_1 -Werte von 1 bis 10, die rechte Hälfte für df_1 -Werte von 12 bis „ ∞ “. Die kritische Grenze ist dann in der entsprechenden Zeile zu dem df_2 -Wert abzulesen. Auch beim F-Test müssen wir voraussetzen, dass die Grundgesamtheiten normalverteilt sind, die Prüfung hierzu wird nicht behandelt. Die Mittelwertverteilungen sind oft so „robust“, dass der F-Test auch dann durchgeführt werden kann, wenn die Populationen nicht normalverteilt sind, aber der Stichprobenumfang n hinreichend groß genug ist.

Beispiel eines zweiseitigen Signifikanztests zur Prüfung auf Gleichheit zweier Stichprobenvarianzen (F-Test)

Angaben: Für den klassischen t-Test zum Mittelwertvergleich zweier unabhängiger Stichproben auf S. 44 soll die Voraussetzung der Varianzhomogenität der zugrundeliegenden Populationen geprüft werden.

1. *Statistische Hypothesen:* $H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$
2. *Festlegung des Signifikanzniveaus:* $\alpha = 0,10$
3. *Auswahl des Verfahrens:* $F = \frac{s_1^2}{s_2^2}; s_1^2 > s_2^2; df_1 = n_1 - 1; df_2 = n_2 - 1$
4. *Stichprobenwerte:*

$s_1 = 12$	$s_2 = 7$
$n_1 = 10$	$n_2 = 30$
5. *Berechnung der Prüfgröße:* $F = \frac{12^2}{7^2} = 2,94$
6. *Ermittlung der kritischen Grenze:* $F_{1-\frac{\alpha}{2}; df_1; df_2} = F_{0,95; 9; 29} = 2,22$
7. *Entscheidung:* $2,94 > 2,22 \Rightarrow H_0$ wird abgelehnt

Die Hypothese, dass die beiden Stichproben aus varianzhomogenen Populationen entstammen, muss abgelehnt werden, was zur Konsequenz hat, dass die Voraussetzung zur Anwendung des klassischen t-Testes auf S. 44 nicht gegeben war.

5. Varianzanalysen

Bei der Besprechung der Signifikanztests zur Prüfung des Mittelwertunterschieds zweier Stichproben gingen wir von der Annahme aus, dass sich die beiden Gruppen nur durch das Wirken *eines* Faktors unterscheiden. Ergibt sich zwischen den beiden Stichprobenmittelwerten ein signifikanter Unterschied, so kann man sagen, dieser Faktor habe einen Einfluss auf das untersuchte Merkmal. Hat man jedoch mehr als zwei Mittelwerte zu vergleichen oder sind weitere Faktoren zu berücksichtigen, so können die im vorigen Kapitel beschriebenen Tests nicht verwendet werden. Wir brauchen daher Verfahren, bei denen die Problemstellung in zwei Richtungen verallgemeinert werden:

- Einfaktorische Varianzanalysen

Die Untersuchung der Signifikanz des Effektes eines Faktors erstreckt sich auf mehr als zwei Stufen.

Beispiel: Es soll die Effektivität von drei Therapierichtungen (Psychoanalyse, Verhaltenstherapie, Gesprächspsychotherapie) untersucht werden. Der Faktor Therapieart wird also in drei Stufen untersucht, es werden drei Stichproben von behandelten Klienten erhoben und der psychische Zustand von allen festgestellt. Es ergeben sich drei Mittelwerte, deren Unterschied auf Signifikanz geprüft werden soll.

- Mehrfaktorielle Varianzanalysen

Es sind mehrere Faktoren zu berücksichtigen, die die Verteilung der untersuchten Zufallsvariablen beeinflussen.

Beispiel: Das obige Experiment wird erweitert, indem der Einfluss der Gleich- bzw. Gegengeschlechtlichkeit der Therapeut-Klient-Beziehung auf den Heilungserfolg mit berücksichtigt wird. Die obigen drei Gruppen werden in sechs Gruppen aufgeteilt, je nachdem, ob die Therapeut-Klient-Beziehung gleich- oder gegengeschlechtlich sind. Es liegen nun zwei Faktoren vor: die Therapierichtung in den drei Stufen Psychoanalyse, Verhaltenstherapie und Gesprächspsychotherapie und der Faktor Geschlechtlichkeit der Therapeut-Klient-Beziehung in den beiden Stufen gleichgeschlechtlich und gegengeschlechtlich. Es ergeben sich verschiedene Mittelwerte, deren Unterschied auf Signifikanz geprüft werden soll.

Der Begriff der Varianzanalyse wurde erstmals 1918 von R. A. FISCHER in einer Arbeit über Fragen der Populationsgenetik eingeführt. Biologie, Landwirtschaft und Astronomie waren die ersten Disziplinen, in denen die Varianzanalyse praktisch angewandt wurde. In der Folgezeit erschien eine Flut von Lehrbüchern über die Varianzanalyse; in den Jahren 1931 bis 1933 wurden über 150 Veröffentlichungen registriert.

5.1. α -Adjustierung

Bevor die Varianzanalyse eingeführt wurde, wurde der Vergleich von mehr als zwei Mittelwerten bei unabhängigen Stichproben mit dem klassischen t-Test durchgeführt. Dabei wurde jede mögliche Kombination der Mittelwerte auf signifikanten Unterschied hin geprüft, um zu entscheiden, ob alle oder zwei der drei Mittelwerte signifikant voneinander verschieden sind.

Beispiel: Bei drei Mittelwerten \bar{X}_1 , \bar{X}_2 und \bar{X}_3 sind folgende statistische Hypothesen möglich:

$$\begin{array}{lll} H_0: \mu_1 = \mu_2 & H_0: \mu_1 = \mu_3 & H_0: \mu_2 = \mu_3 \\ H_1: \mu_1 \neq \mu_2 & H_1: \mu_1 \neq \mu_3 & H_1: \mu_2 \neq \mu_3 \end{array}$$

Die Wahrscheinlichkeit in einem Einzeltest für die Ablehnung der Nullhypothese entspricht der Irrtumswahrscheinlichkeit α . Die Gesamthypothese $H_0: \mu_1 = \mu_2 = \dots = \mu_J$, dass alle drei Mittelwerte sich nur zufällig voneinander unterscheiden, gilt als verworfen, wenn mindestens einer der m Einzeltests die Nullhypothese ablehnt. Die Wahrscheinlichkeit hierfür lautet nach Regel 6 im Skript von Herrn HUBER, STAT -7-:

$$\pi = 1 - (1 - \alpha)^m$$

Beispiel: Im obigen Fall mit $m=3$ Einzeltests und einem Signifikanzniveau von 5% ergibt sich die Wahrscheinlichkeit, dass in mindestens einem der Tests die Nullhypothese dieses Tests und damit die Gesamt-Nullhypothese verworfen wird, mit $1 - (1 - 0,05)^3 = 0,143$. In 14,3% aller Fälle ist zu erwarten, dass mind. ein Ergebnis der gezogenen Stichproben die Gesamt-Nullhypothese verwirft, obwohl sie in der Gesamtpopulation zutrifft.

Es ist offensichtlich, dass mit steigenden Anzahlen von Stichproben der α -Fehler ansteigt. Man spricht daher von α -Fehlerkumulierung. Um den α -Fehler auf dem Niveau der geplanten Signifikanz zu halten, muss daher eine α -Adjustierung durchgeführt werden. Es wird dabei das Signifikanzniveau α' für einen Einzeltest gesucht, das über die α -Kumulierung bewirkt, dass die Gesamtirrtumswahrscheinlichkeit π aller Einzeltests dem geplanten Signifikanzniveau α entspricht. Die Berechnung hierfür lautet:

$$\alpha' = 1 - (1 - \alpha)^{1/m}$$

Beispiel: Im Fall mit $m=3$ Einzeltests und einem Signifikanzniveau von $\alpha=0,05$ ergibt sich eine α -Adjustierung von $\alpha'=0,017$. Das Signifikanzniveau für jeden Einzeltest lautet also 1,7%. In die Formel zur Berechnung der Gesamtirrtumswahrscheinlichkeit π eingesetzt, ergibt sich ein Gesamtsignifikanzniveau von $\pi = 1 - (1 - \alpha')^m = 1 - (1 - 0,017)^3 = 0,05$.

5.2. Einfaktorielle Varianzanalyse

5.2.1. Unabhängige Stichproben

5.2.1.1. Verfahrensweise

Die einfaktorielle Varianzanalyse für unabhängige Stichproben wird im Prinzip nach dem gleichen Schema wie ein Signifikanztest durchgeführt. Die statistischen Hypothesen lauten:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_J \quad H_1: \mu_1 \neq \mu_2 \neq \dots \neq \mu_J$$

J = Anzahl der Stichproben

Die Nullhypothese besagt, dass die Stichproben aus Grundgesamtheiten mit demselben Erwartungswert μ entstammen, d.h. letztendlich aus dergleichen Grundgesamtheit. Der Unterschied zwischen den Mittelwerten der Stichproben ist nicht signifikant. Der untersuchte Faktor übt keinen Effekt auf die abhängige Variable (Messgröße) aus. Die Alternativhypothese besagt, dass die Stichproben aus verschiedenen Grundgesamtheiten mit unterschiedlichen Parametern μ entstammen. Der Unterschied zwischen den Mittelwerten der Stichproben ist signifikant. Der den Unterschied zwischen den Stichproben begründete Faktor übt einen Effekt auf die abhängige Variable aus. Dem Verfahren der Varianzanalyse liegt folgende allgemeine Überlegung zugrunde. Nimmt man die Streuung der Gesamtstichprobe, die aus allen Stichproben gebildet wird, so lässt sich diese nach dem Satz der Streuungszerlegung in die Streuung zwischen den Gruppen (SAQ_Z) und der Streuung innerhalb der Gruppen (SAQ_I) zerlegen (=Varianzanalyse):

$$SAQ_G = SAQ_Z + SAQ_I$$

$$\sum_j \sum_i (x_{ij} - \bar{x})^2 = \sum_j (\bar{x}_j - \bar{x})^2 \cdot n_j + \sum_j \sum_i (x_{ij} - \bar{x}_j)^2$$

- x_{ij} = Messwert Nr. i der Stichprobe Nr. j
- \bar{x} = Arithmetisches Mittel der Gesamtstichprobe
- \bar{x}_j = Arithmetisches Mittel der Stichprobe Nr. j
- n_j = Anzahl der Personen der Stichprobe Nr. j

Die Summe SAQ_G auf der linken Seite der Gleichung resultiert aus den Abweichungen aller Beobachtungswerte vom Gesamtmittel \bar{x} . Das zweite Glied auf der rechten Seite, die Streuung SAQ_I innerhalb der Gruppen, berücksichtigt die Abweichungen der Beobachtungswerte von dem jeweiligen Gruppenmittel. Die Streuung innerhalb der Gruppen lässt sich auf die überall wirkende Zufallsstreuung zurückführen, da innerhalb einer Gruppe vom Versuchsleiter alle möglichen Einflüsse konstant gehalten werden (sollen). Mit dem ersten Glied auf der rechten Seite ist die Streuung SAQ_Z zwischen den Gruppen erfasst, die sich aus den Abweichungen der Gruppenmittel vom Gesamtmittel ergibt. Die Streuung zwischen den Gruppen lässt sich daher nicht nur auf die Zufallsstreuung zurückführen, sondern auch auf den Einfluss des Faktors, der den Unterschied zwischen den Mittelwerten ausmacht. Aus dem SAQ_Z und dem SAQ_I werden nun sogenannte mittlere Quadrate gebildet, welche als eine Art Varianzen der SAQ aufgefasst werden können, indem die SAQ - entsprechend

der Berechnung der Varianz eines SAQ - durch die jeweiligen Freiheitsgrade geteilt werden:

$$MQ_Z = \frac{SAQ_Z}{J-1} \quad MQ_I = \frac{SAQ_I}{N-J}$$

J = Anzahl Stichproben; N = Anzahl Personen der Gesamtstichprobe

MQ_I ist nun als eine Schätzung für die Zufallsstreuung σ^2 aufzufassen, während MQ_Z die Schätzung einer Größe ist, die neben der Zufallsstreuung σ^2 einen Anteil enthält, der für den Einfluss des untersuchten Faktors steht. Dieser Faktor ist für den Unterschied der Mittelwerte der Stichproben verantwortlich und damit auch für die Vergrößerung der Streuung zwischen den Gruppen, was sich im MQ_Z niederschlägt. Daraus ergibt sich unmittelbar der Weg zur Prüfung der Nullhypothese: Ist $H_0: \mu_1 = \mu_2 = \dots = \mu_J$ richtig, so ist auch MQ_Z ebenfalls eine Schätzung der Zufallsstreuung σ^2 , da der Anteil des Faktors an der Streuung im MQ_Z zu Null wird. Damit werden sich MQ_I und MQ_Z „nur zufällig“ unterscheiden. Um die Nullhypothese zu prüfen, muss man folglich die Prüfgröße

$$F = \frac{MQ_Z}{MQ_I}$$

mit der kritischen Grenze der F-Verteilung mit dem Zählerfreiheitsgrad $df_1 = J-1$ und dem Nennerfreiheitsgrad $df_2 = N-J$ vergleichen. Die Hypothesen sind zwar ungerichtet, die kritische Grenze ist aber immer *einseitig*:

$$F_{1-\alpha; J-1; N-J}$$

Die Nullhypothese gilt als angenommen, wenn die Prüfgröße nicht die kritische Grenze überschreitet. Um die Berechnung der Prüfgröße durchzuführen, wird eine sogenannte Tafel der Varianzanalyse aufgestellt, wie sie im Skript von Herrn HUBER, STAT -10-, aufgeführt ist:

	SAQ	df	MQ	F
Zwischen	$\sum_j (\bar{x}_j - \bar{x})^2 \cdot n_j$	J-1	$\frac{SAQ_Z}{J-1}$	$\frac{MQ_Z}{MQ_I}$
Innerhalb	$\sum_j \sum_i (x_{ij} - \bar{x}_j)^2$	N-J	$\frac{SAQ_I}{N-J}$	
Gesamt	$\sum_j \sum_i (x_{ij} - \bar{x})^2$	N-1		

Die Zeile „Gesamt“ ist hierbei zur Ermittlung der Prüfgröße F nicht unbedingt erforderlich, aber nützlich zur Berechnung von SAQ_I bzw. SAQ_Z , wenn SAQ_G bekannt ist: $SAQ_I = SAQ_G - SAQ_Z$, $SAQ_Z = SAQ_G - SAQ_I$. Ist die Standardabweichung s_G der Gesamtgruppe bekannt, so berechnet sich das SAQ_G mit $s_G^2 \cdot (N-1)$. Sind hingegen die einzelnen Standardabweichungen s_j der Stichproben bekannt, so berechnen sich die SAQ_j der jeweiligen Stichproben mit $s_j^2 \cdot (n_j - 1)$, die Summe der SAQ_j ergibt dann das SAQ_I . Das Gesamtmittel \bar{X} berechnet sich mit:

$$\bar{x} = \frac{\sum \bar{x}_j \cdot n_j}{\sum n_j}$$

**Beispiel einer einfaktoriellen Varianzanalyse
für unabhängige Stichproben**

Beispiel 1 von Herrn HUBER

1. *Statistische Hypothesen:* $H_0: \mu_1=\mu_2=\mu_3=\mu_4$ $H_1: \mu_1\neq\mu_2\neq\mu_3\neq\mu_4$
2. *Festlegung des Signifikanzniveaus:* $\alpha = 0,01$
3. *Auswahl des Verfahrens:* $F = \frac{MQ_Z}{MQ_I}$; $df_1 = J-1$; $df_2 = N-J$
- Voraussetzungen:
- normalverteilte Grundgesamtheiten
 - varianzhomogene Grundgesamtheiten
4. *Stichprobenwerte:*

Methode	1	2	3	4
Messwerte	2 1 3 3 1	3 4 3 5 0	6 8 7 4 10	5 5 5 3 2
Σ	10	15	35	20
\bar{x}_j	2	3	7	4
SAQ _j	4	14	20	8

$$\bar{x}_1 = \frac{\sum x_{i1}}{n_1} = \frac{10}{5} = 2$$

$$\bar{x}_2 = \frac{\sum x_{i2}}{n_2} = \frac{15}{5} = 3$$

$$\bar{x}_3 = \frac{\sum x_{i3}}{n_3} = \frac{35}{5} = 7$$

$$\bar{x}_4 = \frac{\sum x_{i4}}{n_4} = \frac{20}{5} = 4$$

$$\bar{x} = \frac{\sum \bar{x}_j \cdot n_j}{\sum n_j} = \frac{2 \cdot 5 + 3 \cdot 5 + 7 \cdot 5 + 4 \cdot 5}{5 + 5 + 5 + 5} = 4$$

$$SAQ_Z = \sum_j (\bar{x}_j - \bar{x})^2 \cdot n_j = (2-4)^2 \cdot 5 + (3-4)^2 \cdot 5 + (7-4)^2 \cdot 5 + (4-4)^2 \cdot 5 = 70$$

$$SAQ_I = \sum_j \sum_i (x_{ij} - \bar{x}_j)^2 = \sum SAQ_j = 4 + 14 + 20 + 8 = 46$$

$$SAQ_1 = \sum (x_{i1} - \bar{x}_1)^2 = (2-2)^2 + \dots + (1-2)^2 = 4$$

$$SAQ_2 = \sum (x_{i2} - \bar{x}_2)^2 = (3-3)^2 + \dots + (0-2)^2 = 14$$

$$SAQ_3 = \sum (x_{i3} - \bar{x}_3)^2 = (6-7)^2 + \dots + (10-7)^2 = 20$$

$$SAQ_4 = \sum (x_{i4} - \bar{x}_4)^2 = (5-4)^2 + \dots + (2-4)^2 = 8$$

5. *Tafel der Varianzanalyse:*

	SAQ	df	MQ	F
Zwischen	70	3	23,33	8,10
Innerhalb	46	16	2,88	

6. *Ermittlung der kritischen Grenze:* $F_{1-\alpha; J-1; N-J} = F_{1-0,01; 4-1; 20-4} = F_{0,99; 3; 16} = 5,29$

7. *Entscheidung:* $8,10 > 5,29 \Rightarrow H_0$ wird abgelehnt

Damit ist mit einer Irrtumswahrscheinlichkeit von 1% die Hypothese angenommen, dass die unterschiedlichen Lehrmethoden unterschiedlichen Lernerfolg bei den Schülern bewirken.

5.2.1.2. Varianzaufklärung und Effektstärke durch Eta-Quadrat

Wie bereits oben erläutert, enthält die Quadratsumme SAQ_Z die Streuung der Messwerte, die auf den Einfluss des untersuchten Faktors zurückzuführen ist. Wenn wir also Eta-Quadrat berechnen, indem wir das SAQ_Z durch das SAQ_G dividieren, erhalten wir den relativen Anteil der Streuung, die auf den unterschiedlichen Stufen des Faktors beruht, an der Gesamtstreuung. Eta-Quadrat berechnet sich mit:

$$\eta^2 = \frac{SAQ_Z}{SAQ_G}$$

Beispiel: In obigen Beispiel 1 berechneten wir ein SAQ_Z von 70 und ein SAQ_I von 46, so dass sich ein SAQ_G von $SAQ_Z + SAQ_I = 116$ ergibt. Eta-Quadrat erhalten wir daher mit $\eta^2 = SAQ_Z/SAQ_G = 70/116 = 0,60$. Das bedeutet, dass 60% der Gesamtvarianz, der Streuung aller Messwerte, auf den Effekt des Faktors 'Lehrmethode' zurückzuführen ist, der Rest von 40% auf andere (zufällige) Faktoren. Damit kann man die Stärke des Effekts speziell in dieser Stichprobe angeben.

5.2.1.3. Multipler Mittelwertvergleich

Die Varianzanalyse gestattet nur den globalen Vergleich der Mittelwerte *aller* Stichproben. Dennoch möchte man wissen, welche Differenzen zwischen zwei Mittelwerten signifikant sind und welche nicht. Wir behandeln im folgenden den SCHEFFÉ-Test, der sich gegenüber Verletzungen von Voraussetzungen als relativ robust erwiesen hat und sich zudem tendenziell eher konservativ für die Nullhypothese entscheidet. Der SCHEFFÉ-Test garantiert weiterhin, dass die Wahrscheinlichkeit des α -Fehlers eines Einzelvergleichs nicht größer ist als das Signifikanzniveau α für den Gesamttest der Varianzanalyse (s. α -Adjustierung, S. 47). Der SCHEFFÉ-Test gilt als Erweiterung der Varianzanalyse, der mit durchgeführt wird. Die Differenz zwischen zwei Mittelwerten ist auf dem für die Varianzanalyse spezifizierten Signifikanzniveau signifikant, wenn ihr Absolutbetrag mindestens genauso groß ist wie die nach folgendem Ausdruck ermittelte kritische Grenze:

$$\sqrt{\left(\frac{2}{n}\right) \cdot (J-1) \cdot MQ_1 \cdot F_{1-\alpha;df_1;df_2}}$$

- n = Anzahl der Personen einer Stichprobe
- J = Anzahl der Stichproben
- MQ_1 = Mittleres Quadrat aus der Varianzanalyse
- $F_{1-\alpha;...}$ = F-Grenze aus der Varianzanalyse

Allerdings gilt diese Formel nur für den Fall gleicher Stichprobenumfänge n. Für unterschiedliche Stichprobengrößen gibt es eine modifizierte Form, die hier nicht weiter behandelt wird.

Beispiel eines multiplen Mittelwertvergleichs

Angaben: Zum Beispiel 1 der einfaktoriellen Varianzanalyse von unabhängigen Stichproben (s.S. 50) wird ein multipler Mittelwertvergleich nach dem Verfahren von SCHEFFÉ durchgeführt.

1. Mittelwertdifferenzen:

	$\bar{X}_1=2$	$\bar{X}_2=3$	$\bar{X}_3=7$	$\bar{X}_4=4$
$\bar{X}_1=2$	0	-1	-5	-2
$\bar{X}_2=3$		0	-4	-1
$\bar{X}_3=7$			0	3
$\bar{X}_4=4$				0

2. Kritische Grenze:

$$\sqrt{\left(\frac{2}{5}\right) \cdot (4-1) \cdot 2,88 \cdot 5,29} = 4,28$$

3. Multipler Mittelwertvergleich:

Differenz	Absolutbetrag	Differenz ist
$\bar{X}_1 - \bar{X}_2$	1	n.s.
$\bar{X}_1 - \bar{X}_3$	5	s.
$\bar{X}_1 - \bar{X}_4$	2	n.s.
$\bar{X}_2 - \bar{X}_3$	4	n.s.
$\bar{X}_2 - \bar{X}_4$	1	n.s.
$\bar{X}_3 - \bar{X}_4$	3	n.s.

Der SCHEFFÉ-Test hat auf dem Globalsignifikanzniveau von 1% der vorangegangenen Varianzanalyse ergeben, dass nur die Mittelwerte \bar{X}_1 und \bar{X}_3 signifikant voneinander verschieden sind. Es besteht ein Unterschied zwischen Lehrmethode 1 und 3 hinsichtlich ihrer Auswirkung auf den Lernerfolg. Zwischen den anderen Lehrmethoden wurde kein statistischer Unterschied im Hinblick auf den Lernerfolg festgestellt.

5.2.2. Abhängige Stichproben

Eine sehr vielseitig einsetzbare Versuchsanordnung sieht vor, dass von jeder Versuchsperson nicht nur eine oder zwei, sondern mehrere Messungen erhoben werden. Messwiederholungen werden in allen Versuchsanordnungen benötigt, wo die Veränderung eines Merkmals über die Zeit untersucht wird: in der Therapieforschung, um die Auswirkungen einer Behandlung durch Untersuchungen vor, während und nach der Therapie zu ermitteln; in der Gedächtnisforschung, um den Erinnerungsverlauf erworbener Lerninhalte zu überprüfen; in der Einstellungsforschung, um die Veränderung von Einstellungen durch Medieneinwirkung zu erkunden; oder in der Wahrnehmungspsychologie, um mögliche Veränderungen in der Bewertung von Kunstprodukten nach mehrmaligem Betrachten herauszufinden.

Eine weitere Indikation der Varianzanalyse für abhängige Stichproben liegt vor, wenn die Stichproben aufgrund eines Kontrollkriteriums parallelisiert wurden. D.h., dass aufgrund der Ergebnisse eines Vortestes, das die Ausprägung des zu kontrollierenden Merkmals erfasst, die Versuchspersonen zu Blöcken mit gleicher oder ähnlicher Ausprägung des getesteten Merkmals zusammengefasst werden, um sie gleichmäßig auf die Stichproben verteilen zu können. Die Größe eines Blockes entspricht der Anzahl der Stichproben, die Personen eines Blockes nennt man Zwillinge, homogene Tripel (3 gleiche P.), homogene Quadrupel (4 gleiche P.) etc. Sie sind hinsichtlich des zu kontrollierenden Merkmals gleich. Ein Merkmal wird auf diese Weise kontrolliert, wenn eine ungleiche Verteilung eines Merkmals in den Stichproben zu störenden Nebeneffekten führt, die die Messergebnisse verfälschen. Auch die zusammengehörigen Messwerte einer Versuchsperson bei Experimenten mit Messwiederholung nennt man aufgrund der Analogie einen Block.

5.2.2.1. Verfahrensweise

Werden n Versuchspersonen unter J Faktorstufen wiederholt beobachtet, ergibt sich das im folgenden dargestellte Datenschema:

Faktorstufe Nr.	1	2	...	J	Σ	$\bar{X}_{i\cdot}$
1	x_{11}	x_{12}	...	x_{1J}	Σx_{1j}	$\bar{X}_{1\cdot}$
2	x_{21}	x_{22}	...	x_{2J}	Σx_{2j}	$\bar{X}_{2\cdot}$
Block Nr.
n	x_{n1}	x_{n2}	...	x_{nJ}	Σx_{nj}	$\bar{X}_{J\cdot}$
Σ	Σx_{i1}	Σx_{i2}	...	Σx_{iJ}	Σx_{ij}	
$\bar{X}_{\cdot j}$	$\bar{X}_{\cdot 1}$	$\bar{X}_{\cdot 2}$...	$\bar{X}_{\cdot J}$		\bar{X}

In einer Zeile ergeben sich die Messwerte x_{ij} zu einer Vp bzw. von J parallelisierten Personen, Block genannt, in einer Spalte die Messwerte x_{ij} zu einer Faktorstufe, Treatment genannt. Mit dem ersten Index i werden die Zeilen bzw. Vpn einer Faktorstufe von 1 bis n , mit dem Index j die Spalten bzw. Faktorstufen einer Vp von 1 bis J durchnummeriert. In der letzten Zeile werden die Durchschnitte $\bar{X}_{i\cdot}$ der Messwerte der Faktorstufen und in der letzten Spalte die Durchschnitte $\bar{X}_{\cdot j}$ der Messwerte der Blöcke dargestellt.

Beispiel: In einer Untersuchung wird ausgezählt, wie viele fehlerfreie Additionen von jeweils 2 einstelligen Zahlen eine Vp pro Minute schafft. Lässt man eine Vp z.B. 30 min hintereinander Zahlen addieren, erhält man pro Minute einen bzw. insgesamt 30 Werte. Diese 30 Werte bilden eine Zeile im obigen Datenschema. Werden mehrere Vpn untersucht, ergibt sich das vollständige Schema.

Wie in der einfaktoriellen Varianzanalyse für unabhängige Stichproben lautet die Nullhypothese, dass mindestens zwei Mittelwerte signifikant verschieden sind: $H_0: \mu_1 = \mu_2 = \dots = \mu_J$; die Alternativhypothese entsprechend: $H_1: \mu_1 \neq \mu_2 \neq \dots \neq \mu_J$.

Beispiel: Bezogen auf den obigen Fall würde die H_0 inhaltlich besagen, dass sich die Rechengenauigkeit der Vpn während der einförmigen Dauerbelastungsaufgabe nicht ändert, die Alternativhypothese entsprechend umgekehrt.

Das SAQ_G wird im Fall abhängiger Stichproben zerlegt in die Streuung zwischen den Faktorstufen (SAQ_T , T wie engl. Treatment=Faktor), der Streuung zwischen den Blöcken (SAQ_B , B wie engl. block) und einer Reststreuung (SAQ_R , R wie engl. rest).

$$SAQ_G = SAQ_T + SAQ_B + SAQ_R$$

$$\sum_j \sum_i (x_{ij} - \bar{x})^2 = n \cdot \sum_j (\bar{x}_{\cdot j} - \bar{x})^2 + J \cdot \sum_i (\bar{x}_{i\cdot} - \bar{x})^2 + \sum_j \sum_i (x_{ij} - \bar{x}_{\cdot j} - \bar{x}_{i\cdot} + \bar{x})^2$$

- x_{ij} = Messwert Nr. i der Faktorstufe Nr. j
- \bar{x} = Arithmetisches Mittel aller Messwerte
- $\bar{x}_{i\cdot}$ = Arithmetisches Mittel des Blocks/Zeile Nr. i
- $\bar{x}_{\cdot j}$ = Arithmetisches Mittel der Faktorstufe/Stichprobe/Spalte Nr. j
- n = Anzahl der Blöcke/Zeilen
- J = Anzahl der Faktorstufen/Spalten

Die Summe SAQ_G auf der linken Seite der Gleichung resultiert aus den Abweichungen aller Messwerte vom Gesamtmittel \bar{x} . Das erste Glied auf der rechten Seite, die

Streuung SAQ_T zwischen den Faktorstufen, berücksichtigt die Abweichungen der Faktorstufenmittel $\bar{x}_{\cdot j}$ vom Gesamtmittel \bar{x} . Das SAQ_T entspricht dem SAQ_Z zwischen den Stichproben in der Varianzanalyse für unabhängige Stichproben und enthält ebenfalls neben der Zufallsstreuung einen Anteil, der auf die Wirkung des untersuchten Faktor zurückzuführen ist, wenn er wirksam ist. (Da alle Stichproben gleich groß sind, kann man in der Formel die Stichprobengrößen n_j als n vor das Summenzeichen ziehen.) Das zweite Glied auf der rechten Seite, die Streuung SAQ_B zwischen den Blöcken, berücksichtigt die Abweichungen der Blockmittel $\bar{x}_{i\cdot}$ vom Gesamtmittel. Das dritte Glied ist die Reststreuung SAQ_R , die sich auf die Zufallsstreuung zurückführen lässt und zu der das SAQ_T in der Berechnung der Prüfgröße in Beziehung zu setzen ist, um einen signifikanten Effekt annehmen zu können.

Das SAQ_I der Varianzanalyse für unabhängige Stichproben entspricht hier der Summe des SAQ_B und des SAQ_R . Das bisherige Verfahren ist nicht mehr anwendbar, da das SAQ_I bei abhängigen Stichproben nicht nur die Zufallsstreuung, sondern auch die interindividuellen Unterschiede zwischen den Blöcken enthält. Der Umstand, dass die Person eines Blockes die Messwerte seines Blockes aufgrund seiner individuellen Eigenschaften gleich beeinflusst, verändert die Streuung innerhalb der Stichproben. Es ist also das SAQ_R anstelle von SAQ_I für die Prüfgröße zu benutzen. Das SAQ_R lässt sich einfacher als oben in der Gesamtgleichung angegeben über folgende Gleichung bestimmen:

$$SAQ_R = SAQ_G - SAQ_T - SAQ_B$$

Bei der Varianzanalyse für abhängige Stichproben berechnet man mittlere Quadrate aus dem SAQ_T und dem SAQ_R :

$$MQ_T = \frac{SAQ_T}{J-1} \quad MQ_R = \frac{SAQ_R}{(J-1) \cdot (n-1)}$$

J = Anzahl Stichproben; n = Anzahl Personen einer Faktorstufe

Wie bei der Varianzanalyse für unabhängige Stichproben ergibt sich die Methode zur Prüfung der Nullhypothese unmittelbar aus dem Vergleich der durch die mittleren Quadrate geschätzten Größen. Ist $H_0: \mu_1 = \mu_2 = \dots = \mu_J$ richtig, so ist sowohl MQ_R als auch MQ_T eine Schätzung der Zufallsstreuung σ^2 , da der Anteil des Faktors an der Streuung im MQ_T zu Null wird. Damit werden sich MQ_I und MQ_R „nur zufällig“ unterscheiden. Um die Nullhypothese zu prüfen, muss man folglich die Prüfgröße

$$F = \frac{MQ_T}{MQ_R}$$

mit der kritischen Grenze der F-Verteilung mit dem Zählerfreiheitsgrad $df_1 = J-1$ und dem Nennerfreiheitsgrad $df_2 = (J-1) \cdot (n-1)$ vergleichen. Auch hier gilt, dass die Hypothesen zwar ungerichtet sind, die kritische Grenze aber immer *einseitig* zu bestimmen ist. Allerdings kommt jetzt ein verkomplizierender Umstand hinzu: Die genannten Freiheitsgrade gelten nur für bestimmte Voraussetzungen, die komplexer Natur sind und daher hier nicht weiter behandelt werden. Wenn diese Voraussetzungen nicht erfüllt sind, müssen die Freiheitsgrade durch einen Korrekturfaktor ε verkleinert werden:

$$df_1 = (J-1) \cdot \varepsilon \quad df_2 = (J-1) \cdot (n-1) \cdot \varepsilon$$

Für diesen Korrekturfaktor ε können folgende Grenzen angegeben werden:

$$\frac{1}{J-1} \leq \varepsilon \leq 1$$

Im Falle von $J=2$ Stichproben ist daher die Sachlage einfach: ε ist genau Eins und die genannten Freiheitsgrade können unkorrigiert verwendet werden. Im Fall von $J>2$ Stichproben allerdings ist es kompliziert, ε zu bestimmen, so dass wir die Berechnung dem Computer überlassen. Für Haus- und Klausuraufgaben greifen wir zu folgender Vorgehensweise: Da ε in einem bestimmten Bereich liegt, liegt auch die kritische F-Grenze in einem bestimmten Bereich. Wir wollen mal nachrechnen, welche Werte die korrigierten Freiheitsgrade in den genannten Extremfällen annehmen können:

- $\varepsilon = \frac{1}{J-1} \Rightarrow df_1 = (J-1) \cdot \varepsilon = \frac{J-1}{J-1} = 1$
 $df_2 = (J-1) \cdot (n-1) \cdot \varepsilon = \frac{(J-1) \cdot (n-1)}{J-1} = n-1$
- $\varepsilon = 1 \Rightarrow df_1 = J-1$
 $df_2 = (J-1) \cdot (n-1)$

Die gesuchte kritische Grenze $F_{1-\alpha}$ liegt dann also im kritischen Bereich von:

$$F_{1-\alpha; J-1; (J-1) \cdot (n-1)} \leq F_{1-\alpha} \leq F_{1-\alpha; 1; n-1}$$

Da wir nicht wissen, wo in diesem Bereich die kritische Grenze letztendlich liegt, befinden wir uns auf der sicheren Seite, wenn die Prüfgröße entweder unterhalb oder oberhalb des kritischen Bereichs liegt. Liegt die Prüfgröße innerhalb des kritischen Bereiches, können wir ohne EDV-Unterstützung keine Aussage treffen. Die Nullhypothese gilt als angenommen, wenn die Prüfgröße nicht den kritischen Bereich, also gleichzeitig beide oben genannten kritischen Grenzen, überschreitet:

$$F < (F_{1-\alpha; J-1; (J-1) \cdot (n-1)} \text{ bis } F_{1-\alpha; 1; n-1})$$

Zur Berechnung der Prüfgröße wird eine Tafel der Varianzanalyse nach Herrn HUBER, STAT -10-, aufgestellt:

	SAQ	df	MQ	F
Treatment	$n \cdot \sum_j (\bar{x}_{\cdot j} - \bar{x})^2$	$J-1$	$\frac{SAQ_T}{J-1}$	$\frac{MQ_T}{MQ_R}$
Blöcke	$J \cdot \sum_i (\bar{x}_{i\cdot} - \bar{x})^2$	$n-1$		
Rest	$SAQ_G - SAQ_T - SAQ_B$	$(J-1) \cdot (n-1)$	$\frac{SAQ_R}{(J-1) \cdot (n-1)}$	
Gesamt	$\sum_j \sum_i (x_{ij} - \bar{x})^2$	$J \cdot n - 1$		

Das Gesamtmittel \bar{x} berechnet sich durch die Summe aller Messwerte, dividiert durch die Anzahl aller Messwerte. Das Blockmittel $\bar{x}_{i\cdot}$ wird durch die Summe der Messwerte einer Zeile, dividiert durch die Anzahl der Messwerte einer Zeile, bestimmt. Das Faktorstufenmittel $\bar{x}_{\cdot j}$ ergibt sich aus der Summe der Messwerte einer Spalte, dividiert durch die Anzahl der Messwerte einer Spalte:

$$\bar{x} = \frac{\sum_j \sum_i x_{ij}}{J \cdot n} \quad \bar{x}_{i\cdot} = \frac{\sum_j x_{ij}}{J} \quad \bar{x}_{\cdot j} = \frac{\sum_i x_{ij}}{n}$$

Beispiel einer einfaktoriellen Varianzanalyse für abhängige Stichproben

Angaben: 6 Schüler bekommen einen zweiteiligen Förderunterricht, wobei die Hypothese aufgestellt wird, dass der Förderunterricht eine Leistungsänderung bei den Schülern bewirkt. Es wird die Leistung der Schüler sowohl vor dem Unterricht, nach dem ersten Teil des Unterrichts sowie nach Abschluss des Förderunterrichts in der Form von Punkten ermittelt. Signifikanzniveau: 5%.

1. *Statistische Hypothesen:* $H_0: \mu_1 = \mu_2 = \mu_3$ $H_1: \mu_1 \neq \mu_2 \neq \mu_3$
2. *Festlegung des Signifikanzniveaus:* $\alpha = 0,05$
3. *Auswahl des Verfahrens:* $F = \frac{MQ_T}{MQ_R}$; $df_1 = (J-1) \cdot \varepsilon$; $df_2 = (J-1) \cdot (n-1) \cdot \varepsilon$
 Voraussetzungen:
 - normalverteilte Grundgesamtheiten
 - varianzhomogene Grundgesamtheiten
4. *Stichprobenwerte:*

Block Nr.	Messzeitpunkt Nr.			Σ	$\bar{X}_{i\cdot}$
	1	2	3		
1	0	7	17	24	8
2	5	14	20	39	13
3	1	8	18	27	9
4	0	8	13	21	7
5	3	12	21	36	12
6	3	11	19	33	11
Σ	12	60	108	180	
$\bar{X}_{\cdot j}$	2	10	18		10

$$SAQ_T = n \cdot \sum_j (\bar{x}_{\cdot j} - \bar{x})^2$$

$$= 6 \cdot [(2 - 10)^2 + (10 - 10)^2 + (18 - 10)^2] = 768$$

$$SAQ_B = J \cdot \sum_i (\bar{x}_{i\cdot} - \bar{x})^2$$

$$= 3 \cdot [(8 - 10)^2 + (13 - 10)^2 + \dots + (12 - 10)^2 + (11 - 10)^2] = 84$$

$$SAQ_G = \sum_j \sum_i (x_{ij} - \bar{x})^2 = 866$$

$$= (0 - 10)^2 + (7 - 10)^2 + (17 - 10)^2$$

$$+ (5 - 10)^2 + (14 - 10)^2 + (20 - 10)^2$$

$$+ \dots$$

$$+ (3 - 10)^2 + (11 - 10)^2 + (19 - 10)^2$$

$$SAQ_R = SAQ_G - SAQ_T - SAQ_B = 866 - 768 - 84 = 14$$

5. Tafel der Varianzanalyse:

	SAQ	df	MQ	F
Treatment	768	2	384	274,3
Block	84	5		
Rest	14	10	1,4	
Gesamt	866			

6. Ermittlung des kritischen Bereichs: $F_{1-\alpha; J-1; (J-1)(n-1)} = F_{0,95; 2; 10} = 4,10$
 $F_{1-\alpha; 1; n-1} = F_{0,95; 1; 5} = 6,61$

7. Entscheidung: $274,3 > (4,10 \text{ bis } 6,61)$
 $\Rightarrow H_0$ wird abgelehnt

Damit ist mit einer Irrtumswahrscheinlichkeit von 5% die erste Hypothese angenommen, dass die unterschiedlichen Lehrmethoden unterschiedlichen Lernerfolg bei den Schülern bewirken.

5.3. Mehrfaktorielle Varianzanalyse

Wir wollen die Problemstellung noch einmal erweitern: Es ist der Einfluss mehrerer Faktoren auf eine abhängige Zufallsvariable X zu untersuchen, und das Experiment soll dabei so angelegt werden, dass auch eine eventuell vorhandene Wechselwirkung zwischen den Faktoren erfasst werden kann. Dass mehrere Faktoren auf einmal geprüft werden, kann zwei Gründe haben:

- Viele Forscher möchten die Effekte mehrerer unabhängiger Variablen, die aufgrund mehrerer theoretisch-inhaltlicher Hypothesen die abhängige Variable beeinflussen sollen, zugleich auf Signifikanz prüfen. Darüber hinaus bietet die mehrfaktorielle Varianzanalyse im Gegensatz zur einfaktoriellen die Möglichkeit, Effekte zu prüfen, die sich aus der Kombination mehrerer unabhängiger Variablen ergeben. Der Effekt einer solchen Kombination wird Interaktion bzw. Wechselwirkung genannt. So kann beispielsweise ein bestimmter Faktor bei Frauen anders als bei Männern wirken.
- Führt eine einfaktorielle Varianzanalyse zu keinem Ergebnis, so kann dies auf folgende Ursachen zurückgeführt werden: Entweder übt der Faktor tatsächlich keinen Einfluss aus (resultiert in einem zu kleinem SAQ_Z gegenüber SAQ_I) oder die Zufallsstreuung ist im Vergleich zum Effekt des Faktors zu groß (resultiert in einem zu großem SAQ_I gegenüber SAQ_Z). Das SAQ_Z kann und sollte versuchs-technisch nicht verändert werden, anders sieht es aber im Fall der Zufallsstreuung SAQ_I aus: Diese entsteht unter anderem auch aus Fehlereinflüssen, die aus unkontrollierten Störvariablen hervorgehen. Wollen wir die Präzision einer Untersuchung verbessern, müssen wir dafür Sorge tragen, dass der Einfluss der Störvariablen möglichst klein gehalten wird. Hierfür bietet es sich an, neben Maßnahmen wie der Konstanthaltung oder der Parallelisierung störender Variablen die Störvariablen systematisch zu variieren. Wir gruppieren die V_{pn} nicht nur nach den Stufen der uns eigentlich interessierenden unabhängigen Variablen (Wirkfaktor), sondern zusätzlich nach Variablen (Störfaktoren), von denen wir annehmen, dass sie neben dem Faktor ebenfalls einen starken Einfluss auf die abhängige Variable

(Messgröße) ausüben. Der Effekt dieser Variablen wird auf diese Weise nicht nur aus der Zufallsstreuung herausgezogen, sondern kann zusätzlich auf seine statistische Bedeutsamkeit überprüft werden. Der Nachteil dieser die Zufallsstreuung reduzierenden Technik liegt darin, dass mit steigender Anzahl der kontrollierten Störfaktoren die Anzahl der benötigten Versuchspersonen rapide anwächst.

5.3.1. Verfahrensweise

Im folgenden werden wir nur die zweifaktorielle Varianzanalyse besprechen. Mit ihr überprüfen wir, wie eine abhängige Variable von zwei unabhängigen Variablen (Faktoren) beeinflusst wird. Den ersten Faktor bezeichnen wir mit A und den zweiten Faktor mit B. Für die Nummerierung des Faktors A vereinbaren wir den Index j und für die Nummerierung des Faktors B den Index k. Der Faktor A habe J Stufen mit den Bezeichnungen A_j , der Faktor B habe K Stufen mit den Bezeichnungen B_k . Insgesamt ergeben sich $J \cdot K$ Faktorstufenkombinationen. Jeder dieser Faktorstufenkombinationen wird eine Zufallsstichprobe des Umfanges n zugewiesen, so dass die Gesamtstichprobe aus $N = J \cdot K \cdot n$ Vpn besteht. Von jeder Vp wird die abhängige Variable erhoben. Die Messwerte x_{ijk} werden nach folgendem allgemeinen Datenschema angeordnet:

	B_1	B_2	...	B_K	$\bar{X}_{j\bullet}$	
A_1	x_{111}	x_{112}	...	x_{11K}	$\bar{X}_{1\bullet}$	
	x_{211}	x_{212}	...	x_{21K}		
	x_{311}	x_{312}	...	x_{31K}		\bar{X}_{1K}
		
	x_{n11}	x_{n12}	...	x_{n1K}		
A_2	x_{121}	x_{122}	...	x_{12K}	$\bar{X}_{2\bullet}$	
	x_{221}	x_{222}	...	x_{22K}		
	x_{321}	x_{322}	...	x_{32K}		\bar{X}_{2K}
		
	x_{n21}	x_{n22}	...	x_{n2K}		
...	
A_J	x_{1J1}	x_{1J2}	...	x_{1JK}	$\bar{X}_{J\bullet}$	
	x_{2J1}	x_{2J2}	...	x_{2JK}		
	x_{3J1}	x_{3J2}	...	x_{3JK}		\bar{X}_{JK}
		
	x_{nJ1}	x_{nJ2}	...	x_{nJK}		
$\bar{X}_{\bullet k}$	$\bar{X}_{\bullet 1}$	$\bar{X}_{\bullet 2}$...	$\bar{X}_{\bullet K}$	\bar{X}	

Die Messwerte x_{ijk} sind hier dreifach indiziert. Der erste Index i kennzeichnet die Nummer des Messwertes in einer Faktorstufenkombination. Der zweite Index j kennzeichnet die Nummer der Stufe des Faktors A und der dritte Index k die Nummer der Stufe des Faktors B, in deren Kombination der Messwert auftritt. Ausgehend von den Einzelmessungen x_{ijk} wird für jede Stichprobe (Faktorstufenkombination oder Zelle) der Zelldurchschnitt \bar{x}_{jk} berechnet. Ausgehend von den Einzelmessungen aller Stichproben der Stufe j des Faktors A wird der Zeilendurchschnitt $\bar{X}_{j\bullet}$ berechnet. Ausgehend von den Einzelmessungen aller Stichproben der Stufe k des Faktors B

wird der Spaltendurchschnitt $\bar{x}_{\cdot k}$ berechnet. Ausgehend von allen Messwerten wird das Gesamtmittel \bar{x} berechnet. Die Zeilen- und Spaltendurchschnitte sowie das Gesamtmittel lassen sich aber auch aus den Zellmittelwerten berechnen:

$$\bar{x}_{j\cdot} = \frac{\sum_k \bar{x}_{jk}}{K} \quad \bar{x}_{\cdot k} = \frac{\sum_j \bar{x}_{jk}}{J} \quad \bar{x} = \frac{\sum_j \bar{x}_{j\cdot}}{J} = \frac{\sum_k \bar{x}_{\cdot k}}{K} = \frac{\sum_j \sum_k \bar{x}_{jk}}{J \cdot K}$$

Beispiel: Wir sehen uns Beispiel 2 von Herrn HUBER an. Es soll die Wirkung von drei verschiedenen Darreichungsformen eines Medikamentes zur Behandlung von Depressionen untersucht werden. Weiterhin soll geklärt werden, ob sich das Medikament bei Männern und Frauen unterschiedlich auswirkt. Faktor A ist das Merkmal Geschlecht mit den $J=2$ Stufen männlich (A_1) und weiblich (A_2). Faktor B ist das Merkmal Darreichungsform mit den $K=3$ Stufen Placebo (B_1), einfache Dosis (B_2) und doppelte Dosis (B_3). Insgesamt gibt es $J \cdot K=6$ Faktorstufenkombinationen: männlich - Placebo (A_1B_1), weiblich - Placebo (A_2B_1), männlich - einfache Dosis (A_1B_2), weiblich - einfache Dosis (A_2B_2), männlich - doppelte Dosis (A_1B_3) und weiblich - doppelte Dosis (A_2B_3). Jeder Faktorstufenkombination werden $n=5$ Personen zugewiesen, es gibt also $J \cdot K \cdot n=30$ Versuchspersonen. Von jeder Versuchsperson wird ein Messwert x_{ijk} für die abhängige Variable Depressivität erhoben. Im ersten Datenschema sind die Messwerte der einzelnen Vpn in den einzelnen Faktorstufenkombinationen dargestellt. Der Messwert x_{321} ist der dritte Messwert der Zelle A_2B_1 (weiblich - Placebo). Im zweiten Datenschema sind die Zelldurchschnitte \bar{x}_{jk} der Messwerte in den einzelnen Faktorstufenkombinationen, die Zeilendurchschnitte $\bar{x}_{j\cdot}$ der Messwerte des Faktors A und die Spaltendurchschnitte $\bar{x}_{\cdot k}$ der Messwerte des Faktors B dargestellt. Der Zelldurchschnitt \bar{x}_{12} berechnet sich aus den Messwerten der Faktorstufenkombination A_1B_2 und beträgt 15,6, der Zeilendurchschnitt $\bar{x}_{1\cdot}$ betrifft die Messwerte der Faktorstufe A_1 und lautet 16,8 und der Spaltendurchschnitt $\bar{x}_{\cdot 2}$ bezieht sich auf die Messwerte der Faktorstufe B_2 und ist 16,6. Das Gesamtmittel beruht auf allen Messwerten und beträgt 16,9. Insgesamt lässt sich erkennen, dass die Zelldurchschnitte \bar{x}_{jk} verschieden voneinander sind. Lässt sich überprüfen, ob es signifikante Mittelwertunterschiede zwischen den Zellmitteln gibt, d.h. gibt es in diesem Experiment überhaupt irgendeinen Effekt? Für Faktor A lässt sich erkennen, dass der Zeilendurchschnitt $\bar{x}_{1\cdot}=16,8$ der Männer kleiner ist als der Zeilendurchschnitt $\bar{x}_{2\cdot}=17,0$ der Frauen. Übt Faktor A einen Effekt aus, so dass diese Durchschnitte signifikant verschieden voneinander sind, d.h. hat das Geschlecht einen Einfluss auf die Depressivität? Für Faktor B lässt sich erkennen, dass das Spaltenmittel $\bar{x}_{\cdot 1}=20,6$ für das Placebo höher ist als das Spaltenmittel $\bar{x}_{\cdot 2}=16,6$ für die einfache Dosis und dieses ist wiederum größer als der Spaltendurchschnitt $\bar{x}_{\cdot 3}=13,5$ für die doppelte Dosis. Sind diese Mittelwerte signifikant verschieden voneinander, so dass man annehmen kann, dass Faktor B einen Effekt auf die Depressivität ausübt? Hat das Medikament überhaupt eine Wirkung und welche Dosis ist wirksamer? Gibt es vielleicht eine Wechselwirkung zwischen den beiden Faktoren, d.h. wirkt das Medikament bei den Frauen anders als bei den Männern?

Die zweifaktorielle Varianzanalyse beantwortet im allgemeinen vier Fragen:

1. Gibt es überhaupt einen signifikanten Effekt?
2. Übt Faktor A einen signifikanten Effekt aus?
3. Übt Faktor B einen signifikanten Effekt aus?
4. Gibt es eine Wechselwirkung zwischen den beiden Faktoren, die einen signifikanten Effekt ausübt?

Im Prinzip wird für jede dieser Fragen eine eigenständige einfaktorielle Varianzanalyse durchgeführt, die bei den der jeweiligen Fragestellung entsprechenden Durchschnitten überprüft, ob diese sich signifikant voneinander unterscheiden.

5.3.1.1. Gesamteffekt

Ob es überhaupt einen signifikanten Effekt gibt, lässt sich daran überprüfen, ob sich die Zelldurchschnitte signifikant voneinander unterscheiden. Entsprechend lautet die Nullhypothese:

$$H_0: \sum_j \sum_k (\mu_{jk} - \mu)^2 = 0$$

Wenn die Erwartungswerte μ_{jk} der den Zellstichproben zugehörigen Grundgesamtheiten alle gleich sind, sind die Differenzen zum μ der alle Stichproben umfassenden Grundgesamtheit gleich Null und damit auch die Abweichungsquadrate und die Summe der Abweichungsquadrate: Die Zellmittel \bar{x}_{jk} sind also nicht signifikant voneinander verschieden. Um dies zu prüfen, wird eine einfaktorielle Varianzanalyse (für unabhängige Stichproben) durchgeführt, wobei die J·K Stichproben (Zellen) als J·K Stufen eines Faktors betrachtet werden. Dazu wird das MQ_Z zwischen allen Stichproben, das neben der Zufallsstreuung die Effekte der Faktoren enthält, und das MQ_I innerhalb aller Stichproben, das als Schätzwert für die Zufallsstreuung genommen wird, berechnet:

$$\begin{aligned} SAQ_Z &= n \cdot \sum_j \sum_k (\bar{x}_{jk} - \bar{x})^2 & MQ_Z &= \frac{SAQ_Z}{J \cdot K - 1} \\ SAQ_I &= \sum_j \sum_k \sum_i (x_{ijk} - \bar{x}_{jk})^2 & MQ_I &= \frac{SAQ_I}{J \cdot K \cdot (n - 1)} \end{aligned}$$

Wie bei der einfaktoriellen Varianzanalyse ergibt sich die Methode zur Prüfung der Nullhypothese unmittelbar aus dem Vergleich der durch die mittleren Quadrate geschätzten Größen. Ist die Nullhypothese richtig, so ist sowohl MQ_Z als auch MQ_I eine Schätzung der Zufallsstreuung σ^2 , da die Anteile der Faktoren an der Streuung im MQ_Z zu Null wird. Damit werden sich MQ_Z und MQ_I „nur zufällig“ unterscheiden. Um die Nullhypothese beizubehalten, muss folglich die Prüfgröße

$$F = \frac{MQ_Z}{MQ_I}$$

kleiner als die kritische Grenze der F-Verteilung mit dem Zählerfreiheitsgrad $df_1 = J \cdot K - 1$ und dem Nennerfreiheitsgrad $df_2 = J \cdot K \cdot (n - 1)$ sein.

Wird die Nullhypothese bestätigt, ist die zweifaktorielle Varianzanalyse hiermit beendet. Es gibt keinen signifikanten Effekt, sowohl von Faktor A als auch von Faktor B als auch von der Wechselwirkung zwischen beiden Faktoren nicht. Wird die Nullhypothese abgelehnt, sind folgende weitere Analysen notwendig, um den Sachverhalt näher aufzuklären.

5.3.1.2. Haupteffekt Faktor A

Ob Faktor A einen Haupteffekt auf die abhängige Variable ausübt, lässt sich daran prüfen, ob es einen signifikanten Mittelwertunterschied zwischen den Zeilenmitteln $\bar{x}_{j\cdot}$ gibt. Die Nullhypothese lautet:

$$H_0: \sum_j (\mu_{j\cdot} - \mu)^2 = 0$$

Hier besagt die Nullhypothese, dass die Zeilenmittel $\bar{x}_{j\cdot}$ nicht signifikant voneinander verschieden sind. Um dies zu prüfen, wird eine einfaktorielle Varianzanalyse durchgeführt, wobei die Stichproben der Zeilen zu größeren Stichproben zusammengefasst und als die der hier durchgeführten einfaktoriellen Varianzanalyse zugrundeliegenden Stichproben betrachtet werden. Es gibt also J Zeilenstichproben mit J Zeilenmittelwerten $\bar{x}_{j\cdot}$ zu den J Stufen des Faktors A. Zur Durchführung wird das MQ_Z zwischen den Zeilenstichproben, das neben der Zufallsstreuung den Effekt des Faktors A enthält und daher nun MQ_A genannt wird, berechnet:

$$SAQ_A = n \cdot K \cdot \sum_j (\bar{x}_{j\cdot} - \bar{x})^2 \quad MQ_A = \frac{SAQ_A}{J-1}$$

Ist die Nullhypothese richtig, so ist sowohl MQ_A als auch das vorher berechnete MQ_I eine Schätzung der Zufallsstreuung σ^2 , da der Anteil des Faktors A an der Streuung im MQ_A zu Null wird. Damit werden sich MQ_A und MQ_I „nur zufällig“ unterscheiden. Um die Nullhypothese beizubehalten, muss folglich die Prüfgröße

$$F = \frac{MQ_A}{MQ_I}$$

kleiner als die kritische Grenze der F-Verteilung mit dem Zählerfreiheitsgrad $df_1=J-1$ und dem Nennerfreiheitsgrad $df_2=J \cdot K \cdot (n-1)$ sein.

5.3.1.3. Haupteffekt Faktor B

Das gleiche gilt entsprechend für Faktor B: Ob dieser einen Haupteffekt auf die abhängige Variable ausübt, lässt sich daran prüfen, ob es einen signifikanten Mittelwertunterschied zwischen den Spaltendurchschnitten $\bar{x}_{\cdot k}$ gibt. Die Nullhypothese lautet also:

$$H_0: \sum_k (\mu_{\cdot k} - \mu)^2 = 0$$

Die Nullhypothese besagt, dass die Spaltenmittel $\bar{x}_{\cdot k}$ nicht signifikant voneinander verschieden sind. Nun werden die Stichproben der Spalten zu größeren Stichproben zusammengefasst und als die der hier durchgeführten einfaktoriellen Varianzanalyse zugrundeliegenden Stichproben betrachtet. Es gibt also K Spaltenstichproben mit K Spaltenmittelwerten $\bar{x}_{\cdot k}$ zu den K Stufen des Faktors B. Bei der Durchführung wird das MQ_Z zwischen den Spaltenstichproben, das neben der Zufallsstreuung den Effekt des Faktors B enthält und daher nun MQ_B genannt wird, berechnet:

$$SAQ_B = n \cdot J \cdot \sum_k (\bar{x}_{\cdot k} - \bar{x})^2 \quad MQ_B = \frac{SAQ_B}{K-1}$$

Ist die Nullhypothese richtig, so werden sich MQ_B und MQ_I „nur zufällig“ voneinander unterscheiden, wobei folglich die Prüfgröße

$$F = \frac{MQ_B}{MQ_I}$$

kleiner als die kritische Grenze der F-Verteilung mit dem Zählerfreiheitsgrad $df_1=K-1$ und dem Nennerfreiheitsgrad $df_2=J \cdot K \cdot (n-1)$ sein muss.

5.3.1.4. Wechselwirkung AxB

Ob eine Wechselwirkung zwischen den Faktoren A und B einen Effekt auf die abhängige Variable ausübt, lässt sich daran prüfen, ob es einen signifikanten Mittelwertunterschied in den Beziehungen zwischen den Zell-, Zeilen- und Spaltendurchschnitten und zum Gesamtdurchschnitt gibt, was hier nicht weiter erklärt wird. Die Nullhypothese lautet:

$$H_0: \sum_k (\mu_{jk} - \mu_{j\cdot} - \mu_{\cdot k} + \mu)^2 = 0$$

Die Nullhypothese besagt, dass die Wechselwirkung zwischen den Faktoren A und B keinen signifikanten Effekt auf die abhängige Variable ausübt. Um ein mittleres Quadrat MQ_{AxB} , das neben der Zufallsstreuung eine Streuung enthält, die aufgrund des Effekts einer Wechselwirkung zwischen den beiden Faktoren entsteht, zu erhalten, macht man sich folgende Gleichungen der Varianzzerlegung in einer zweifaktoriellen Varianzanalyse zunutze:

$$\begin{aligned} SAQ_G &= SAQ_Z + SAQ_I \\ SAQ_Z &= SAQ_A + SAQ_B + SAQ_{AxB} \end{aligned}$$

Das gesuchte mittlere Quadrat MQ_{AxB} berechnet sich dann wie folgt:

$$SAQ_{AxB} = SAQ_Z - SAQ_A - SAQ_B \quad MQ_{AxB} = \frac{SAQ_{AxB}}{(J-1) \cdot (K-1)}$$

Ist die Nullhypothese richtig, so werden sich MQ_{AxB} und MQ_I „nur zufällig“ voneinander unterscheiden, so dass folglich die Prüfgröße

$$F = \frac{MQ_{AxB}}{MQ_I}$$

kleiner als die kritische Grenze der F-Verteilung mit dem Zählerfreiheitsgrad $df_1=(J-1) \cdot (K-1)$ und dem Nennerfreiheitsgrad $df_2=J \cdot K \cdot (n-1)$ sein muss.

5.3.2. Tafel der Varianzanalyse

Um sich die Durchführung der zweifaktoriellen Varianzanalyse zu vereinfachen, wird entsprechend HUBER (STAT -11-) eine Tafel der Varianzanalyse aufgestellt:

	SAQ	df	MQ	F
Faktor A	$n \cdot K \cdot \sum_j (\bar{x}_{j\cdot} - \bar{x})^2$	J-1	$\frac{SAQ_A}{J-1}$	$\frac{MQ_A}{MQ_I}$
Faktor B	$n \cdot J \cdot \sum_k (\bar{x}_{\cdot k} - \bar{x})^2$	K-1	$\frac{SAQ_B}{K-1}$	$\frac{MQ_B}{MQ_I}$
Wechselw. AxB	$SAQ_Z - SAQ_A - SAQ_B$	(J-1)·(K-1)	$\frac{SAQ_{AxB}}{(J-1) \cdot (K-1)}$	$\frac{MQ_{AxB}}{MQ_I}$
Zwischen	$n \cdot \sum_j \sum_k (\bar{x}_{jk} - \bar{x})^2$	(J·K)-1	$\frac{SAQ_Z}{J \cdot K - 1}$	$\frac{MQ_Z}{MQ_I}$
Innerhalb	$\sum_j \sum_k \sum_i (x_{ijk} - \bar{x})^2$	J·K·(n-1)	$\frac{SAQ_I}{J \cdot K \cdot (n-1)}$	
Gesamt	$SAQ_Z + SAQ_I$	N-1		

Es empfiehlt sich, eine weitere Spalte mit den jeweiligen kritischen Grenzen anzulegen, um sich die Entscheidung für oder gegen einen signifikanten Effekt zu erleichtern. Bei allen Varianzanalysen gilt, dass die Hypothesen zwar ungerichtet sind, die kritische F-Grenze aber immer *einseitig* zu bestimmen ist.

5.3.3. Varianzaufklärung und Effektstärken

Ebenso wie in der einfaktoriellen Varianzanalyse kann man über η^2 für die Gesamtvarianz angeben, inwieweit diese auf die Effekte der Faktoren A und B zurückgeführt werden kann. Dies ist ein Maß für die Stärke der Faktoreffekte:

$$\eta^2 = \frac{SAQ_Z}{SAQ_G}$$

Weiterhin kann man ermitteln, wie sich diese Effektstärke auf die Faktoren A und B sowie die Wechselwirkung zwischen den Faktoren aufteilt:

$$\eta^2 = \eta_A^2 + \eta_B^2 + \eta_{AxB}^2$$

$$\eta_A^2 = \frac{SAQ_A}{SAQ_G} \quad \eta_B^2 = \frac{SAQ_B}{SAQ_G} \quad \eta_{AxB}^2 = \frac{SAQ_{AxB}}{SAQ_G}$$

Es lassen sich also die Effektstärken von den Faktoren A und B und von der Wechselwirkung zwischen den Faktoren einzeln berechnen, deren Summe die Gesamteffektstärke ergibt.

Beispiel einer zweifaktoriellen Varianzanalyse

Beispiel 2 von Herrn HUBER

\bar{x}_{jk}	B ₁	B ₂	B ₃	$\bar{x}_{j\cdot}$
A ₁	22,4	15,6	12,4	16,8
A ₂	18,8	17,6	14,6	17,0
$\bar{x}_{\cdot k}$	20,6	16,6	13,5	16,9

$$SAQ_1 = \sum_j \sum_k \sum_i (x_{ijk} - \bar{x}_{jk})^2 = \sum_k \sum_j SAQ_{jk}$$

$$= \left[\begin{array}{l} (x_{111} - \bar{x}_{11})^2 + \dots + (x_{511} - \bar{x}_{11})^2 + \\ (x_{112} - \bar{x}_{12})^2 + \dots + (x_{512} - \bar{x}_{12})^2 + \\ \vdots \\ (x_{123} - \bar{x}_{23})^2 + \dots + (x_{523} - \bar{x}_{23})^2 \end{array} \right]$$

$$= \left[\begin{array}{l} (22 - 22,4)^2 + \dots + (22 - 22,4)^2 + \\ (16 - 15,6)^2 + \dots + (15 - 15,6)^2 + \\ \vdots \\ (16 - 14,6)^2 + \dots + (14 - 14,6)^2 \end{array} \right] = 40,8$$

$$J = 2; K = 3$$

$$n = 5; N =$$

$$SAQ_z = n \cdot \sum_j \sum_k (\bar{x}_{jk} - \bar{x})^2$$

$$= 5 \cdot \left[\begin{array}{l} [(\bar{x}_{11} - \bar{x})^2 + (\bar{x}_{12} - \bar{x})^2 + (\bar{x}_{13} - \bar{x})^2] + \\ [(\bar{x}_{21} - \bar{x})^2 + (\bar{x}_{22} - \bar{x})^2 + (\bar{x}_{23} - \bar{x})^2] \end{array} \right]$$

$$= 5 \cdot \left[\begin{array}{l} [(22,4 - 16,9)^2 + (15,6 - 16,9)^2 + (12,4 - 16,9)^2] + \\ [(18,8 - 16,9)^2 + (17,6 - 16,9)^2 + (14,6 - 16,9)^2] \end{array} \right] = 307,9$$

$$\begin{aligned}
 \text{SAQ}_A &= n \cdot K \cdot \sum_j (\bar{x}_{j\cdot} - \bar{x})^2 \\
 &= 5 \cdot 3 \cdot \left[(\bar{x}_{1\cdot} - \bar{x})^2 + (\bar{x}_{2\cdot} - \bar{x})^2 \right] \\
 &= 5 \cdot 3 \cdot \left[(16,8 - 16,9)^2 + (17,0 - 16,9)^2 \right] = 0,3
 \end{aligned}$$

$$\begin{aligned}
 \text{SAQ}_B &= n \cdot J \cdot \sum_k (\bar{x}_{\cdot k} - \bar{x})^2 \\
 &= 5 \cdot 2 \cdot \left[(\bar{x}_{\cdot 1} - \bar{x})^2 + (\bar{x}_{\cdot 2} - \bar{x})^2 + (\bar{x}_{\cdot 3} - \bar{x})^2 \right] \\
 &= 5 \cdot 2 \cdot \left[(20,6 - 16,9)^2 + (16,6 - 16,9)^2 + (13,5 - 16,9)^2 \right] = 253,4
 \end{aligned}$$

$$\begin{aligned}
 \text{SAQ}_{A \times B} &= \text{SAQ}_Z - \text{SAQ}_A - \text{SAQ}_B \\
 &= 307,9 - 0,3 - 253,4 = 54,2
 \end{aligned}$$

Tafel der Varianzanalyse:

	SAQ	df	MQ	F	F _{0,95}	Ergebnis
A	0,3	1	0,3	0,18	4,26	n. s.
B	253,4	2	126,7	74,53	3,40	s.
AxB	54,2	2	27,1	15,95	3,40	s.
Z	307,9	5	61,6	36,22	2,62	s.
I	40,8	24	1,7			
G	348,7	29				

Gesamteffekt Z:

$$H_0: \sum_j \sum_k (\mu_{jk} - \mu)^2 = 0 \quad H_1: \sum_j \sum_k (\mu_{jk} - \mu)^2 > 0$$

Die Nullhypothese wird verworfen: Es konnte ein signifikanter Mittelwertunterschied zwischen den Zellmittelwerten festgestellt werden. Der Faktor A, der Faktor B und/oder die Wechselwirkung zwischen den Faktoren A und B haben einen signifikanten Effekt.

Haupteffekt Faktor A:

$$H_0: \sum_j (\mu_{j\cdot} - \mu)^2 = 0 \quad H_1: \sum_j (\mu_{j\cdot} - \mu)^2 > 0$$

Die Nullhypothese wird beibehalten: Es konnte kein signifikanter Mittelwertunterschied zwischen den Durchschnitten der Männer und der Frauen festgestellt werden. Der Faktor A, das Geschlecht, übt keinen Effekt auf die abhängige Variable Depressivität aus.

Haupteffekt Faktor B:

$$H_0: \sum_k (\mu_{\cdot k} - \mu)^2 = 0 \quad H_1: \sum_k (\mu_{\cdot k} - \mu)^2 > 0$$

Die Nullhypothese wird abgelehnt: Es konnte ein signifikanter Mittelwertunterschied zwischen den Durchschnitten der verschiedenen Darreichungsformen des Medikaments bestätigt werden. Der Faktor B, die Darreichungsform des Medikamentes, übt einen Effekt auf die abhängige Variable Depressivität aus.

Effekt der Wechselwirkung AxB:

$$H_0: \sum_k (\mu_{jk} - \mu_{j\cdot} - \mu_{\cdot k} + \mu)^2 = 0 \quad H_1: \sum_k (\mu_{jk} - \mu_{j\cdot} - \mu_{\cdot k} + \mu)^2 > 0$$

Die Nullhypothese wird nicht beibehalten: Es konnte ein Effekt der Wechselwirkung zwischen den Faktoren A und B auf die abhängige Variable Depressivität bestätigt werden. Männer scheinen anders auf verschiedene Dosen des Medikaments zu reagieren als Frauen.

Varianzaufklärung / Effektstärke:

$$\eta^2 = \frac{SAQ_Z}{SAQ_G} = \frac{307,9}{348,7} = 0,883$$

Die Gesamtvarianz kann zu 88,3% auf die Effekte der Faktoren A und B und der Wechselwirkung zwischen den Faktoren A und B erklärt werden. Dieser Gesamtwert teilt sich auf wie folgt:

$$\eta_A^2 = \frac{SAQ_A}{SAQ_G} = \frac{0,3}{348,7} = 0,001$$

Zu 0,1% kann die Gesamtvarianz auf den Einfluss des Geschlechts zurückgeführt werden.

$$\eta_B^2 = \frac{SAQ_B}{SAQ_G} = \frac{253,4}{348,7} = 0,727$$

Zu 72,7% kann die Gesamtvarianz durch den Einfluss der Darreichungsform des Medikaments erklärt werden.

$$\eta_{AxB}^2 = \frac{SAQ_{AxB}}{SAQ_G} = \frac{54,2}{348,7} = 0,155$$

Zu 15,5% beruht die Gesamtvarianz auf den Effekt der Wechselwirkung.

5.3.4. Interaktion zwischen den Faktoren

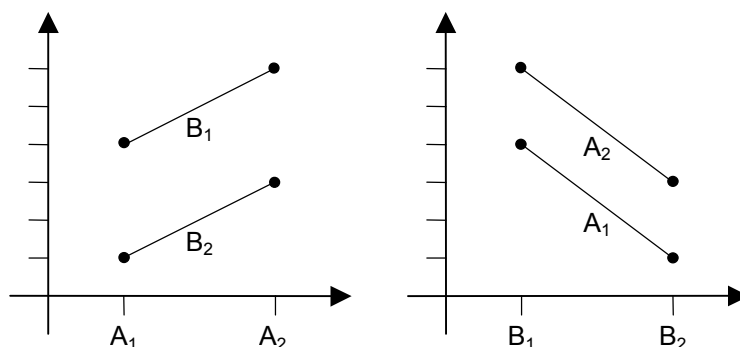
Die Wechselwirkung zwischen den Faktoren A und B wird auch Interaktion genannt und ist wie folgt definiert:

Der Effekt eines Faktors in einer einzelnen Stufe des anderen Faktors wird einfacher Haupteffekt genannt. Eine Wechselwirkung liegt vor, wenn die einfachen Haupteffekte eines der beiden Faktoren auf mindestens zwei Stufen des anderen Faktors signifikant unterschiedlich ist. Keine Wechselwirkung liegt also vor, wenn die einfachen Haupteffekte beider Faktoren auf allen Stufen des anderen Faktors gleich groß sind.

Wir wollen uns diesen Sachverhalt einmal grafisch in einem Interaktionsdiagramm verdeutlichen. Dazu werden die Zellmittelwerte in ein Koordinatenkreuz eingetragen, wobei auf der x-Achse die Stufen des einen Faktors, auf der y-Achse die Stufen des anderen Faktors eingetragen werden. Wichtig dabei ist, dass beide Diagrammmöglichkeiten betrachtet werden: Auf der x-Achse werden in einem Diagramm die Stufen des Faktors A aufgetragen, in dem anderem Diagramm die Stufen des Faktors B, auf der y-Achse die Messwertskala. Die Zellmittelwerte, die hinsichtlich des nicht auf der x-Achse dargestellten Faktors zu derselben Stufe gehören, verdeutlichen einen einfachen Haupteffekt und werden daher mit einer Linie verbunden. Als Beispiel nehmen wir an, Faktor A sei das Geschlecht (1-weiblich, 2-männlich), Faktor B ein blutdrucksenkendes Medikament (1-Placebo, 2-Wirkstoff). Weiterhin nehmen wir an, dass alle Mittelwertunterschiede signifikant seien, beide Faktoren zeigen einen Effekt auf den Blutdruck.

5.3.4.1. Nullinteraktion

	B ₁	B ₂	$\bar{x}_{j\cdot}$	$d_{j\cdot}$
A ₁	4	1	2,5	3
A ₂	6	3	4,5	3
$\bar{x}_{\cdot k}$	5	2		
$d_{\cdot k}$	-2	-2		



Die Abbildungen zeigen, dass die Linien jeweils den gleichen Trend aufweisen (links monoton steigend, rechts monoton fallend). Die Rangfolge der A-Stufen ist für B₁ und B₂ identisch (links), und die Rangfolge der B-Stufen ist für A₁ und A₂ identisch (rechts). Die Haupteffekte beider Faktoren A und B sind damit eindeutig interpretierbar: Die Rangfolge der Zeilenmittelwerte des Haupteffekts A ($\bar{x}_{1\cdot} < \bar{x}_{2\cdot}$) gilt für beide Stufen des Faktors B: $\bar{x}_{11} < \bar{x}_{21}$ (einfacher Haupteffekt A in Stufe B₁) und $\bar{x}_{12} < \bar{x}_{22}$ (einfacher Haupteffekt A in Stufe B₂), und die Rangfolge der Spaltenmittelwerte des Haupteffekts B ($\bar{x}_{\cdot 1} > \bar{x}_{\cdot 2}$) gilt für beide Stufen des Faktors A: $\bar{x}_{11} > \bar{x}_{12}$ (einfacher Haupteffekt B in Stufe A₁) und $\bar{x}_{21} > \bar{x}_{22}$ (einfacher Haupteffekt B in Stufe A₂). Frauen haben einen niedrigeren Blutdruck als Männer, unabhängig davon, ob ein Placebo oder ein blutdrucksenkendes Mittel gegeben wird (Faktor A). Das Medikament bewirkt einen niedrigeren Blutdruck als das Placebo, also eine Blutdrucksenkung, gleichgültig, ob es Männern oder Frauen verabreicht wird (Faktor B).

Die linke Abbildung zeigt die einfachen Haupteffekte von Faktor A auf den einzelnen Stufen des Faktors B: In der Placebo-Stufe (Linie B_1) zeigen die Frauen (Punkt A_1B_1) einen niedrigeren Blutdruck als die Männer (Punkt A_2B_1), die Differenz $\bar{d}_{\cdot 1}$ beträgt -2 , was sich in der wachsenden Steigung der Linie B_1 auswirkt. In der Wirkstoff-Stufe (Linie B_2) zeigt sich das gleiche Bild: Frauen (Punkt A_1B_2) haben einen niedrigeren Blutdruck als die Männer (Punkt A_2B_2), die Differenz $\bar{d}_{\cdot 2}$ ist auch hier -2 , was sich in der gleichen Steigung der Linie B_2 wie in Linie B_1 auswirkt. Die einfachen Haupteffekte des Geschlechts sind in beiden Stufen des Faktors B gleich, d.h. die Zellmittelwertdifferenzen sind in beiden Stufen gleich hoch, was sich in einer Parallelität der Linien auswirkt. (Insgesamt ist die Linie B_1 höher als die Linie B_2 . Das bedeutet, dass die Blutdruckwerte in der Placebo-Stufe insgesamt höher sind als in der Wirkstoff-Stufe.)

Die rechte Abbildung zeigt die einfachen Haupteffekte von Faktor B auf den einzelnen Stufen des Faktors A: In der Frauen-Stufe (Linie A_1) zeigt sich beim Placebo (Punkt B_1A_1) ein generell höherer Blutdruck als beim Wirkstoff (Punkt B_2A_1), die Differenz $\bar{d}_{1\cdot}$ beträgt 3 , was sich in der fallenden Steigung der Linie A_1 auswirkt. In der Männer-Stufe (Linie A_2) zeigt sich das gleiche Bild: Beim Placebo (Punkt B_1A_2) zeigt sich ein höherer Blutdruck als beim Wirkstoff (Punkt B_2A_2), die Differenz $\bar{d}_{2\cdot}$ ist auch hier 3 , was sich in der gleichen Steigung der Linie A_2 wie in Linie A_1 auswirkt. Die einfachen Haupteffekte des Faktors B sind in beiden Stufen des Faktors A gleich, d.h. die Zellmittelwertdifferenzen sind in beiden Stufen gleich hoch, was sich in einer Parallelität der Linien auswirkt. (Insgesamt ist die Linie A_1 niedriger als die Linie A_2 . Das bedeutet, dass die Blutdruckwerte bei Frauen niedriger sind als bei Männern.)

Letztendlich lässt sich feststellen, dass keine Interaktion vorliegt, wenn die Differenzen der Zellmittelwerte jeder Stufe eines Faktors auf allen Stufen des anderen Faktors gleich sind, und sich dies im Interaktionsdiagramm in einer Parallelität der Linien auswirkt: Die einfachen Haupteffekte eines Faktors sind auf allen Stufen des anderen Faktors gleich. Umgekehrt können wir daraus schlussfolgern, dass eine Interaktion vorliegen muss, wenn die Linien nicht mehr parallel sind, also wenn die Zellmittelwertdifferenzen unterschiedlich sind: Die einfachen Haupteffekte eines Faktors sind in mindestens zwei Stufen des anderen Faktors unterschiedlich. Nun ist das aufgrund zufälliger Einflüsse fast immer der Fall, aber wann wird diese Unterschiedlichkeit signifikant? Dies trifft zu, wenn in der zweifaktoriellen Varianzanalyse für die Wechselwirkung $A \times B$ ein signifikantes Ergebnis festgestellt werden konnte.

5.3.4.2. Interaktionstypen

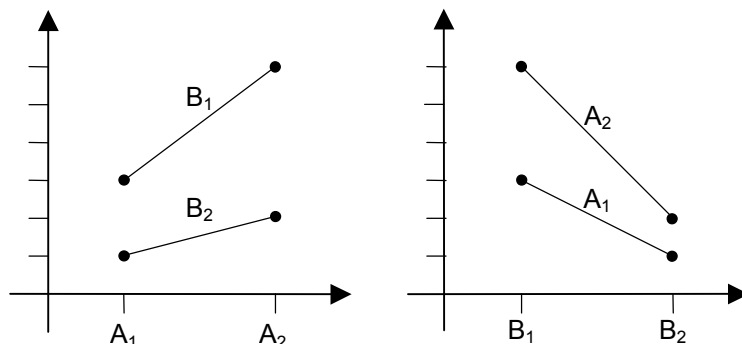
Es lassen sich drei Typen von Interaktionen klassifizieren: *ordinale*, *semiordinale (hybride)* und *disordinale Interaktionen*. Die Unterscheidung beruht nicht darauf, inwieweit die einfachen Haupteffekte in den einzelnen Stufen unterschiedlich stark sind, was sich in unterschiedlichen starken Zellmittelwertdifferenzen in den einzelnen Stufen bemerkbar macht, obwohl die Interpretation der Unterschiedlichkeit der Mittelwertdifferenzen in den einzelnen Stufen eines Faktors die Interaktion sichtbar macht und daher notwendig ist. Die Unterscheidung beruht im wesentlichen darauf, ob sich der Haupteffekt eines betrachteten Faktors auf *allen* einzelnen Stufen des anderen Faktors in die *gleiche* oder in *mindestens* einer Stufe in die *umgekehrte* Richtung auswirkt. Wirkt sich der Haupteffekt eines Faktors in allen Stufen des ande-

ren Faktors gleich aus, so weisen die Zellmittelwerte der Stufen des betrachteten Faktors in allen Stufen des anderen Faktors die gleiche Reihenfolge auf. Die Zellmittelwertdifferenzen zeigen dadurch alle das gleiche Vorzeichen wie die Mittelwertdifferenz des Haupteffektes, wodurch die zugehörigen Linien die gleiche Steigung (Trend) besitzen. Damit ist der Haupteffekt eindeutig interpretierbar. Wirkt sich aber der Haupteffekt eines Faktors in mindestens einer Stufe in die umgekehrte Richtung aus, so weisen die Zellmittelwerte dieser Stufe eine andere Rangfolge auf. Die Zellmittelwertdifferenzen zeigen dann in dieser Stufe ein umgekehrtes Vorzeichen als die Mittelwertdifferenz des Haupteffektes. Die zugehörige Linie zeigt einen anderen Trend als die Linien der anderen Stufen. Der Haupteffekt ist damit nicht mehr eindeutig interpretierbar und daher müssen in allen einzelnen Stufen die einfachen Haupteffekte interpretiert werden.

Der Typ der Interaktion zwischen zwei Faktoren ergibt sich aus der kombinierten Betrachtung der Interaktionsdiagramme beider Faktoren: Für jeden Faktor wird ein Interaktionsdiagramm angelegt und mit dem anderen Faktor verglichen. Wirken sich beide Faktoren auf allen Stufen des anderen jeweils in die gleiche Richtung aus, so ist die Interaktion zwischen beiden Faktoren ordinal, wirken sich beide Faktoren auf jeweils mindestens einer Stufe in die entgegengesetzte Richtung aus, so liegt eine disordinale Interaktion vor, zeigen beide Faktoren einen jeweils anderen Wirkungstyp, so ergibt sich daraus eine semiordinale (hybride) Interaktion. Letztendlich lässt sich dies in den Interaktionsdiagrammen am einfachsten erkennen: Überkreuzen sich die Linien in beiden Interaktionsdiagrammen nicht, liegt eine ordinale Interaktion vor, überkreuzen sich die Linien nur in einem der beiden Diagramme, so liegt eine semiordinale (hybride) Interaktion vor, und überkreuzen sich die Linien in beiden Diagrammen, so liegt eine disordinale Interaktion vor.

Ordinale Interaktion

	B ₁	B ₂	$\bar{x}_{j\cdot}$	$d_{j\cdot}$
A ₁	3	1	2	2
A ₂	6	2	4	4
$\bar{x}_{\cdot k}$	4,5	1,5		
$d_{\cdot k}$	-3	-1		

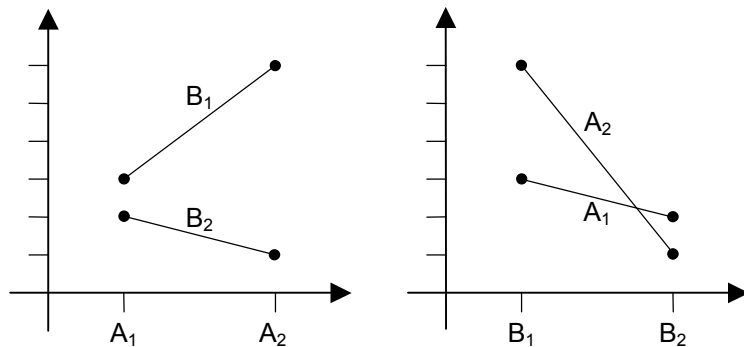


In beiden Abbildungen zeigen die Linien jeweils für sich den gleichen Trend auf. Hinsichtlich der Haupteffekte kann man wie bei der Nullinteraktion sagen, dass diese eindeutig sind (s.S. 67).

Die Interaktion macht sich in den unterschiedlichen Zellmittelwertdifferenzen der Faktorstufen bemerkbar: Das Medikament $\bar{d}_{1\cdot}$ bewirkt bei den Männern eine stärkere Blutdrucksenkung als bei den Frauen: $\bar{d}_{1\cdot} < \bar{d}_{2\cdot}$. Dies wird in der rechten Abbildung deutlich, wo die Linie A₂ schneller abfällt als Linie A₁. Eine analoge Verlaufsbeobachtung für das Geschlecht ist an diesem Beispiel nicht möglich, da das Geschlecht ein personengebundener Faktor ist: Man kann nicht willkürlich Personen den Stufen A₁ u. A₂ zuweisen und ihnen dann das entsprechende Geschlecht geben. Allerdings kann man sagen, dass die Blutdruckdifferenz zwischen den Geschlechtern sich verkleinert, wenn man ihnen das Medikament gibt: $\bar{d}_{\cdot 1} > \bar{d}_{\cdot 2}$ (B₁ steiler als B₂).

Semiordinale (hybride) Interaktion

	B ₁	B ₂	$\bar{x}_{j\bullet}$	$d_{j\bullet}$
A ₁	3	2	2,5	1
A ₂	6	1	3,5	5
$\bar{x}_{\bullet k}$	4,5	1,5		
$d_{\bullet k}$	-3	1		

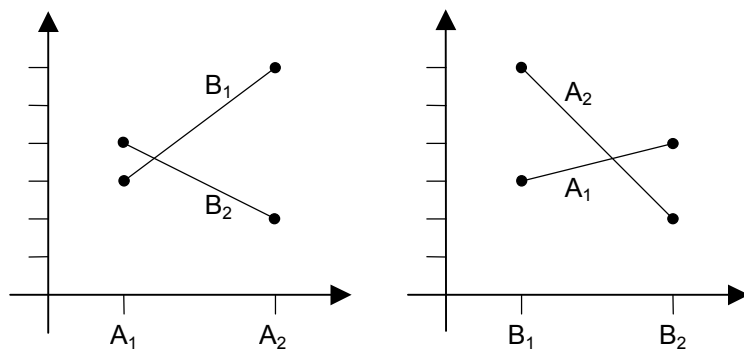


Das linke Diagramm zeigt zwei Linienzüge mit gegenläufigem Trend, was zwangsläufig dazu führt, dass sich die Linien im rechten Diagramm überschneiden. Dennoch sind die Trends im rechten Diagramm gleichsinnig. Die Rangfolge der Mittelwerte des Haupteffekts B ($\bar{x}_{\bullet 1} > \bar{x}_{\bullet 2}$) gilt für beide Stufen des Faktors A: $\bar{x}_{11} > \bar{x}_{12}$ und $\bar{x}_{21} > \bar{x}_{22}$. Der Haupteffekt ist damit eindeutig interpretierbar: Das Medikament bewirkt eine höhere Blutdrucksenkung als das Placebo, unabhängig davon, ob es den Frauen oder den Männern verabreicht wurde. Die Interaktion macht sich in den Mittelwertdifferenzen $\bar{d}_{k\bullet}$ in den einzelnen Faktorstufen bemerkbar. Das Medikament bewirkt bei den Männern eine stärkere Blutdrucksenkung als bei den Frauen: $\bar{d}_{1\bullet} < \bar{d}_{2\bullet}$, was sich im Interaktionsdiagramm damit ausdrückt, dass die Linie A₂ stärker abfällt als die Linie A₁.

Haupteffekt A sollte hingegen nicht interpretiert werden. Die Aussage $\bar{x}_{1\bullet} < \bar{x}_{2\bullet}$ gilt nur für die Stufe B₁: $\bar{x}_{11} < \bar{x}_{21}$. Für die Stufe B₂ ist der Trend genau umgekehrt: $\bar{x}_{12} > \bar{x}_{22}$. Statt dessen sollte die Interaktion in den einzelnen Stufen beschrieben werden: Frauen haben bei Placeboeinnahme einen niedrigeren Blutdruck als Männer (Linie B₁: $\bar{d}_{\bullet 1} = -3$), während sie bei Einnahme des Wirkstoffes einen höheren Blutdruck als Männer besitzen (Linie B₂: $\bar{d}_{\bullet 2} = 1$).

Disordinale Interaktion

	B ₁	B ₂	$\bar{x}_{j\bullet}$	$d_{j\bullet}$
A ₁	3	4	3,5	-1
A ₂	6	2	4	4
$\bar{x}_{\bullet k}$	4,5	3		
$d_{\bullet k}$	-3	2		



Beide Abbildungen verdeutlichen divergierende Linien sowohl im linken als auch im rechten Diagramm, d.h. beide Haupteffekte sind für sich genommen inhaltlich bedeutungslos. Haupteffekt A ($\bar{x}_{1\bullet} < \bar{x}_{2\bullet}$) findet sich nur in Stufe B₁ wieder, in Stufe B₂ können wir einen gegenteiligen Effekt beobachten. Haupteffekt B ($\bar{x}_{\bullet 1} > \bar{x}_{\bullet 2}$) findet sich nur in Stufe A₂ wieder, in Stufe A₁ ist der Effekt umgekehrt.

Unterschiede zwischen den Faktorstufen A₁ und A₂ sind nur jeweils in den einzelnen Stufen B₁ und B₂ und Unterschiede zwischen den Faktorstufen B₁ und B₂ sind nur jeweils in den einzelnen Stufen A₁ und A₂ sinnvoll interpretierbar. Die einfachen Haupteffekte B in A machen deutlich, dass die Einnahme des Medikamentes bei den

Männern eine Blutdrucksenkung (Linie A₂ steigend: $\bar{d}_{2\bullet} = 4$), bei den Frauen hingegen eine Blutdruckerhöhung (Linie A₁ fallend: $\bar{d}_{1\bullet} = -1$) zur Folge hat. Die einfachen Haupteffekte A in den Stufen von B besagen auch hier, dass Frauen bei Placeboeinnahme einen niedrigeren Blutdruck als Männer aufweisen (Linie B₁: $\bar{d}_{\bullet 1} = -3$), während sie bei Einnahme des Wirkstoffes einen höheren Blutdruck als Männer besitzen (Linie B₂: $\bar{d}_{\bullet 2} = 2$).

5.3.4.3. Signifikanztests für einfache Haupteffekte

In den obigen Beispielen zu den einzelnen Interaktionstypen wurde vorausgesetzt, dass alle Mittelwertunterschiede signifikant seien. Im Einzelfall, wenn eine Interaktion in der Varianzanalyse festgestellt wurde, muss man jedoch bestimmen können, welche einfachen Haupteffekte signifikant sind und welche nicht. Dies macht man insbesondere dann, wenn der Fall eintritt, dass ein Faktor nicht signifikant ist, die Wechselwirkung AxB aber. Dann ermittelt man, in welchen Stufen des anderen Faktors dieser Faktor signifikante einfache Haupteffekte aufweist, da sonst keine Wechselwirkung aufgetreten wäre. Im Prinzip wird hierzu in allen Stufen des anderen Faktors eine einfaktorische Varianzanalyse durchgeführt, so dass wir mittlere Quadrate für jede einzelne Stufe ermitteln müssen.

	B ₁	B ₂	$\bar{X}_{j\bullet}$
A ₁	\bar{X}_{11}	\bar{X}_{12}	$\bar{X}_{1\bullet}$
A ₂	\bar{X}_{21}	\bar{X}_{22}	$\bar{X}_{2\bullet}$
$\bar{X}_{\bullet k}$	$\bar{X}_{\bullet 1}$	$\bar{X}_{\bullet 2}$	

Für den Fall, dass wir die einfachen Haupteffekte des Faktors A in den Stufen von Faktor B auf Signifikanz prüfen wollen, wollen wir die Zellmittelwerte in den einzelnen Stufen von B auf signifikante Mittelwertunterschiede testen:

Einfache Haupteffekte A in den Stufen von B				
$SAQ_{A(B_k)}$	$MQ_{A(B_k)}$	F	df ₁	df ₂
$n \cdot \sum_j (\bar{x}_{jk} - \bar{x}_{\bullet k})^2$	$\frac{SAQ_{A(B_k)}}{J-1}$	$\frac{MQ_{A(B_k)}}{MQ_I}$	J-1	J·K·(n-1)

Beispiel: Wenn wir obigen Versuchsplan betrachten und die einfachen Haupteffekte des Faktors A testen wollen, stellen wir folgende Hypothesen und SAQ auf:

Einfacher Haupteffekt A in Stufe B₁:

Statistische Hypothesen:

$$H_0: \mu_{11} = \mu_{21} \quad H_1: \mu_{11} \neq \mu_{21}$$

Berechnung des SAQ:

$$SAQ_{A(B_1)} = n \cdot [(\bar{X}_{11} - \bar{X}_{\bullet 1})^2 + (\bar{X}_{21} - \bar{X}_{\bullet 1})^2]$$

Einfacher Haupteffekt A in Stufe B₂:

Statistische Hypothesen:

$$H_0: \mu_{12} = \mu_{22} \quad H_1: \mu_{12} \neq \mu_{22}$$

Berechnung des SAQ:

$$SAQ_{A(B_2)} = n \cdot [(\bar{X}_{12} - \bar{X}_{\bullet 2})^2 + (\bar{X}_{22} - \bar{X}_{\bullet 2})^2]$$

Für den Fall, dass wir die einfachen Haupteffekte des Faktors B in den Stufen von Faktor A auf Signifikanz prüfen wollen, wollen wir die Zellmittelwerte in den einzelnen Stufen von A auf signifikante Mittelwertunterschiede testen:

Einfache Haupteffekte B in den Stufen von A				
$SAQ_{B(A_k)}$	$MQ_{B(A_k)}$	F	df ₁	df ₂
$n \cdot \sum_k (\bar{x}_{jk} - \bar{x}_{j\cdot})^2$	$\frac{SAQ_{B(A_k)}}{K-1}$	$\frac{MQ_{B(A_k)}}{MQ_1}$	K-1	J·K·(n-1)

Beispiel: Wenn wir obigen Versuchsplan betrachten und die einfachen Haupteffekte des Faktors B testen wollen, stellen wir folgende Hypothesen und SAQ auf:

Einfacher Haupteffekt B in Stufe A₁:

Statistische Hypothesen:

$$H_0: \mu_{11} = \mu_{12} \quad H_1: \mu_{11} \neq \mu_{12}$$

Berechnung der SAQ:

$$SAQ_{B(A_1)} = n \cdot [(\bar{X}_{11} - \bar{X}_{1\cdot})^2 + (\bar{X}_{12} - \bar{X}_{1\cdot})^2]$$

Einfacher Haupteffekt B in Stufe A₂:

Statistische Hypothesen:

$$H_0: \mu_{21} = \mu_{22} \quad H_1: \mu_{21} \neq \mu_{22}$$

Berechnung der SAQ:

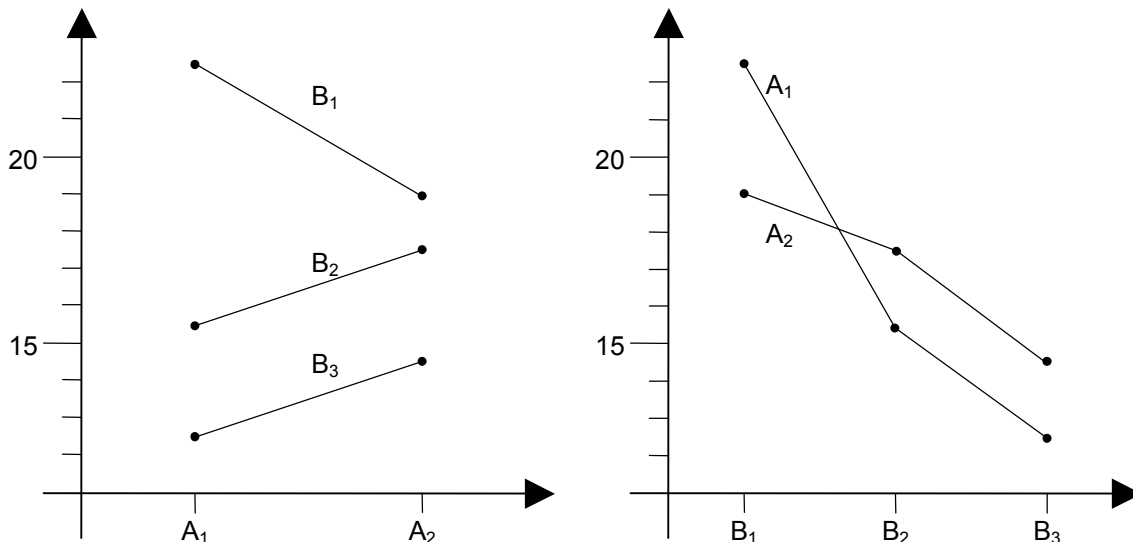
$$SAQ_{B(A_2)} = n \cdot [(\bar{X}_{21} - \bar{X}_{2\cdot})^2 + (\bar{X}_{22} - \bar{X}_{2\cdot})^2]$$

Beispiel zur Ermittlung der Interaktion zwischen den Faktoren

Beispiel 2 von Herrn HUBER:

\bar{X}_{jk}	B ₁	B ₂	B ₃	$\bar{X}_{j\cdot}$
A ₁	22,4	15,6	12,4	16,8
A ₂	18,8	17,6	14,6	17,0
$\bar{X}_{\cdot k}$	20,6	16,6	13,5	16,9

$$J = 2; K = 3; n = 5; N = 30$$



Wie man auf dem ersten Blick erkennen kann, handelt es sich um eine semiordinale Interaktion: Das Interaktionsdiagramm zu Haupteffekt A zeigt keine Überschneidung, während das Diagramm zu Haupteffekt B eine solche aufweist. In der linken Abbildung sieht man einen gegenteiligen einfachen Haupteffekt A in B₁ gegenüber den gleichgerichteten einfachen Haupteffekten A in B₂ und B₃, während in der rechten zwei gleichläufige Trends der einfachen Haupteffekte B in A₁ und A₂ zu erkennen sind.

Da der Haupteffekt von Faktor B signifikant ist, sind auch die einfachen Haupteffekte von Faktor B signifikant. Sie zeigen alle den gleichen Trend, der Haupteffekt B kann sinnvoll interpretiert werden: Mit steigender Wirkstoffkonzentration sinkt die Depressivität, unabhängig davon, ob Männer oder Frauen den Wirkstoff einnehmen. Die einfachen Haupteffekte von Faktor B können wir wie folgt interpretieren: Von Stufe B_1 zu B_2 sinkt die Depressivität bei Männern schneller als bei Frauen, Frauen scheinen eher auf den Placebo-Effekt zu reagieren als Männer oder sind an sich weniger depressiv. Die Linien verlaufen hingegen von Stufe B_2 zu Stufe B_3 fast parallel: Hier ist keine Interaktion feststellbar, die doppelte Dosis zeigt fast dieselbe Wirkungssteigerung sowohl bei den Männern als bei den Frauen. (Die Linie A_1 der Männer liegt unter der Linie A_2 der Frauen, d.h. das Medikament ist bei den Männern wirksamer.)

Der Haupteffekt A ist nicht signifikant, es müssen die einfachen Haupteffekte auf den einzelnen Stufen von B getestet werden (Signifikanzniveau 5%):

$$\begin{aligned} \text{A in } B_1: \quad \text{SAQ}_{A(B_1)} &= n \cdot \sum_j (\bar{x}_{j1} - \bar{x}_{\cdot 1})^2 \\ &= n \cdot [(\bar{x}_{11} - \bar{x}_{\cdot 1})^2 + (\bar{x}_{21} - \bar{x}_{\cdot 1})^2] \\ &= 5 \cdot [(22,4 - 20,6)^2 + (18,8 - 20,6)^2] = 32,4 \end{aligned}$$

$$\begin{aligned} \text{A in } B_2: \quad \text{SAQ}_{A(B_2)} &= n \cdot \sum_j (\bar{x}_{j2} - \bar{x}_{\cdot 2})^2 \\ &= n \cdot [(\bar{x}_{12} - \bar{x}_{\cdot 2})^2 + (\bar{x}_{22} - \bar{x}_{\cdot 2})^2] \\ &= 5 \cdot [(15,6 - 16,6)^2 + (17,6 - 16,6)^2] = 10,0 \end{aligned}$$

$$\begin{aligned} \text{A in } B_3: \quad \text{SAQ}_{A(B_3)} &= n \cdot \sum_j (\bar{x}_{j3} - \bar{x}_{\cdot 3})^2 \\ &= n \cdot [(\bar{x}_{13} - \bar{x}_{\cdot 3})^2 + (\bar{x}_{23} - \bar{x}_{\cdot 3})^2] \\ &= 5 \cdot [(12,4 - 13,5)^2 + (14,6 - 13,5)^2] = 12,1 \end{aligned}$$

Tafel der Varianzanalyse:

	SAQ	df	MQ	F	$F_{0,95}$	Ergebnis
A in B_1	32,4	1	32,4	19,1	4,26	s.
A in B_2	10,0	1	10,0	5,9	4,26	s.
A in B_3	12,1	1	12,1	7,1	4,26	s.
I	40,8	24	1,7			

Die einfachen Haupteffekte A sind auf allen Stufen von Faktor B signifikant. Wenn man sich das Interaktionsdiagramm zu Faktor A in der linken Abbildung oben anschaut, so zeigen die einfachen Haupteffekte nicht den gleichen Trend: Linie B_1 fällt, während die Linien B_2 und B_3 ansteigen. Der Haupteffekt von Faktor A könnte also auch dann nicht sinnvoll interpretiert werden, wenn er signifikant wäre. Es müssen die einzelnen Haupteffekte gedeutet werden: Bei Einnahme eines Placebos sind die Frauen weniger depressiv als die Männer (Linie B_1). Entweder sind Frauen generell weniger depressiv oder bei Ihnen wirkt sich der Placebo-Effekt stärker aus. Bei Einnahme des Medikaments sind die Männer weniger depressiv als die Frauen, sowohl

bei einfacher als bei doppelter Dosis (Linie B_2 und B_3). Männer reagieren also eher auf das Medikament als Frauen. Linie B_2 und B_3 sind parallel: Hier liegt keine Wechselwirkung zwischen dem Geschlecht und der einfachen bzw. doppelten Dosis vor. (Die Linie B_3 verläuft unterhalb der Linie B_2 und diese unterhalb von B_1 : Die doppelte Dosis zeigt eine stärkere Wirkung als die einfache Dosis und diese eine stärkere Wirkung als das Placebo.) Generell erfüllt das Medikament seinen Sinn bei einer Irrtumswahrscheinlichkeit von 5%.

5.4. Voraussetzungen zu den Varianzanalysen

Sollen Mittelwertunterschiede mit der Varianzanalyse auf Signifikanz geprüft werden, müssen hinsichtlich der Stichproben zu den einzelnen Faktorstufen (einfaktoriell) bzw. Faktorstufenkombinationen (mehrfaktoriell) folgende Bedingungen erfüllt sein:

1. Die den Stichproben zugehörigen Grundgesamtheiten müssen normalverteilt sein.
2. Die den Stichproben zugehörigen Grundgesamtheiten müssen varianzhomogen sein.

Um die Voraussetzungen zur Varianzanalyse zu prüfen, müssen Testverfahren eingesetzt werden, die z.T. nicht so robust wie die Varianzanalyse sind. D.h. dass die Durchführung einer Varianzanalyse auf Grund eines signifikanten Vortests kontraindiziert erscheint, obwohl der F-Test als robustes Verfahren durchaus noch zu richtigen Entscheidungen führen würde. Zu der Frage, wie die Varianzanalyse reagiert, wenn eine ihrer Voraussetzungen verletzt sind, wurden zahlreiche Untersuchungen durchgeführt. Generell gilt, dass die Voraussetzungen der Varianzanalyse mit wachsenden Umfang der untersuchten Stichproben an Bedeutung verlieren. Heterogene Varianzen beeinflussen die Varianzanalyse nur unerheblich, wenn die untersuchten Stichproben gleich groß sind. Bei ungleich großen Stichproben und heterogenen Varianzen ist die Gültigkeit des F-Tests vor allem bei kleineren Stichprobenumfängen erheblich gefährdet.

Zusammenfassend ist festzustellen, dass die Varianzanalyse bei

- hinreichend großen ($n > 10$) und
- gleich großen Stichproben

gegenüber Verletzungen ihrer Voraussetzungen relativ robust ist. Dies gilt weitgehend auch für die Durchführung von Einzelvergleichen nach dem SCHEFFÉ-Test.

5.5. Varianzanalyse und klassischer t-Test

Die einfaktorielle Varianzanalyse für $J=2$ Stichproben ist mit dem klassischen t-Test für unabhängige Stichproben identisch. Für die kritischen Grenzen gilt:

$$\left(t_{1-\frac{\alpha}{2}; df} \right)^2 = F_{1-\alpha; 1; df}$$

Das Quadrat einer t-Größe ist eine F-Größe. Dabei ist zu beachten, dass die t-Größe zweiseitig und die F-Größe einseitig ist. Der Zählerfreiheitsgrad der F-Größe ist immer 1, der Nennerfreiheitsgrad immer der Freiheitsgrad der t-Größe.

6. Nichtparametrische, verteilungsfreie Verfahren

Bisher haben wir parametrische Verfahren kennengelernt, die alle auf der Normalverteilung mit ihren Parametern μ und σ beruhen. Daher sind alle bisherigen Tests nur auf *metrischen Skalenniveaus* sinnvoll. Nichtparametrische und verteilungsfreie Verfahren sind eigentlich irreführende Bezeichnungen. Sie dienen zur Untersuchung von Sachverhalten auf *nichtmetrischen Skalenniveaus*. Hier interessieren uns insbesondere Stichprobenverteilungen von Häufigkeiten für nominale Merkmale mit dem Parameter π : die Binomial- und χ^2 -Verteilung.

6.1. Verfahren zur Binomialverteilung

6.1.1. Die Binomialverteilung

Die Binomialverteilung ist eine diskrete Wahrscheinlichkeitsverteilung, die Ereignisse betrifft, die in *zwei Alternativen* auftreten. Diese können dabei gleich oder ungleich wahrscheinlich sein. Gleich wahrscheinliche Alternativen sind z.B. Adler vs. Zahl beim Münzwurf, gerade Zahl vs. ungerade Zahl beim Würfeln usw.; ungleich wahrscheinliche Alternativen sind z.B. Zahl 1 vs. andere Augenzahl, Ledige vs. anderer Familienstatus usw.

Führen wir mehrmals hintereinander ein Zufallsexperiment mit zwei alternativen Ausgängen A und B durch, so erhalten wir eine Zufallsabfolge von zwei Ereignissen A und B mit konstanter Wahrscheinlichkeit π für das Ereignis A und $1-\pi$ für das Ereignis B. Wenn n Versuche durchgeführt werden, kann das Ereignis A 0mal, 1mal, ..., n -mal auftreten. Die Binomialverteilung ist die Wahrscheinlichkeitsdichtefunktion $f(X)$ für die Zufallsvariable $X = \{\text{Häufigkeit des Auftretens des Ereignisses A bei } n \text{ Versuchen}\}$. Diese Wahrscheinlichkeitsverteilung ist abhängig von den Wahrscheinlichkeiten der beiden alternativen Ereignissen und von der Anzahl n der Versuche.

Beispiel: Eine Münze wird 5mal geworfen und es wird mitgezählt, wie oft Adler gefallen ist. Diese Prozedur wird unendlich oft durchgeführt. Anschließend wird eine Häufigkeitsverteilung aufgestellt, die angibt, wie häufig eine bestimmte Anzahl von Adlern beim 5maligem Werfen aufgetreten ist. Damit gibt diese Verteilung an, wie wahrscheinlich es ist, bei einem Versuchsdurchgang eine bestimmte Anzahl von Adlern zu erhalten. Dieselbe Verteilung ergibt sich, wenn man 5 Münzen gleichzeitig wirft und dann die Anzahl der gefallenen Adler ermittelt.

Eine Binomialverteilung wird aufgestellt, indem für jede mögliche Auftretenshäufigkeit f die zugehörige Wahrscheinlichkeit $W(f)$ berechnet wird. Dies kann mit Hilfe der Wahrscheinlichkeitsrechnung oder durch folgendes Bildungsgesetz berechnet werden:

$$W(f) = \binom{n}{f} \cdot \pi^f \cdot (1-\pi)^{n-f}$$

n = Anzahl der Versuche
 f = Auftretenshäufigkeit von A
 π = Wahrscheinlichkeit, dass A eintritt

Das erste Glied auf der rechten Seite der Gleichung wird Binomialkoeffizient genannt und wie folgt ermittelt:

$$\binom{n}{f} = \frac{n \cdot (n-1) \cdot \dots \cdot (n-f+1)}{1 \cdot 2 \cdot \dots \cdot f} \quad \text{wobei} \quad \binom{n}{0} = \binom{n}{n} = 1$$

Am einfachsten wird der Binomialkoeffizient aufgestellt, indem man einen Bruchstrich zieht und zuerst im Nenner die Produktreihe aufsteigend von 1 bis f bildet und dann im Zähler von n ab absteigend die Produktreihe so lange weiterführt, bis genauso viele Elemente wie im Nenner miteinander multipliziert werden.

$$\binom{4}{0} = 1$$

$$\binom{4}{1} = \frac{4}{1} = 4$$

Beispiele:

$$\binom{4}{2} = \frac{4 \cdot 3}{1 \cdot 2} = 6$$

$$\binom{4}{3} = \frac{4 \cdot 3 \cdot 2}{1 \cdot 2 \cdot 3} = 4$$

$$\binom{4}{4} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{1 \cdot 2 \cdot 3 \cdot 4} = 1$$

Wir wollen nun die Binomialverteilung aufstellen für den Fall, dass ein Prüfling in einem Multiple-Choice-Test die Antworten zufällig ankreuzt. Der Test besteht aus vier Aufgaben. Jede Aufgabe hat vier Lösungsmöglichkeiten, von denen nur eine richtig ist. Die Wahrscheinlichkeit, in einer Aufgabe zufällig die richtige Lösungsmöglichkeit zu raten, beträgt $W(R) = \pi = 1/4$ (HUBER, STAT -6-, Regel 2). Die Wahrscheinlichkeit für die andere Alternative, eine Aufgabe falsch zu erraten, beträgt $W(F) = 1 - \pi = 3/4$. Die Wahrscheinlichkeiten, im Gesamttest keine, eine, zwei, drei oder vier Antworten zufällig richtig anzukreuzen, sind nach den Regeln der Wahrscheinlichkeit bzw. dem Bildungsgesetz für Binomialwahrscheinlichkeiten wie folgt zu bestimmen:

Keine Aufgabe richtig gelöst: $f=0$

Alle Aufgaben sind falsch gelöst: Die erste *und* die zweite *und* die dritte *und* die vierte. Nach dem Multiplikationssatz der Wahrscheinlichkeiten für unabhängige Ereignisse (Regel 4, STAT -7-) ergibt sich:

$$W(F \cap F \cap F \cap F) = 3/4 \cdot 3/4 \cdot 3/4 \cdot 3/4 = (3/4)^4 = 0,316$$

Nach dem Bildungsgesetz der Binomialverteilung können wir errechnen:

$$W(0) = \binom{4}{0} \cdot (1/4)^0 \cdot (1 - 1/4)^{4-0} = 1 \cdot 1 \cdot (3/4)^4 = 0,316$$

Eine Aufgabe richtig gelöst: $f=1$

Es gibt vier Möglichkeiten: Entweder die erste *oder* die zweite *oder* die dritte *oder* die vierte ist richtig gelöst. Für jeden Fall müssen wir erst mal eine Einzelwahrscheinlichkeit errechnen:

$$W(R \cap F \cap F \cap F) = 1/4 \cdot 3/4 \cdot 3/4 \cdot 3/4 = 1/4 \cdot (3/4)^3 = 0,1055$$

$$W(F \cap R \cap F \cap F) = 3/4 \cdot 1/4 \cdot 3/4 \cdot 3/4 = 1/4 \cdot (3/4)^3 = 0,1055$$

$$W(F \cap F \cap R \cap F) = 3/4 \cdot 3/4 \cdot 1/4 \cdot 3/4 = 1/4 \cdot (3/4)^3 = 0,1055$$

$$W(F \cap F \cap F \cap R) = 3/4 \cdot 3/4 \cdot 3/4 \cdot 1/4 = 1/4 \cdot (3/4)^3 = 0,1055$$

Nach dem Additionssatz der Wahrscheinlichkeit (HUBER, STAT -6-, Regel 3) ergibt die Summe der Einzelwahrscheinlichkeiten die Gesamtwahrscheinlichkeit für das Ereignis, in dem Test eine Aufgabe zufällig richtig zu lösen:

$$W(1) = 0,1055 + 0,1055 + 0,1055 + 0,1055 = 4 \cdot 0,1055 = 0,422$$

Nach dem Bildungsgesetz gilt:

$$W(1) = \binom{4}{1} \cdot \left(\frac{1}{4}\right)^1 \cdot \left(1 - \frac{1}{4}\right)^{4-1} = 4 \cdot \frac{1}{4} \cdot \left(\frac{3}{4}\right)^3 = 0,422$$

Zwei Aufgaben richtig gelöst: f=2

Es gibt sechs Möglichkeiten:

$$\begin{aligned} W(R \cap R \cap F \cap F) &= \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} = \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^2 = 0,0352 \\ W(R \cap F \cap R \cap F) &= \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} = \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^2 = 0,0352 \\ W(R \cap F \cap F \cap R) &= \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} = \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^2 = 0,0352 \\ W(F \cap R \cap R \cap F) &= \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} = \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^2 = 0,0352 \\ W(F \cap R \cap F \cap R) &= \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} = \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^2 = 0,0352 \\ W(F \cap F \cap R \cap R) &= \frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^2 = 0,0352 \end{aligned}$$

$$W(2) = 6 \cdot 0,035 = 0,211$$

Bildungsgesetz:
$$W(2) = \binom{4}{2} \cdot \left(\frac{1}{4}\right)^2 \cdot \left(1 - \frac{1}{4}\right)^{4-2} = 6 \cdot \left(\frac{1}{4}\right)^2 \cdot \left(\frac{3}{4}\right)^2 = 0,211$$

Drei Aufgaben richtig gelöst: f=3

Es gibt vier Möglichkeiten:

$$\begin{aligned} W(F \cap R \cap R \cap R) &= \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{3}{4} \cdot \left(\frac{1}{4}\right)^3 = 0,0117 \\ W(R \cap F \cap R \cap R) &= \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{3}{4} \cdot \left(\frac{1}{4}\right)^3 = 0,0117 \\ W(R \cap R \cap F \cap R) &= \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{4} \cdot \left(\frac{1}{4}\right)^3 = 0,0117 \\ W(R \cap R \cap R \cap F) &= \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{4} \cdot \left(\frac{1}{4}\right)^3 = 0,0117 \end{aligned}$$

$$W(3) = 4 \cdot 0,0117 = 0,047$$

Bildungsgesetz:
$$W(3) = \binom{4}{3} \cdot \left(\frac{1}{4}\right)^3 \cdot \left(1 - \frac{1}{4}\right)^{4-3} = 4 \cdot \left(\frac{1}{4}\right)^3 \cdot \frac{3}{4} = 0,047$$

Vier Aufgaben richtig gelöst: f=4

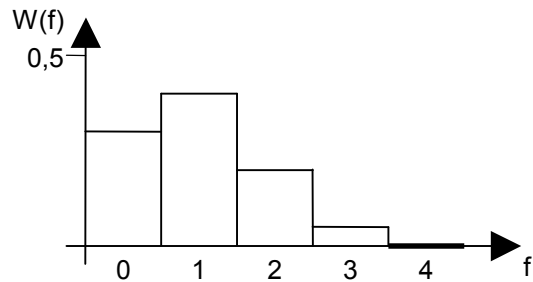
$$W(R \cap R \cap R \cap R) = \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \left(\frac{1}{4}\right)^4 = 0,004$$

Bildungsgesetz:
$$W(4) = \binom{4}{4} \cdot \left(\frac{1}{4}\right)^4 \cdot \left(1 - \frac{1}{4}\right)^{4-4} = 1 \cdot \left(\frac{1}{4}\right)^4 \cdot 1 = 0,004$$

Demnach sieht die gesuchte Binomialverteilung wie folgt aus:

Stoff

f	W(f)
0	0,316
1	0,422
2	0,211
3	0,047
4	0,004
Σ	1



Es liegt eine asymmetrische Verteilung vor, weil $\pi \neq 1/2$. Bei $\pi = 1/2$ erhält man eine symmetrische Wahrscheinlichkeitsverteilung.

Der Erwartungswert μ der Binomialverteilung – auch Erfolgszahl genannt – lässt sich anhand folgender Überlegung bestimmen: Die Häufigkeit f ergibt sich aus der Summe der Zufallsvariablen x_i , die mit der Wahrscheinlichkeit π den Wert 1 annimmt, sonst den Wert 0. Wie wir in Hausaufgabe Nr. 43 bereits bestimmt haben, ist der Erwartungswert μ der Zufallsvariablen x_i gleich π selbst, und die Varianz der Zufallsvariablen x_i gleich $\pi \cdot (1-\pi)$. Damit ergibt sich:

$$f = x_1 + x_2 + x_3 + \dots$$

$$\begin{aligned} \mu_f &= \mu_{x_1} + \mu_{x_2} + \mu_{x_3} + \dots \\ &= \pi + \pi + \pi + \dots \\ &= n \cdot \pi \end{aligned}$$

$$\begin{aligned} \sigma_f^2 &= \sigma_{x_1}^2 + \sigma_{x_2}^2 + \sigma_{x_3}^2 + \dots \\ &= \pi \cdot (1-\pi) + \pi \cdot (1-\pi) + \pi \cdot (1-\pi) + \dots \\ &= n \cdot \pi \cdot (1-\pi) \end{aligned}$$

Beispiel: Für die obige Binomialverteilung können wir die Erfolgszahl μ und die Varianz σ^2 wie folgt bestimmen: $\mu = n \cdot \pi = 4 \cdot 1/4 = 1$; $\sigma^2 = n \cdot \pi \cdot (1-\pi) = 4 \cdot 1/4 \cdot 3/4 = 3/4$.

Für die symmetrische Verteilungen mit $\pi=1/2$ gibt es bei HUBER eine Tabelle, die die einzelnen Wahrscheinlichkeiten $W(f)$ zu einem gegebenen n bis zu 25 auflistet. Zu berücksichtigen ist dabei allerdings, dass es sich hierbei um kumulierte Wahrscheinlichkeiten handelt. Um die Wahrscheinlichkeit zu einem bestimmten f zu erhalten, muss von der Wahrscheinlichkeit, die angegeben ist, die links davon befindliche abgezogen werden. Ist ein Zahlenpaar zur Subtraktion unvollständig, ist in der Tabelle von Herrn HUBER links eine Null bzw. rechts eine Eins zu vervollständigen. Die Leerfläche unten links besteht aus Nullen, die Leerfläche oben rechts aus Einsen.

Beispiel: Für die Binomialverteilung der Auftretenshäufigkeit des Adlers bei fünf geworfenen Münzen sehen die einzelnen Wahrscheinlichkeiten wie folgt symmetrisch aus:

f	cum W(f)	Berechnung	W(f)
0	0,031	0,031–0,000	0,031
1	0,188	0,188–0,031	0,157
2	0,500	0,500–0,188	0,312
3	0,812	0,812–0,500	0,312
4	0,969	0,969–0,812	0,157
5	1,000	1,000–0,969	0,031

6.1.2. Approximativer Binomialtest

Die Binomialverteilung nähert sich mit wachsendem Stichprobenumfang n der Normalverteilung an und wird durch die Normalverteilung unter folgender bereits bekannter Voraussetzung hinreichend nahe approximiert (s. S 10):

$$n > \frac{9}{\pi \cdot (1 - \pi)}$$

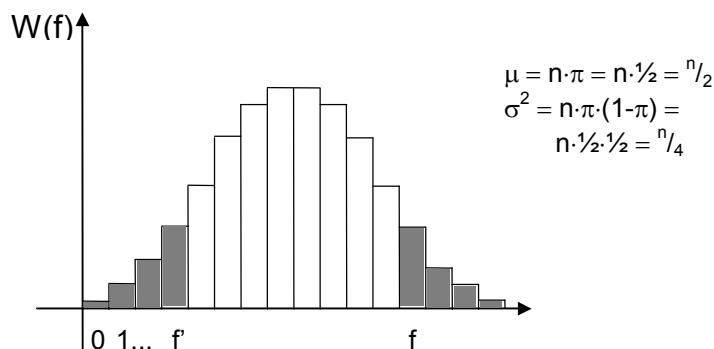
Unter dieser Voraussetzung wird der bereits bekannte approximative Binomialtest (s. S. 28) gerechnet, ist sie nicht erfüllt, wird der exakte Binomialtest benutzt.

6.1.3. Exakter Binomialtest

Im Verfahren des exakten Binomialtests wird die Entscheidung für ein signifikantes Ergebnis über die Überschreitungswahrscheinlichkeit P gefällt. Prüfgröße ist dabei die absolute Häufigkeit f der in der Stichprobe betrachteten Alternative. Je nachdem, ob eine symmetrische ($\pi=1/2$) oder eine asymmetrische Verteilung ($\pi \neq 1/2$) vorliegt, ist eine andere Vorgehensweise zur Bestimmung der Überschreitungswahrscheinlichkeit P angebracht.

6.1.3.1. Symmetrische Verteilung

Im zweiseitigen Verfahren für symmetrische Verteilungen mit $\pi=1/2$ sieht die Überschreitungswahrscheinlichkeit wie folgt aus:



Da in der Tabelle zur symmetrischen Binomialverteilung die Wahrscheinlichkeiten kumuliert aufgestellt sind, gibt die dort angegebene Wahrscheinlichkeit zu der Prüfgröße f und dem Stichprobenumfang n bereits die untere Hälfte der Überschreitungswahrscheinlichkeit an. Diese braucht dann nur noch verdoppelt zu werden. Zum Ablesen in der Tabelle muss jedoch eine rechts von $\mu = n/2$ befindliche Prüfgröße immer in sein linksseitiges Spiegelbild transformiert werden:

$$\begin{aligned} f' &= n - f && \text{falls } f > \mu \\ f' &= f && \text{falls } f \leq \mu \end{aligned}$$

Die Überschreitungswahrscheinlichkeit lautet dann

$$P = 2 \cdot W(f \leq f')$$

Beispiel: Es wird vermutet, dass der Anteil der Frauen im Psychologiestudium nicht dem Anteil der Männer entspricht. In der Bevölkerung sind die Frauen mit $\pi=1/2$ vertreten. Die Alternativhypothese lautet also: $H_1: \pi \neq 1/2$. Eine Stichprobe mit 15 Personen ergibt einen Frauenanteil von $f=12$. Der Stichprobenwert mit $f=12$ ist größer als $\mu=n/2=7,5$. Er muss also in sein kleineres Pendant mit $f' = n-f = 15-12 = 3$ transformiert werden. In der Tabelle mit den kumulierten Wahrscheinlichkeiten zu der Binomialverteilung mit $\pi=1/2$ können wir in der Zeile $n=15$ und der Spalte $f=3$ den Wert $W(f \leq 3) = 0,018$ ablesen. Die Überschreitungswahrscheinlichkeit lautet dann: $P = 2 \cdot W(f \leq 3) = 2 \cdot 0,018 = 0,032$. Die Aussage, Frauen und Männer sind im Psychologiestudium nicht gleich häufig vertreten, wäre damit auf dem 5%-Niveau der Verlässlichkeit signifikant.

Im einseitigen Verfahren besteht die Überschreitungswahrscheinlichkeit nur aus der unteren oder der oberen Hälfte. Beide Hälften sind von der Fläche her gleich. Es kann also bequemerweise immer die untere Hälfte bestimmt werden, um die Tabelle der Binomialverteilung zu $\pi=1/2$ verwenden zu können. Auch hier muss dazu die Prüfgröße in sein kleineres Spiegelbild überführt werden, wenn es größer als die Mitte der Verteilung ist. Die Überschreitungswahrscheinlichkeit P lautet in der einseitigen Fragestellung:

$$P = W(f \leq f')$$

Beispiel: Aufgrund des obigen Stichprobenergebnisses kommen wir zu der Vermutung, dass das Psychologiestudium überdurchschnittlich häufig von Frauen begonnen wird. Die Alternativhypothese lautet jetzt: $H_1: \pi > 1/2$. Eine neue Stichprobe mit 15 Personen ergibt einen Frauenanteil von $f=13$. Der Stichprobenwert mit $f=13$ ist größer als $\mu=n/2$. Er muss also in sein linksseitiges Spiegelbild mit $f' = n-f = 15-13 = 2$ transformiert werden. In der Tabelle mit den kumulierten Wahrscheinlichkeiten zu der Binomialverteilung mit $\pi=1/2$ können wir in der Zeile $n=15$ und der Spalte $f=2$ den Wert $W(f \leq 2) = 0,004$ ablesen. Die Überschreitungswahrscheinlichkeit lautet dann: $P = W(f \leq 2) = 0,004$. Die Aussage, Frauen wären im Psychologiestudium überdurchschnittlich häufig vertreten, wäre damit auf dem 1%-Niveau signifikant.

Beispiele für den exakten Binomialtest bei symmetrischer Verteilung

Angaben: Es wird eine Stichprobe aus einer Berufsgruppe entnommen und die Häufigkeit f der Frauen ermittelt, da vermutet wird, dass die Verteilung von Frauen und Männern in dieser Gesamtheit den Verhältnissen in der Bevölkerung nicht entspricht. Signifikanzniveau: 5%.

- | | |
|----------------------------------------------|-----------------------------------------------------------------------------------------|
| 1. <i>Statistische Hypothesen:</i> | $H_0: \pi = 1/2$ $H_1: \pi \neq 1/2$ |
| 2. <i>Festlegung des Signifikanzniveaus:</i> | $\alpha = 0,05$ |
| 3. <i>Auswahl des Verfahrens:</i> | Approximativer Binomialtest |
| <i>Voraussetzung</i> | $n > \frac{9}{\pi \cdot (1-\pi)} = 36$, falls nicht erfüllt:
→ Exakter Binomialtest |
| 4. <i>Stichprobenwerte:</i> | $n = 10$ → Exakter Binomialtest |
| 5. <i>Ermittlung der Prüfgröße:</i> | $f = 9$, da $f > n/2$: $f' = n-f = 10-9 = 1$ |

6. *Überschreitungswahrscheinlichkeit:* $P = 2 \cdot W(f \leq 1) = 2 \cdot 0,011 = 0,022$
 7. *Entscheidung:* $0,022 \leq 0,05 \Rightarrow H_0$ ablehnen

Es konnte auf dem 5%-Niveau der Verlässlichkeit bestätigt werden, dass in der untersuchten Berufsgruppe Männer und Frauen nicht gleich verteilt sind.

Angaben: In einer bestimmten Berufsgruppe sollen Frauen unterrepräsentiert sein. Eine Stichprobe von 5 Personen ergab eine Frau. (5%-Signifikanzniveau)

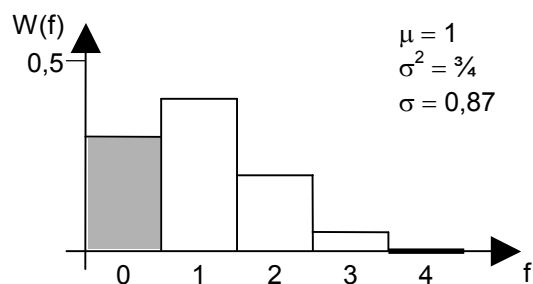
1. *Statistische Hypothesen:* $H_0: \pi \geq \frac{1}{2}$ $H_1: \pi < \frac{1}{2}$
 2. *Festlegung des Signifikanzniveaus:* $\alpha = 0,05$
 3. *Auswahl des Verfahrens:* Approximativer Binomialtest
Voraussetzung $n > \frac{9}{\pi \cdot (1 - \pi)} = 36$, falls nicht erfüllt:
 \rightarrow Exakter Binomialtest
 4. *Stichprobenwerte:* $n = 5 \rightarrow$ Exakter Binomialtest
 5. *Ermittlung der Prüfgröße:* $f = 1$, da $f < \frac{n}{2}$: $f' = f = 1$
 6. *Überschreitungswahrscheinlichkeit:* $P = W(f \leq 1) = 0,188$
 7. *Entscheidung:* $0,188 > 0,05 \Rightarrow H_0$ beibehalten

Es konnte auf einem Signifikanzniveau von 5% nicht bestätigt werden, dass in der untersuchten Berufsgruppe Frauen unterrepräsentiert sind.

6.1.3.2. Asymmetrische Verteilungen

Bei asymmetrischen Verteilungen liegt das Problem vor, dass die benötigten Wahrscheinlichkeiten mit dem Bildungsgesetz zur Binomialverteilung aufgestellt werden müssen und weiterhin, dass die Flächen nicht spiegelbildlich sind:

f	W(f)
0	0,316
1	0,422
2	0,211
3	0,047
4	0,004
Σ	1



Im einseitigen Fall besteht die Überschreitungswahrscheinlichkeit aus der Fläche, die sich von f weg entsprechend der Richtung der Alternativhypothese erstreckt:

- Linksseitiger Test: $P = W(f \leq f') = W(0) + W(1) + \dots + W(f')$
 Rechtsseitiger Test: $P = W(f \geq f') = W(f') + W(f'+1) + \dots + W(n)$

Beispiel: Ergibt sich bei einem linksseitigen Test in einer Stichprobe mit $n=5$ eine Häufigkeit von $f=0$, so lautet die Überschreitungswahrscheinlichkeit dann: $P = W(f \leq 0) = 0,316$. Dieses Ergebnis ist damit auf keinem Signifikanzniveau bedeutsam. Bei einem rechtsseitigen Test gilt für ein Stichprobenergebnis

von $f=3$ eine Überschreitungswahrscheinlichkeit von $P = W(f \geq 3) = 0,047 + 0,004 = 0,051$, was ebenfalls nicht signifikant ist. Wird bei einem rechtsseitigen Test aber eine Häufigkeit von $f=4$ festgestellt, so ergibt sich eine Überschreitungswahrscheinlichkeit von $P = W(f \geq 4) = 0,004$, ein Ergebnis, das auf den Signifikanzniveaus von 5% und 1% bedeutsam ist.

Im zweiseitigen Fall wird es aufgrund der Asymmetrie kompliziert. Im Grunde müsste man beide Flächen einzeln berechnen und addieren. Allerdings soll die Verdopplung einer Hälfte auch hier zu zufriedenstellenden Ergebnissen führen:

$$P = 2 \cdot W(f \leq f'), \text{ falls } f' < \mu$$

$$P = 2 \cdot W(f \geq f'), \text{ falls } f' \geq \mu$$

Beispiel für den exakten Binomialtest bei asymmetrischer Verteilung

Angaben: In einem Multiple-Choice-Test mit drei Aufgaben zu jeweils vier Antwortmöglichkeiten, wovon eine richtig ist, erreichte eine Person 2 Punkte. Ist dieses Ergebnis durch blindes Raten entstanden oder auf das Wissen des Probanden zurückführbar? Signifikanzniveau: 1%

1. **Statistische Hypothesen:** $H_0: \pi \leq 1/4$ $H_1: \pi > 1/4$
(H_1 : Die Wahrscheinlichkeit, dass die Person eine Aufgabe löst, ist größer als $1/4$, wenn er sie aufgrund seines Wissens und nicht zufällig löst.)
2. **Festlegung des Signifikanzniveaus:** $\alpha = 0,01$
3. **Auswahl des Verfahrens:** Approximativer Binomialtest
Voraussetzung $n > \frac{9}{\pi \cdot (1 - \pi)} = 36$, falls nicht erfüllt:
→ Exakter Binomialtest
4. **Stichprobenwerte:** $n = 3$ → Exakter Binomialtest
5. **Ermittlung der Prüfgröße:** $f = 2$
6. **Überschreitungswahrscheinlichkeit:** $P = W(f \geq 2) = W(2) + W(3)$, da rechtsseitig

$$W(2) = \binom{3}{2} \cdot \left(\frac{1}{4}\right)^2 \cdot \left(1 - \frac{1}{4}\right)^{3-2} = 3 \cdot \left(\frac{1}{4}\right)^2 \cdot \frac{3}{4} = 0,141$$

$$W(3) = \binom{3}{3} \cdot \left(\frac{1}{4}\right)^3 \cdot \left(1 - \frac{1}{4}\right)^{3-3} = 1 \cdot \left(\frac{1}{4}\right)^3 \cdot 1 = 0,015$$

$$P = 0,141 + 0,015 = 0,156$$

7. **Entscheidung:** $0,156 \geq 0,01 \Rightarrow H_0$ beibehalten

Es konnte auf dem 1%-Niveau der Verlässlichkeit bestätigt werden, dass das Ergebnis des Tests durch blindes Raten erklärt werden kann.

6.2. Verfahren zur Chi-Quadrat-Verteilung

6.2.1. χ^2 -Unabhängigkeitstest

Der χ^2 -Unabhängigkeitstest dient zur Prüfung, ob der durch Cramer's V ermittelte Zusammenhang in einer Stichprobe signifikant ist, also auf die Grundgesamtheit verallgemeinert werden kann. Für die Berechnung von Cramer's V wird χ^2 benötigt:

$$V = \sqrt{\frac{\chi^2}{n \cdot (\min[J, K] - 1)}}$$

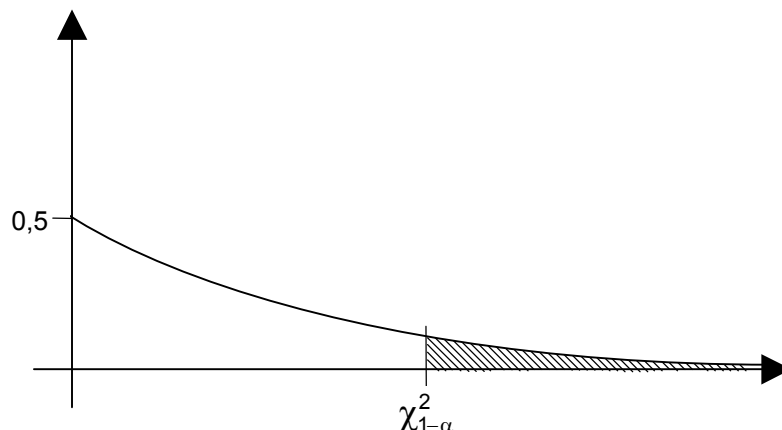
In die Berechnung von χ^2 gehen die erwarteten Häufigkeiten $E(f_{jk})$ für den Fall ein, dass kein Zusammenhang zwischen zwei Merkmalen besteht:

$$E(f_{jk}) = \frac{f_{j\cdot} \cdot f_{\cdot k}}{N} = \frac{\text{Zeilensumme} \cdot \text{Spaltensumme}}{\text{Gesamtsumme}}$$

Die beobachteten Häufigkeiten f_{jk} können im Rahmen der Zufallsstreuung um einen bestimmten Betrag von der erwarteten Häufigkeiten $E(f_{jk})$ abweichen. Diese Abweichung $f_{jk} - E(f_{jk})$ summiert sich in χ^2 :

$$\chi^2 = \sum_j \sum_k \frac{(f_{jk} - E(f_{jk}))^2}{E(f_{jk})}$$

χ^2 -Werte verteilen sich nun in einer Grundgesamtheit, in der die beobachteten den erwarteten Häufigkeiten entsprechen, wie folgt:



Die χ^2 -Verteilung wird benutzt, um die Entscheidung zu fällen, ob der Zusammenhang zwischen zwei Variablen A und B in der Stichprobe signifikant ist, also auch für die Grundgesamtheit angenommen werden kann. Die statistischen Hypothesen sind ungerichtet und lauten dementsprechend:

$$H_0: A, B \text{ unabhängig} \quad H_1: A, B \text{ abhängig}$$

Prüfgröße des Verfahrens ist χ^2 . Die kritische Grenze ist einseitig mit dem Freiheitsgrad $df = (J-1) \cdot (K-1)$. Die kritischen χ^2 -Grenzen können zu den verschiedenen Freiheitsgraden und den zwei Signifikanzniveaus bei HUBER in der χ^2 -Tabelle rechts neben der t-Wert-Tabelle abgelesen werden. Die Nullhypothese wird beibehalten, wenn die Prüfgröße χ^2 unter der kritischen Grenze $\chi^2_{1-\alpha;df}$ bleibt, ein Zusammenhang ist also dann signifikant, wenn die kritische Grenze überschritten wird.

Voraussetzung aller χ^2 -Tests ist, dass alle erwarteten Häufigkeiten größer als 1 sind und dass höchstens $1/5$ der erwarteten Häufigkeiten kleiner als 5 sind.

Beispiel zum χ^2 -Unabhängigkeitstest

Angaben: Es wurde nach dem Zusammenhang zwischen Einkommen und Kaffeesorte gefragt und eine Stichprobe mit 400 Personen gezogen. Die Berechnung von Cramer's V ergab eine äußerst geringe Abhängigkeit von $V = 0,065$. Ist dieser Zusammenhang zufällig? Signifikanzniveau: 1%.

1. *Statistische Hypothesen:* H_0 : A, B unabhängig H_1 : A, B abhängig

2. *Festlegung des Signifikanzniveaus:* $\alpha = 0,01$

3. *Auswahl des Verfahrens:*

χ^2 -Unabhängigkeitstest:

$$\chi^2 = \sum_j \sum_k \frac{(f_{jk} - E(f_{jk}))^2}{E(f_{jk})}; \quad df=(J-1)(K-1)$$

$$E(f_{jk}) = \frac{\text{Zeilensumme} \cdot \text{Spaltensumme}}{\text{Gesamtsumme}}$$

Voraussetzung:

alle $E(f_{jk}) > 1$, max. $1/5$ der $E(f_{jk}) < 5$

4. *Stichprobenwerte:*

Kaffeesorte	Sozialer Status		Σ
	niedrig	hoch	
A	31 [36]	49 [44]	80
B	30 [28,8]	34 [35,2]	64
C	119 [115,2]	137 [140,8]	256
Σ	180	220	400

5. *Prüfgröße:*

$$\begin{aligned} \chi^2 &= \frac{(31-36)^2}{36} + \frac{(49-44)^2}{44} \\ &+ \frac{(30-28,8)^2}{28,8} + \frac{(34-35,2)^2}{35,2} = 1,96 \\ &+ \frac{(119-115,2)^2}{115,2} + \frac{(137-140,8)^2}{140,8} \end{aligned}$$

6. *Kritische Grenze:*

$$\chi^2_{1-\alpha;(J-1)(K-1)} = \chi^2_{0,99;2} = 9,2$$

7. *Entscheidung:*

$1,69 < 9,2 \Rightarrow H_0$ beibehalten

Auf einem Signifikanzniveau von 1% konnte eine signifikante Abhängigkeit zwischen sozialem Status und Kaffeesorte für die Grundgesamtheit nicht bestätigt werden.

6.2.2. χ^2 -Anpassungstest

Der χ^2 -Anpassungstest kann hinsichtlich der Anwendung als Erweiterung des exakten Binomialtests verstanden werden. Während beim Binomialtest Hypothesen über die Auftretenswahrscheinlichkeit von zwei Alternativen in der Grundgesamtheit geprüft werden (dichotomes Merkmal), werden beim χ^2 -Anpassungstest Hypothesen über die Auftretenswahrscheinlichkeit von mehr als zwei Alternativen getestet (nominales Merkmal). Die statistischen Hypothesen für K Alternativen lauten dementsprechend:

$$\begin{array}{ll} H_0: \pi_1 = \pi_{10} & H_1: \pi_1 \neq \pi_{10} \\ \dots & \dots \\ H_0: \pi_K = \pi_{K0} & H_1: \pi_K \neq \pi_{K0} \end{array}$$

Auch in diesem Verfahren ist χ^2 die Prüfgröße. Allerdings gehen in die Formel nicht die erwarteten Häufigkeiten für den Fall ein, dass kein Zusammenhang besteht, sondern die erwarteten Häufigkeiten für den Fall, dass die vermuteten Wahrscheinlichkeiten in der Grundgesamtheit sich als abbildungstreue Häufigkeiten in der Stichprobe widerspiegeln:

$$E(f_k) = n \cdot \pi_{k0}$$

Beispiel: Bei einem idealen Würfel treten alle Augenzahlen gleich wahrscheinlich mit $\pi = 1/6$ auf. Wenn 60mal hintereinander gewürfelt wird, so kann erwartet werden, dass jede Augenzahl mit einer Häufigkeit von $E(f_k) = n \cdot \pi_{k0} = 60 \cdot 1/6 = 10$ auftritt.

Die beobachteten Häufigkeiten können im Rahmen der Zufallsstreuung um einen bestimmten Betrag von den erwarteten Häufigkeiten abweichen. Diese Abweichungen summieren sich in χ^2 :

$$\chi^2 = \sum_k \frac{(f_k - E(f_k))^2}{E(f_k)}$$

Überschreitet χ^2 die kritische Grenze mit einem Freiheitsgrad von $df=K-1$, so weichen die beobachteten Häufigkeiten der einzelnen Alternativen von den erwarteten signifikant ab und die Nullhypothese muss verworfen werden.

Voraussetzung aller χ^2 -Tests ist, dass alle erwarteten Häufigkeiten größer als 1 sind und dass höchstens $1/5$ der erwarteten Häufigkeiten kleiner als 5 sind.

6.2.3. χ^2 - und z-Werte

Das Quadrat einer z-Größe ist eine χ^2 -Größe. Dabei ist zu beachten, dass die z-Größe zweiseitig und die χ^2 -Größe einseitig ist. Der Freiheitsgrad der χ^2 -Größe muss dabei 1 betragen:

$$\left(z_{1-\alpha/2}\right)^2 = \chi_{1-\alpha;1}^2$$

Beispiel zum χ^2 -Anpassungstest

Angaben: Es wird vermutet, dass ein Würfel gezinkt sei. Um das zu klären, wird 60mal hintereinander gewürfelt und die Häufigkeiten des Auftretens jeder Augenzahl notiert. Es wird auf dem 5%-Niveau der Signifikanz geprüft.

1. *Statistische Hypothesen:* $H_0: \pi_k = 1/6$ $H_1: \pi_k \neq 1/6$ (für alle $k=1\dots 6$)

2. *Festlegung des Signifikanzniveaus:* $\alpha = 0,05$

3. *Auswahl des Verfahrens:* χ^2 -Anpassungstest:

$$\chi^2 = \sum_k \frac{(f_k - E(f_k))^2}{E(f_k)}; \quad df = K-1$$

$$E(f_k) = n \cdot \pi_{k0}$$

Voraussetzung

alle $E(f_k) > 1$, max. $1/5$ der $E(f_k) < 5$

4. *Stichprobenwerte:*

Augen	π_k	$n \cdot \pi_{k0}$	f_k	$\frac{(f_k - n \cdot \pi_{k0})^2}{n \cdot \pi_{k0}}$
1	$1/6$	10	15	2,5
2	$1/6$	10	8	0,4
3	$1/6$	10	7	0,9
4	$1/6$	10	10	0
5	$1/6$	10	8	0,4
6	$1/6$	10	12	0,4
Σ	1	(n=60)	(n=60)	4,6

5. *Prüfgröße:* $\chi^2 = 4,6$

6. *Kritische Grenze:* $\chi_{1-\alpha; K-1}^2 = \chi_{1-0,05; 6-1}^2 = \chi_{0,95; 5}^2 = 11,1$

7. *Entscheidung:* $4,6 < 11,1 \Rightarrow H_0$ wird beibehalten

Auf einem Signifikanzniveau von 5% konnte die Hypothese nicht aufrecht erhalten werden, dass der Würfel gezinkt sei.

7. Klausurvorbereitung

7.1. Anleitung

Wiederhole für jeden der folgenden Abschnitte das entsprechende Kapitel, arbeite Deine Aufzeichnungen und Hausaufgaben noch einmal durch und fertige kleine Zusammenfassungen mit den wichtigsten Definitionen, Verfahrensweisen, Formeln und Voraussetzungen an. Bearbeite dann erst die jeweiligen Fragen und Klausuraufgaben. Als Schlussvorbereitung nimmst Du Dir dann Deine Zusammenfassungen noch mal vor, lernst diese und übst anhand der Klausuren den Ernstfall (Zeit: 90 min).

7.1.1. Stichprobenverteilungen

Fragen:

1. Was ist eine Stichprobenverteilung? Und was ist eine Durchschnittsverteilung?
2. Was besagt der zentrale Grenzwertsatz?
3. Wie kann eine Normalverteilung von Stichprobenmittelwerten in eine Standardnormalverteilung transformiert werden?

Klausuraufgaben:

- Klausur 89, Aufgabe 3, Aufgabe 6
- Klausur 90, Aufgabe 3
- Klausur 91, Aufgabe 6
- Klausur 92, Aufgabe 6
- Klausur 96, Aufgabe 6, a) und b)
- Klausur 97, Aufgabe 6

7.1.2. Signifikanztests I

Fragen:

1. Was ist der Unterschied zwischen
 - a) einer inhaltlichen und einer statistischen Hypothese,
 - b) einer Nullhypothese und einer Alternativhypothese,
 - c) einer gerichteten und einer ungerichteten Hypothese?
2. Formuliere zu den folgenden Beispielen die Nullhypothese und die Alternativhypothese, und zwar jeweils sowohl inhaltlich als auch statistisch:
 - a) Es wird aus der Theorie der Verwahrlosung Minderjähriger abgeleitet, dass die Intelligenzleistungen verwahrloster Jugendlicher insbesondere bei solchen Aufgaben unterdurchschnittlich sind, die das Erkennen von ordnenden Strukturen und Redundanzen voraussetzen.
 - b) Es wird behauptet, dass die Population der Blinden durch überdurchschnittliche Fähigkeiten zur akustischen Reizdiskriminierung gekennzeichnet sei.
 - c) Eineiige Zwillinge seien einander ähnlicher als zweieiige.
3. Nenne Beispiele für Entscheidungen, bei denen nach Deiner Ansicht
 - a) ein möglicher α -Fehler schwerwiegender ist als ein möglicher β -Fehler,
 - b) ein möglicher β -Fehler schwerwiegender ist als ein möglicher α -Fehler.

4. In einer Untersuchung möge unter der Annahme einer gültigen H_0 ein Mittelwert von $\mu=80$ erwartet werden. Empirisch ergibt sich jedoch der Wert $\bar{x}=85$. Die Abweichung sei bei zweiseitigem Test auf dem 5%-Niveau signifikant. Wäre die gleiche Abweichung auch bei einseitigem Test signifikant?

Klausuraufgaben:

- Klausur 89, Aufgabe 5
- Klausur 90, Aufgabe 5
- Klausur 91, Aufgabe 5
- Klausur 92, Aufgabe 5 (wie Klausur 89, Aufgabe 5)
- Klausur 94, Aufgabe 5 (wie Klausur 91, Aufgabe 5)
- Klausur 96, Aufgabe 2
- Klausur 97, Aufgabe 5 (wie Klausur 91, Aufgabe 5)

7.1.3. Konfidenzintervalle

Fragen:

1. Wie verändert sich das Konfidenzintervall des Mittelwertes
 - a) bei Vergrößerung des Konfidenzniveaus,
 - b) bei Vergrößerung des Stichprobenumfangs,
 - c) bei Vergrößerung der Populationsstreuung σ_x ?
 - d) bei Vergrößerung der Stichprobenstreuung $\sigma_{\bar{x}}$?
2. Eine Verteilung von $n=200$ Beobachtungen sei durch $\bar{x}=100$ und $s=10$ gekennzeichnet. Wie lautet das Konfidenzintervall des Mittelwertes für
 - a) eine Sicherheitswahrscheinlichkeit von 95%,
 - b) einen Konfidenzkoeffizienten von 99%?
3. In einem Konditionierungsexperiment lernen Hunde, auf ein akustisches Signal hin einen Mechanismus zu bedienen, um Futter zu erhalten. Nach einer einwöchigen Trainingsphase zeigen die Hunde in 200 Versuchen 160mal das gewünschte Verhalten. Wie lautet das 95%ige Konfidenzintervall?

Klausuraufgaben:

- Klausur 89, Aufgabe 4
- Klausur 90, Aufgabe 1, Aufgabe 6
- Klausur 91, Aufgabe 1
- Klausur 92, Aufgabe 1, Aufgabe 3
- Klausur 93, Aufgabe 5 (wie Klausur 90, Aufgabe 6)
- Klausur 94, Aufgabe 4
- Klausur 96, Aufgabe 6 (wie Klausur 90, Aufgabe 6)
- Klausur 97, Aufgabe 4

7.1.4. Signifikanztests II

Fragen:

1. Welcher Korrelationskoeffizient kann mit dem uns bekannten Verfahren zur Prüfung von Korrelationskoeffizienten getestet werden?
2. Welche sind die für das psychologische Experiment wichtigsten Signifikanztests?
3. Was kennzeichnet ein Experiment?

4. Was ist der Unterschied zwischen abhängigen und unabhängigen Stichproben?
5. Was sind die Voraussetzungen des klassischen t-Tests?
6. Ist der klass. t-Test oder der WELCH-Test zu bevorzugen?
7. Wie wird einer der Voraussetzungen zum klass. t-Test überprüft?
8. 12 Kinder reicher Eltern und 12 Kinder armer Eltern werden aufgefordert, den Durchmesser eines Markstücks zu schätzen. Folgende Schätzungen wurden abgegeben (in mm):
 - reich: 20, 23, 23, 21, 22, 25, 19, 24, 20, 26, 24, 25
 - arm: 24, 23, 26, 28, 27, 27, 25, 18, 21, 26, 25, 29
 Überprüfe mit Hilfe des kl. t-Testes (Voraussetzung!), ob die durchschnittlichen Schätzwerte der armen Kinder signifikant größer sind als die der reichen Kinder!
9. Nach einer Untersuchung von MILLER und BUGELSKI (1948) ist zu erwarten, dass Personen in ihren Einstellungen gegenüber neutralen Personen negativer werden, wenn sie zwischenzeitlich frustriert wurden (Sündenbockfunktion). Für 9 Jungen mögen sich vor und nach einer Frustration folgende Einstellungswerte ergeben haben:
 - vorher: 38, 32, 33, 28, 29, 37, 35, 35, 34
 - nachher: 33, 28, 34, 26, 27, 31, 32, 36, 30
 Sind die registrierten Einstellungsänderungen statistisch signifikant, wenn man davon ausgeht, dass die Einstellungen normalverteilt sind?

Klausuraufgaben:

- Klausur 89, Aufgabe 2
- Klausur 90, Aufgabe 2, Aufgabe 4
- Klausur 91, Aufgabe 2*, Aufgabe 4
- Klausur 92, Aufgabe 2
- Klausur 93, Aufgabe 2 (wie Klausur 90, Aufgabe 2), Aufgabe 4
- Klausur 94, Aufgabe 1
- Klausur 96, Aufgabe 1
- Klausur 97, Aufgabe 1

7.1.5. Varianzanalysen

Fragen:

1. Wann werden Varianzanalysen eingesetzt und welchen Stellenwert haben sie in der Experimentalpsychologie?
2. Was ist der Unterschied zwischen einer ein- und einer mehrfaktoriellen Varianzanalyse?
3. Welche H_0 wird mit der einfaktoriellen Varianzanalyse überprüft?
4. Was versteht man unter einer Fehlervarianz?
5. In welche Anteile wird die Gesamtvarianz SAQ_G in einer einfaktoriellen Varianzanalyse zerlegt?
6. Was bedeuten die Begriffe Varianzaufklärung und Effektstärke?
7. Durch welches Verfahren kann die Differenz zwischen zwei einzelnen Stichprobenmittelwerten innerhalb der Varianzanalyse auf Signifikanz geprüft werden?
8. Was ist bei der einfaktoriellen Varianzanalyse für abhängige Stichproben hinsichtlich der kritischen Grenze zu berücksichtigen?

* Zur Übung sollte hier mal der klassische t-Test gerechnet werden (Anmerkung des Verfassers).

9. Welchen versuchstechnischen Grund kann es geben, eine zweifaktorielle Varianzanalyse durchzuführen, obwohl nur der Effekt eines Faktors interessiert?
10. Was ist der Unterschied zwischen einem Haupteffekt und einem einfachen Haupteffekt bei der zweifaktoriellen Varianzanalyse?
11. Was ist eine Wechselwirkung und welche Typen gibt es? Worin unterscheiden sich diese Typen?
12. Welche Voraussetzungen müssen bei Varianzanalysen gegeben sein? Wie kann man die Prüfung der Voraussetzungen umgehen?
13. 4 Stichproben á 20 Vpn werden varianzanalytisch untersucht. Wie müssten die Daten der Vpn aussehen, damit folgende F-Werte resultieren?
 - a) Die F-Prüfgröße ist null: $F = 0$.
 - b) Die F-Prüfgröße ist unendlich groß: $F \rightarrow \infty$.

Klausuraufgaben:

- Klausur 93, Aufgabe 3, Aufgabe 6
- Klausur 94, Aufgabe 3
- Klausur 96, Aufgabe 3 und 4
- Klausur 97, Aufgabe 3

7.1.6. Nichtparametrische, verteilungsfreie Verfahren

Fragen:

1. Auf welchen Skalenniveaus werden nichtparametrische Verfahren eingesetzt?
2. Welche Verteilungen liegen den verteilungsfreien Verfahren zugrunde?
3. Was ist eine Binomialverteilung? Wie lautet das Bildungsgesetz, um eine Binomialverteilung aufstellen zu können? Wie lassen sich der Erwartungswert und die Varianz einer Binomialverteilung berechnen?
4. Wann wird ein Binomialtest durchgeführt? Welche Alternativen gibt es hierzu? Welche Voraussetzungen müssen jeweils erfüllt sein?
5. Was ist eine χ^2 -Verteilung? Welche Tests zur χ^2 -Verteilung gibt es und wann werden sie angewendet? Welche Voraussetzungen müssen für die Durchführung eines χ^2 -Tests erfüllt sein?
6. GLEISS et al. (1973) berichten über eine Auszählung, nach der eine Stichprobe von 450 neurotischen Patienten mit folgenden (geringfügig modifizierten) Häufigkeiten in folgenden Therapiearten behandelt wurden:

• Klassische Analyse und analytische Psychotherapie	82
• Direkte Psychotherapie	276
• Gruppenpsychotherapie:	15
• Somatische Behandlung:	48
• Custodial care:	29

Überprüfe die Hypothese, dass sich die 450 Patienten auf die 5 Therapieformen gleich verteilen!

Klausuraufgaben:

- Klausur 89, Aufgabe 1, Aufgabe 6
- Klausur 92, Aufgabe 4
- Klausur 93, Aufgabe 1
- Klausur 94, Aufgabe 2, Aufgabe 6
- Klausur 96, Aufgabe 5, c)
- Klausur 97, Aufgabe 2

7.2. Klausuren

7.2.1. Klausur 1989

Aufgabe 1: Es wurde behauptet, dass zwischen Bildungsniveau und der Einstellung zu einem bestimmten politischen Thema (PRO oder CONTRA) ein signifikanter Zusammenhang besteht. Bei 60 zufällig ausgewählten Personen wurden folgende Ergebnisse festgestellt:

	PRO	CONTRA
Niedriges Bildungsniveau	5	10
Mittleres Bildungsniveau	26	4
Hohes Bildungsniveau	9	6

Ist damit obige Behauptung nachgewiesen? (Signifikanzniveau: 5%)

Aufgabe 2: Formulieren Sie ein konkretes inhaltliches Beispiel, wo ein Signifikanztest für den Vergleich zweier Mittelwerte (unabhängige Stichproben) sinnvoll sein könnte! Führen Sie den Signifikanztest mit fiktiven Daten durch!

Aufgabe 3: Der Anteil einer bestimmten Merkmalsausprägung in einer Grundgesamtheit betrage $\pi=0,2$. Es wird eine Stichprobe vom Umfang 60 gezogen und der entsprechende Anteil p berechnet.

- Welche Werte kann p annehmen?
- Bestimmen Sie die Wahrscheinlichkeitsverteilung von p !
- Zeichnen Sie diese Verteilung!
- Sind die nötigen Voraussetzungen erfüllt?

Aufgabe 4: In welchem Bereich liegt der „wahre“ IQ einer Person, bei der ein IQ von 115 gemessen wurde? Sicherheitswahrscheinlichkeit: 95%. Von den IQ's wird angenommen, dass sie um den „wahren“ IQ normalverteilt sind mit einer Standardabweichung von 5.

Aufgabe 5: Wie heißen die Fehler, die bei einem Signifikanztest auftreten können? Worin bestehen diese? Wovon hängen die zugehörigen Fehlerwahrscheinlichkeiten ab?

Aufgabe 6: Wie verändern sich bei zunehmenden Freiheitsgrad

- die t-Verteilungen,
- die χ^2 -Verteilungen?

In welchem Zahlenbereich liegen χ^2 -verteilte Größen?

7.2.2. Klausur 1990

Aufgabe 1: Um das Durchschnittsalter einer Bevölkerungsgruppe zu bestimmen, wurden aus ihr 150 Personen zufällig ausgewählt und für diese ein Mittelwert von 37 Jahren und eine Standardabweichung von 12 Jahren berechnet. Wie groß ist das Durchschnittsalter der ganzen Bevölkerungsgruppe? (Sicherheitswahrscheinlichkeit: 99%)

Aufgabe 2: Formulieren Sie ein konkretes inhaltliches Beispiel, wo ein Signifikanztest für den Vergleich zweier Mittelwerte (verbundene Stichproben) sinnvoll sein könnte! Führen Sie den Signifikanztest mit fiktiven Daten durch!

Aufgabe 3: Aus einer Gesamtheit von Personen, die bezüglich einer bestimmten Testleistung normalverteilt ist mit einem Mittelwert von 50 und einer Standardabweichung von 10, werden zufällig zwei Personen ausgewählt („mit Zurücklegen“). Wie groß ist die Wahrscheinlichkeit, dass

- a) beide einen Testwert von über 60 haben,
- b) mindestens eine von beiden einen Testwert von über 60 hat,
- c) der Durchschnitt von beiden Testwerten über 60 liegt?

Aufgabe 4: Es wurde behauptet, dass zwischen zwei Merkmalen ein signifikanter Zusammenhang besteht. Bei einer Stichprobe vom Umfang $n=70$ wurde ein Korrelationskoeffizient $r=0,20$ gefunden. Ist obige Behauptung richtig? (Signifikanzniveau: 5%)

Aufgabe 5: Stellen Sie

a) für den Test $H_0: \mu=\mu_0; H_1: \mu\neq\mu_0$

b) für den Test $H_0: \mu\geq\mu_0; H_1: \mu<\mu_0$

die Wahrscheinlichkeitsverteilung der Testgrößen, die kritische(n) Grenze(n) und das Signifikanzniveau α grafisch dar! (σ wird als bekannt, n als hinreichend groß angenommen.)

Aufgabe 6: Es soll eine Aussage über einen Parameter der Grundgesamtheit mittels eines Konfidenzintervalls gemacht werden. Von welchen Faktoren hängt die Präzision einer solchen Aussage ab? Wie kann die Präzision erhöht werden?

7.2.3. Klausur 1991

Aufgabe 1: Von 80 zufällig ausgewählten Studierenden sagten 60, dass sie mit ihrer Studiensituation zufrieden seien. Zwischen welchen Grenzen liegt der Anteil der mit ihrer Studiensituation zufriedenen Studierenden in der zugehörigen Grundgesamtheit? (Sicherheitswahrscheinlichkeit: 95%)

Aufgabe 2: Eine Gruppe von 60 Kindern wurde nach Lehrmethode A unterrichtet, eine andere gleich große Gruppe nach Methode B. Bei einem anschließenden Leistungstest ergab sich für die Methode A eine durchschnittliche Punktzahl von 38 bei einer Standardabweichung von 6, bei der Gruppe B eine durchschnittliche Punktzahl von 43 bei einer Standardabweichung von 7. (Signifikanzniveau: 1%)

Aufgabe 3: Von einer Maschine sei bekannt, wie viele Unfälle mit welchen Wahrscheinlichkeiten bei ihrer Bedienung pro Jahr passieren, und zwar:

Zahl der Unfälle	Wahrscheinlichkeit
0	0,3
1	0,4
2	0,2
3	0,1

Bestimmen sie
 (a) den Erwartungswert und
 (b) die Standardabweichung der Unfallzahl!

Aufgabe 4: Formulieren sie ein konkretes inhaltliches Beispiel, wo ein Signifikanztest für den PEARSONSchen Korrelationskoeffizienten (1-seitige Fragestellung) zur Anwendung kommen könnte! Führen Sie den Signifikanztest mit fiktiven Daten durch!

Aufgabe 5: Kreuzen Sie bitte bei jeder der folgenden Aussagen zu Signifikanztests an, ob Sie sie für richtig oder für falsch halten!

	<u>richtig</u>	<u>falsch</u>
a) Der Fehler 1. Art besteht darin, dass eine richtige H_0 nicht abgelehnt wird.	<input type="radio"/>	<input type="radio"/>
b) Der Fehler 2. Art besteht darin, dass eine richtige Nullhypothese abgelehnt wird.	<input type="radio"/>	<input type="radio"/>
c) Der Fehler 2. Art besteht darin, dass eine falsche H_0 nicht abgelehnt wird.	<input type="radio"/>	<input type="radio"/>
d) Je größer α , desto größer β	<input type="radio"/>	<input type="radio"/>
e) Je kleiner σ , desto kleiner β	<input type="radio"/>	<input type="radio"/>
f) Je kleiner n , desto größer α	<input type="radio"/>	<input type="radio"/>
g) je größer n , desto kleiner β	<input type="radio"/>	<input type="radio"/>
h) α und β ergänzen sich zu 1	<input type="radio"/>	<input type="radio"/>

Aufgabe 6:

- a) Zeichnen Sie die Dichtekurve der Standardnormalverteilung und die Dichtekurve einer t-Verteilung!
- b) Worin unterscheiden sich t-Verteilungen von der Standardnormalverteilung?
- c) Wie verändern sich t-Verteilungen mit zunehmenden Freiheitsgrad?
- d) Nennen Sie t-verteilte Zufallsgrößen!

7.2.4. Klausur 1992

Aufgabe 1: Bei einem Diagnostiker wurde festgestellt, dass er in 68 von 80 Fällen eine richtige Diagnose stellte. Zwischen welchen Grenzen liegt der Anteil der richtigen Diagnosen in der zugehörigen Grundgesamtheit?
(Sicherheitswahrscheinlichkeit: 95%)

Aufgabe 2: In einem bestimmten Land wurde das Fernsehprogramm erweitert. Es wurde vermutet, dass dadurch die tägliche Einschaltzeiten verlängert würden. Um diese Vermutung zu prüfen, wurden 40 zufällig ausgewählte Haushalte sowohl vor als auch nach der Programmerweiterung befragt. Bei Ihnen ergab sich vor der Programmerweiterung ein Durchschnitt von 6 Stunden, nach der Programmerweiterung ein Durchschnitt von 5,5 Stunden. Für die Zeitdifferenzen ergab sich eine Standardabweichung von 1,5 Stunden. Kann damit obige Vermutung statistisch nachgewiesen werden? (Signifikanzniveau: 5%)

Aufgabe 3: Angenommen, es würden zur Bestimmung des Mittelwertes μ einer Grundgesamtheit unabhängig voneinander zwei Konfidenzintervalle berechnet, jeweils mit einer Sicherheitswahrscheinlichkeit von 0,95. Wie groß ist dann die Wahrscheinlichkeit, dass

- alle beide Intervalle μ enthalten
- keines der Intervalle μ enthält
- mindestens eines der Intervalle μ nicht enthält?

Aufgabe 4: Formulieren Sie ein konkretes inhaltliches Beispiel, wo ein χ^2 -Test zur Anwendung kommen könnte! Führen Sie den Signifikanztest mit fiktiven Daten durch!

Aufgabe 5: Wie heißen die Fehler, die bei einem Signifikanztest auftreten können? Worin bestehen diese? Wovon hängen die zugehörigen Fehlerwahrscheinlichkeiten ab?

Aufgabe 6: Was versteht man unter den Begriffen Grundgesamtheit, Stichprobe, Stichprobenverteilung? Wie sieht die Stichprobenverteilung des arithmetischen Mittels aus? (Verteilungsform, Erwartungswert, Standardabweichung)

7.2.5. Klausur 1993

Aufgabe 1: Von einer bestimmten Heilmethode wird behauptet, dass mit ihr signifikant weniger als die Hälfte der behandelten Patienten geheilt werden. Prüfen Sie diese Behauptung auf dem 5%-Signifikanzniveau, wenn Sie wissen, dass von 18 behandelten Personen 6 geheilt wurden!

Aufgabe 2: Formulieren Sie ein konkretes inhaltliches Beispiel, wo ein Signifikanztest für den Vergleich zweier Mittelwerte (abhängige Stichproben) sinnvoll sein könnte! Führen Sie den Signifikanztest mit fiktiven Daten durch!

Aufgabe 3: Um zu untersuchen, ob sich drei Diäten signifikant in ihrer Wirkung unterscheiden, wurde 15 Personen per Zufall je einer Diät zugewiesen und die Gewichtsabnahme in kg festgestellt:

Diät 1	Diät 2	Diät 3
10	12	10
5	10	8
3	15	12
7	20	14
5	13	6

Führen Sie auf dem 1%-Signifikanzniveau, ob sich die drei Diäten signifikant unterscheiden!

Aufgabe 4: Es wurde behauptet, dass der Zusammenhang zwischen zwei Merkmalen positiv ist. Bei einer Stichprobe von 80 Personen wurde ein Korrelationskoeffizient $r=0,20$ gefunden.

Ist obige Behauptung richtig? (Signifikanzniveau: 5%)

Aufgabe 5: Es soll eine Aussage über einen Parameter der Grundgesamtheit mittels eines Konfidenzintervalls gemacht werden. Von welchen Faktoren hängt die Präzision einer solchen Aussage ab? Wie kann die Präzision erhöht werden?

Aufgabe 6: Beantworten Sie folgende Fragen zur zweifaktoriellen Varianzanalyse:

- Welche Voraussetzungen müssen gegeben sein, damit eine solche überhaupt durchgeführt werden kann?
- Wie kann erreicht werden, dass eine Verletzung dieser Voraussetzungen nicht sehr gravierend sind?
- Was versteht man unter einer Wechselwirkung?

7.2.6. Klausur 1994

Aufgabe 1: Um die Behauptung zu prüfen, dass von zwei Lehrmethoden die Methode 2 zu höheren Leistungen führt, wurden 100 Kinder durch Zufall auf die beiden Lehrmethoden aufgeteilt (je 50 Kinder). Bei einem abschließenden Leistungstest ergab sich für die Methode 1 eine durchschnittliche Punktzahl von 38 bei einer Standardabweichung von 6, für die Methode 2 eine durchschnittliche Punktzahl von 41 bei einer Standardabweichung von 9. kann man damit obige Behauptung nachweisen? (Signifikanzniveau: 1%)

Aufgabe 2: Bei 400 Urlaubern wurde festgestellt, ob sie ihren Urlaub im Ausland oder Inland verbringen und welches Bildungsniveau sie aufweisen:

	niedriges Bildungsniveau	hohes Bildungsniveau
Ausland	102	98
Inland	58	142

Prüfen Sie, ob in der zugehörigen Grundgesamtheit ein Zusammenhang zwischen den beiden Merkmalen besteht! (Signifikanzniveau: 1%)

Aufgabe 3: Die Leistungen von 4 Schülern wurden zu drei verschiedenen Zeitpunkten gemessen:

Schüler	Zeitpunkt 1	Zeitpunkt 2	Zeitpunkt 3
1	7	9	14
2	10	10	19
3	12	10	14
4	5	9	13

Gibt es signifikante Leistungsunterschiede zwischen den drei Zeitpunkten? (Signifikanzniveau: 5%; Hinweis: SAQ_G beträgt 150)

Aufgabe 4: In welchem Bereich liegt der „wahre“ IQ einer Person, bei der ein IQ von 105 gemessen wurde? Sicherheitswahrscheinlichkeit: 95%. Von den IQ's wird angenommen, dass sie um den „wahren“ IQ normalverteilt sind mit einer Standardabweichung von 5.

Aufgabe 5: Kreuzen Sie bitte bei jeder der folgenden Aussagen zu Signifikanztests an, ob Sie sie für richtig oder für falsch halten! (s. Klausur 91, Aufgabe 5)

Aufgabe 6:

- a) Zeichnen Sie die Dichtekurve der Standardnormalverteilung und die Dichtekurve einer t-Verteilung!
- b) Worin unterscheiden sich t-Verteilungen von der Standardnormalverteilung?
- c) Wie ist das Quadrat einer normalverteilten z-Größe verteilt?
- d) Wie ist das Quadrat einer t-Größe mit $df=10$ verteilt?

7.2.7. Klausur 1996

Aufgabe 1: Um die Behauptung zu prüfen, dass eine bestimmte Schlankheitsdiät eine signifikante Gewichtsveränderung zur Folge hat, wurde bei 25 Personen das Gewicht vor einer entsprechenden Kur (Durchschnitt: 90 kg) sowie nach der Kur (Durchschnitt: 92 kg) gemessen. Die Standardabweichung der Differenzen betrug 4 kg. Ist obige Behauptung richtig? (Signifikanzniveau: 1%)

Aufgabe 2: Es wurde die Behauptung aufgestellt, dass mehr Leute in der Stadt leben wollen als auf dem Land. Bei einer Befragung von 80 Personen sprachen sich 52 für die Stadt und nur 28 für das Land aus. Gilt damit obige Behauptung auch für die zugehörige Grundgesamtheit? (Signifikanzniveau: 5%)

Aufgabe 3: Im Rahmen einer zweifaktoriellen Varianzanalyse wurden folgende SAQ's ermittelt: $SAQ_A = 20$; $SAQ_B = 45$; $SAQ_{A \times B} = 5$; $SAQ_I = 40$. Die beiden Faktoren hatten je zwei Faktorstufen. Jeder Faktorstufenkombination lag eine Stichprobe von $n=5$ Elementen zugrunde. Prüfen Sie auf dem 5%-Signifikanzniveau, ob es signifikante Haupteffekte der beiden Faktoren sowie eine signifikante Wechselwirkung gibt!

Aufgabe 4: Bestimmen Sie für die vorige Aufgabe die Effektstärken η^2 und interpretieren Sie diese Zahlen!

Aufgabe 5:

- a) Zeichnen Sie die Dichtekurve der Standardnormalverteilung und die Dichtekurve einer t-Verteilung!
- b) Wie verändern sich t-Verteilungen mit zunehmenden Freiheitsgrad?
- c) Wie verändern sich χ^2 -Verteilungen mit zunehmenden Freiheitsgrad?

Aufgabe 6: Es soll eine Aussage über einen Parameter der Grundgesamtheit mittels eines Konfidenzintervalls gemacht werden. Von welchen Faktoren hängt die Präzision einer solchen Aussage ab? Wie kann die Präzision erhöht werden?

7.2.8. Klausur 1997

Aufgabe 1: Bei 2 Gruppen von Personen wurden Leistungswerte bestimmt und daraus Standardabweichungen berechnet. Dabei ergab sich bei der 1. Gruppe mit 25 Personen eine Standardabweichung von 8, bei der 2. Gruppe mit 30 Personen eine Standardabweichung von 6. Überprüfen Sie auf dem 10%-Signifikanzniveau, ob sich die Varianzen der beiden Gruppen signifikant unterscheiden! Nennen Sie auch die dabei unterstellten Voraussetzungen!

Aufgabe 2: Es wurde behauptet, dass es in einer bestimmten Bevölkerungsgruppe weniger Raucher als Nichtraucher gibt. Bei einer Zufallsauswahl von 24 Personen wurden nur 9 Raucher, aber 15 Nichtraucher gefunden. Gilt damit obige Behauptung für die ganze Bevölkerungsgruppe? (Signifikanzniveau: 5%)

Aufgabe 3: Die Leistungen von 3 Schülern wurden zu drei verschiedenen Zeitpunkten gemessen:

Schüler	Zeitpunkt 1	Zeitpunkt 2	Zeitpunkt 3
1	8	12	16
2	11	12	19
3	11	12	16

Gibt es signifikante Mittelwertunterschiede zwischen den 3 Zeitpunkten? (Signifikanzniveau: 5%) Hinweis: SAQ_G beträgt: 90

Aufgabe 4: Um in einem bestimmten Land die Durchschnittsgröße der Wohnungen zu bestimmen, wurden 400 Haushalte zufällig ausgewählt und für diese ein Mittelwert von 80 qm und eine Standardabweichung von 20 qm berechnet. Wie groß ist die durchschnittliche Wohngröße aller Haushalte dieses Landes? (Sicherheitswahrscheinlichkeit: 99%)

Aufgabe 5: Kreuzen Sie bitte bei jeder der folgenden Aussagen zu Signifikanztests an, ob Sie sie für richtig oder für falsch halten! (s. Klausur 91, Aufgabe 5)

Aufgabe 6: Eine Gesamtheit von Personen sei hinsichtlich ihrer IQ's normalverteilt mit einem Mittelwert von 100 und einer Standardabweichung von 15.

- Wie groß ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person einen IQ zwischen 85 und 115 hat?
- Wie groß ist die Wahrscheinlichkeit, dass der Durchschnitts-IQ von 9 zufällig ausgewählten Personen zwischen 85 und 115 liegt?

Stellen Sie die Wahrscheinlichkeiten auch grafisch dar!

7.3. Lösungen zu den Fragen

7.3.1. Stichprobenverteilungen

1. Eine Stichprobenverteilung ist die Wahrscheinlichkeitsverteilung eines statistischen Kennwertes (Maß der zentralen Tendenz: Arithmetisches Mittel, Median; Streuungsmaß: Standardabweichung, Varianz, durchschnittliche Abweichung etc.) von allen möglichen Stichproben (mit dem Umfang n), die aus einer Grundgesamtheit gezogen werden können.
2. Eine Durchschnittsverteilung ist eine Stichprobenverteilung des Mittelwertes, d.h. die Wahrscheinlichkeitsverteilung der an der Gesamtheit aller möglichen Zufallsstichproben (des Umfangs n) erhaltenen Mittelwerte.
3. Der zentrale Grenzwertsatz besagt, dass die Verteilung von Mittelwerten aus Stichproben gleichen Umfangs n , die aus derselben Population stammen, bei wachsendem Stichprobenumfang n in eine Normalverteilung übergeht. Diese Aussage ist unabhängig von der Verteilungsform der Messwerte in der Grundgesamtheit.
4. Mit der z-Transformation (wie jede andere Normalverteilung auch).

7.3.2. Signifikanztests I

1. a) Statistische Hypothesen betreffen Aussagen, die über statistische Kennwerte einer Stichprobe aufgestellt werden. Von den statistischen Hypothesen zu unterscheiden sind die inhaltlichen Hypothese, die der Forscher über seinen Untersuchungsgegenstand aufgestellt hat. Die inhaltliche Hypothese ist in die statistischen Hypothesen H_0 und H_1 umzusetzen.
 - b) Die zu prüfende statistische Hypothese nennt man *Nullhypothese* und besagt folgendes: *Die untersuchte Stichprobe entstammt aus einer Grundgesamtheit mit der in der Nullhypothese formulierten Aussage.* Die der Nullhypothese entgegengesetzte Hypothese nennt man *Alternativhypothese*, und sie besagt: *Die untersuchte Stichprobe entstammt nicht aus einer Grundgesamtheit mit der in der Nullhypothese formulierten Aussage.*
 - c) Gerichtete Hypothesen geben die Richtung des behaupteten Zusammenhangs oder Unterschiedes vor, ungerichtete nicht. Eine gerichtete Hypothese wird mit einem einseitigen, eine ungerichtete mit einem zweiseitigen Test geprüft.
2. a) *Nullhypothese*: Die Intelligenz von verwahrlosten Jugendlichen (μ_0) ist genauso hoch oder höher als die Intelligenz von nicht verwahrlosten Jugendlichen (μ_1):
 $H_0: \mu_0 \geq \mu_1$
Alternativhypothese: Die Intelligenz von verwahrlosten Jugendlichen (μ_0) ist niedriger als die Intelligenz v. nicht verwahrlosten Jugendlichen (μ_1): $H_1: \mu_0 < \mu_1$
 - b) *Nullhypothese*: Die Fähigkeit zur akust. Reizdiskriminierung ist bei Blinden (μ_0) genauso hoch bzw. geringer als bei nicht blinden Personen (μ_1): $H_0: \mu_0 \leq \mu_1$
Alternativhypothese: Die Fähigkeit zur akustischen Reizdiskriminierung ist bei Blinden (μ_0) höher als bei nicht blinden Personen (μ_1): $H_1: \mu_0 > \mu_1$
 - c) *Nullhypothese*: Unterschiede zwischen eineiigen Zwillingen (μ_0) sind genauso hoch bzw. kleiner als bei zweieiigen Zwillingen (μ_1): $H_0: \mu_0 \leq \mu_1$
Alternativhypothese: Unterschiede zwischen eineiigen Zwillingen (μ_0) sind größer als zwischen zweieiigen Zwillingen (μ_1): $H_1: \mu_0 > \mu_1$

3. a) Das Kultusministerium vergibt den Auftrag, die Wirksamkeit einer neuen Unterrichtsmethode zu prüfen. $H_0: \mu_0 \geq \mu_1$ (die Unterrichtsmethoden unterscheiden sich nicht bzw. die neue Methode (μ_1) ist schlechter), $H_1: \mu_0 < \mu_1$ (die neue Unterrichtsmethode (μ_1) ist besser als eine herkömmliche Unterrichtsmethode). Hier besagt der α -Fehler, dass fälschlicherweise angenommen wird, dass die neue Lehrmethode besser sei als die alte Methode, was für das Ministerium die schwerwiegendere Konsequenz hat, dass die Neuanschaffung von Lehrmaterial, Umschulung der Lehrer, Neugestaltung der Unterrichtspläne etc. rausgeworfenes Geld bedeuten. Die Konsequenz eines für das Ministerium geringeren β -Fehler wäre, dass zwar keine Fehlinvestitionen riskiert werden, aber die Chance, den Unterricht zu verbessern.
- b) Es wird ein Zusammenhang zwischen gesteigertem Fernsehkonsum und verminderter Konzentrationsfähigkeit vermutet. $H_0: \rho \geq 0$ (zwischen der Dauer des Fernsehens und der Konzentrationsfähigkeit besteht kein oder sogar ein positiver Zusammenhang), $H_1: \rho < 0$ (mit zunehmender Dauer des Fernsehens sinkt die Konzentrationsfähigkeit). Ein α -Fehler hätte zur Konsequenz, dass Erziehungsberater Eltern, die sie aufgrund von Konzentrationsschwächen ihrer Kinder aufsuchen, empfehlen, die Fernsehzeiten der Kinder einzuschränken. Diese Maßnahme wird zwar die Konzentrationsfähigkeit der Kinder nicht verbessern, sie dürfte aber darüber hinaus keine erheblichen negativen Konsequenzen auf die Kinder haben. Anders sähe es beim β -Fehler aus: Die Erziehungsberater werden den Eltern mitteilen, dass die Konzentrationsschwächen ihrer Kinder nichts mit dem Fernsehen zu tun haben, was zur Folge hat, dass die Konzentrationsschwächen der Kinder weiter zunehmen.
4. Ja, wenn die Alternativhypothese $H_1: \mu_1 > \mu$ heißt.

7.3.3. Konfidenzintervalle

1. a) Es wird größer (unpräziser, ungenauer).
 b) Es wird kleiner (präziser, genauer).
 c) Es wird größer (unpräziser, ungenauer).
 d) Es wird größer (unpräziser, ungenauer).
2. a) $100 \pm 1,96 \cdot \sqrt{1/2} \Rightarrow 98,61 \leq \mu \leq 101,39$
 b) $100 \pm 2,58 \cdot \sqrt{1/2} \Rightarrow 98,18 \leq \mu \leq 101,82$
3. $74,46\% \leq \pi \leq 85,54$

7.3.4. Signifikanztests II

1. Der Produkt-Moment-Korrelationskoeffizient von PEARSON.
2. Es sind die Verfahren zur Prüfung von Unterschiedshypothesen, insbesondere der WELCH-Test und der klassische t-Test.
3. Bei einem Experiment variiert der Versuchsleiter systematisch einen Faktor, um seinen Effekt auf die Messvariable zu ermitteln, während er versucht, Störeinflüsse zu minimieren.
4. Der Unterschied besteht darin, dass unabhängige/unverbundene Stichproben nicht aus denselben Versuchspersonen bestehen und auch nicht parallelisiert sind.

5. unabhängige Stichproben sowie varianzhomogene und normalverteilte Grundgesamtheiten (s.S. 43)
6. Der WELCH-Test ist dem klass. t-Test zu bevorzugen (obwohl der klass. t-Test eine größere Teststärke besitzt), da der klass. T-Test aufgrund seiner Voraussetzungen problematisch ist.
7. Zur Prüfung der Varianzhomogenität der beiden Stichproben zugrundeliegenden Populationen wird der F-Test verwendet.
8. 1. $H_0: \mu_1 - \mu_2 \geq 0$ $H_1: \mu_1 - \mu_2 < 0$ (μ_1 reich)
2. $\alpha = 0,01/0,05$
3. Klassischer t-Test

Prüfung auf Gleichheit der Varianzen (Vor. zum klass. t-Test):

1. $H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$
2. $\alpha = 0,10$
5. $F = 1,85$
6. $F_{0,95;11;11} = 2,85$
7. $1,85 < 2,85 \Rightarrow H_0$ beibehalten \Rightarrow Varianzen sind homogen
 \Rightarrow Voraussetzung für den klassischen t-Test ist erfüllt

4. $n_1 = 11, \bar{x}_1 = 22,67, s_1 = 2,27; n_2 = 11, \bar{x}_2 = 24,92, s_2 = 3,09$
 5. $t = -2,03$
 6. $t_{0,99;22} = -2,51$ oder $t_{0,95;22} = -1,72$
 7. 1%: $-2,03 > -2,85 \Rightarrow H_0$ beibehalten
 5%: $-2,03 < -1,72 \Rightarrow H_1$ beibehalten
- Es konnte nur auf dem 5%-Niveau die inhaltliche Hypothese beibehalten werden, dass arme Kinder Markstücke in ihrer Größe besser schätzen können als reiche.
9. 1. $H_0: \mu_1 - \mu_2 \leq 0$ $H_1: \mu_1 - \mu_2 > 0$ (μ_1 vorher)
 2. $\alpha = 0,01/0,05$
 3. z-Test für abhängige Stichproben (Formel (4)-(I), STAT -9-)
 4. $\bar{d} = 2,67, s_d = 2,45, n = 9$
 5. $t = 3,26$
 6. $t_{0,99;8} = 2,90$ oder $t_{0,95;8} = 1,86$
 7. 1%: $3,26 > 2,90 \Rightarrow H_1$ beibehalten
 5%: $3,26 > 1,86 \Rightarrow H_1$ beibehalten
- Es konnte auf beiden Signifikanzniveaus bestätigt werden, dass nach zwischenzeitlichen Frustrationen neutrale Personen negativer bewertet werden als vorher.

7.3.5. Varianzanalysen

1. Varianzanalysen werden eingesetzt, um die Unterschiede der Mittelwerte von mehr als zwei Stichproben auf Signifikanz zu prüfen. Aufgrund der Tatsache, dass die meisten heutigen psychologischen Experimente mehr als einen Faktoren berücksichtigen, gehören die Varianzanalysen zu den bedeutsamsten Verfahren in der Experimentalpsychologie überhaupt.
2. Bei einer einfaktoriellen Varianzanalyse wird der Effekt eines Faktors, bei einer mehrfaktoriellen die Effekte von mehr als einem Faktor untersucht.
3. $H_0: \mu_1 = \mu_2 = \dots = \mu_J$ oder $H_0: \sum(\mu_j - \mu)^2 = 0$
4. Die Fehlervarianz (Fehlerstreuung, Zufallsvarianz, Zufallsstreuung, Standardfehler) ist die nicht auf den Effekt des Faktors zurückführbare Streuung der Messwer-

te und errechnet sich aus der Summe der quadrierten Abweichungen aller Messwerte von ihrem jeweiligen Gruppenmittel (SAQ_i).

5. $SAQ_G = SAQ_Z + SAQ_i$
6. Sie sind synonyme Begriffe und bezeichnen den relativen Anteil der Streuung, die auf den unterschiedlichen Stufen des Faktors beruhen, an der Gesamtstreuung. Der Rest der Gesamtstreuung ist auf zufällige Einflüsse zurückzuführen und nicht auf den Effekt des untersuchten Faktors. Je größer ein Varianzanteil auf den Einfluss des Faktors zurückgeführt werden kann, desto stärker ist die Wirkung des Faktors.
7. Durch das Verfahren des multiplen Mittelwertvergleichs von SCHEFFÉ.
8. Die kritischen Grenzen der einfaktoriellen Varianzanalyse müssen durch einen Korrekturfaktor ε verkleinert werden. Um die recht komplizierte Berechnung zu umgehen, bestimmen wir in Haus- und Klausuraufgaben einen die kritische Grenze enthaltenden kritischen Bereich, der zur Entscheidung für oder gegen die H_0 herangezogen wird.
9. Führt eine einfaktorielle Varianzanalyse zu keinem Ergebnis, kann die möglicherweise zu hohe Zufallsstreuung verkleinert werden, indem eine (vermutlich) maßgebliche Störvariable in die Varianzanalyse mit einbezogen wird. Dadurch wird nicht nur die Zufallsstreuung verkleinert, sondern der Effekt eines vermuteten Störfaktors kann zusätzlich auf seine statistische Bedeutsamkeit überprüft werden.
10. Der Haupteffekt eines Faktors wirkt sich auf allen Stufen des anderen Faktors aus, der einfache Haupteffekt nur auf einer einzelnen Stufe des anderen Faktors. Oder: Signifikante Mittelwertunterschiede, die sich über alle Stufen eines Faktors ergeben, werden auf den Haupteffekt des anderen Faktors, signifikante Mittelwertunterschiede in einer einzelnen Stufe eines Faktors werden auf den einfachen Haupteffekt des anderen Faktors zurückgeführt.
11. Eine Wechselwirkung zwischen zwei Faktoren liegt vor, wenn sich die einfachen Haupteffekte eines der beiden Faktoren auf mindestens zwei Stufen des anderen Faktors signifikant unterscheiden, d.h. die Wirkung eines Faktors ist auf mindestens zwei Stufen des anderen Faktors unterschiedlich. Es gibt folgende Typen: Eine ordinale Interaktion liegt vor, wenn sich beide Faktoren auf den Stufen des anderen Faktors zwar unterschiedlich, aber in die gleiche Richtung auswirken. Eine disordinale Interaktion liegt vor, wenn sich beide Faktoren auf den Stufen des anderen Faktors in die jeweilig entgegengesetzte Richtung auswirken. Eine semiordinale Interaktion liegt vor, wenn sich der eine Faktor in die gleiche Richtung und der andere in die entgegengesetzte Richtung auswirkt.
12. Die Voraussetzung zu allen Varianzanalysen ist, dass die den Stichproben zugehörigen Populationen sowohl normalverteilt als auch varianzhomogen sind. Die inferenzstatistische Prüfung dieser Voraussetzungen kann man umgehen, indem man dafür sorgt, dass die zu erhebenden Stichproben hinreichend (mehr als zehn V_{pn}) und gleich groß sind.
13. a) Die Mittelwerte der 4 Stichproben müssen identisch sein.
b) Die Werte der V_{pn} innerhalb einer jeden Stichprobe müssen identisch sein.

7.3.6. Nichtparametrische, verteilungsfreie Verfahren

1. auf nichtmetrischen Skalenniveaus
2. Wir haben die Binomial- und die χ^2 -Verteilung kennengelernt.

3. Eine Binomialverteilung ist eine diskrete Wahrscheinlichkeitsverteilung, die die Wahrscheinlichkeit für die Auftretenshäufigkeit für eine von zwei möglichen Alternativen bei einer bestimmten Anzahl von Versuchen wiedergibt. Das Bildungsgesetz für die Wahrscheinlichkeit $W(f)$ einer bestimmten Auftretenshäufigkeit f einer Alternativen, die mit der Wahrscheinlichkeit π bei einem einmaligen Versuch auftritt, findet sich auf S. 75. Erwartungswert einer Binomialverteilung: $\mu = n \cdot \pi$; Varianz: $\sigma^2 = n \cdot \pi \cdot (1 - \pi)$.
4. Ein Binomialtest wird durchgeführt, wenn eine Hypothese über den Anteil an der Population oder die Auftretenshäufigkeit eines Merkmals geprüft werden soll. Hierzu wird der exakte oder der approximative Binomialtest gerechnet. Während der exakte Binomialtest immer gerechnet werden kann, muss für den approximativen die Voraussetzung $n > 9/(\pi \cdot (1 - \pi))$ erfüllt sein.
5. Eine χ^2 -Verteilung ist eine stetige Wahrscheinlichkeitsverteilung, die die Verteilung der möglichen Abweichungen von beobachteten zu erwarteten Häufigkeiten wiedergibt. Zu der χ^2 -Verteilung gibt es den χ^2 -Unabhängigkeitstest und den χ^2 -Anpassungstest. Der χ^2 -Unabhängigkeitstest dient der Prüfung, ob der in einer Stichprobe durch Cramer's V ermittelte nominale Zusammenhang signifikant ist. Mit dem χ^2 -Anpassungstest werden Hypothesen über die Auftretenswahrscheinlichkeit von verschiedenen Alternativen geprüft. Für alle χ^2 -Tests müssen die Voraussetzungen erfüllt sein, dass alle erwarteten Häufigkeiten größer als Eins und höchstens ein Fünftel aller erwarteten Häufigkeiten größer als Fünf sind.
6. 1. $H_0: \pi_k = 0,2 \quad H_1: \pi_k \neq 0,2 \quad \text{für alle } k = 1 \dots 5$
 2. $\alpha = 0,01/0,05$
 3. χ^2 -Anpassungstest: Voraussetzung $E(f_k) > 1$, max. $1/5$ der $E(f_k) < 5$ gegeben
 4. $n = 450$
 5. $\chi^2 = 508,56$
 6. $\chi^2_{0,99;4} = 13,3$ bzw. $\chi^2_{0,95;4} = 9,5$
 7. 1%: $508,56 > 13,3 \Rightarrow H_1$ annehmen
 5%: $508,56 > 9,5 \Rightarrow H_1$ annehmen
 Auf beiden Signifikanzniveaus wurde die Hypothese, dass sich die Klienten auf alle Therapieformen gleichmäßig verteilen, widerlegt.

7.4. Lösungen zu den Klausuraufgaben

7.4.1. Klausur 1989

Aufgabe 1:

1. H_0 : A; B unabhängig H_1 : A, B abhängig
2. $\alpha = 0,05$
3. χ^2 -Unabhängigkeitstest, Vor.: alle $E(f_{jk}) > 1$, höchstens $1/5$ der $E(f_{jk}) < 5$
4. $J=2, K=3, n=60; 5[10], 10[5], 26[20], 4[10], 9[10], 6[5]$
5. $\chi^2 = 13,2$
6. $\chi^2_{0,95;2} = 6,0$
7. $13,2 > 6,0 \Rightarrow H_0$ verwerfen, H_1 annehmen

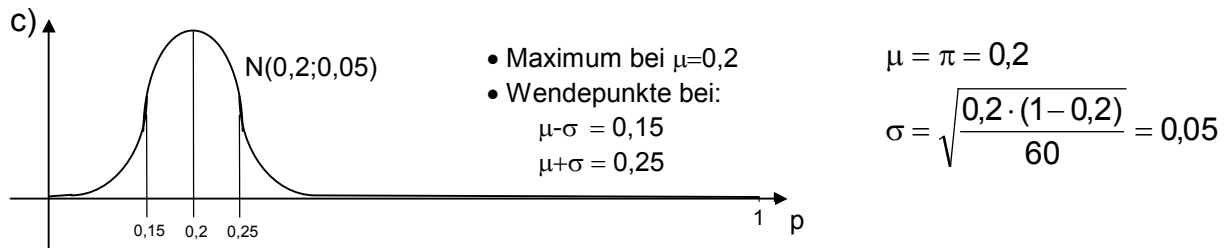
Die Behauptung, dass zwischen Bildungsniveau und politischer Einstellung ein signifikanter Zusammenhang besteht, konnte auf dem 5%-Signifikanzniveau bestätigt werden.

Aufgabe 2:

Es muss ein eigenständiges Beispiel in der Art von Hausaufgabe 64 oder 65 gebracht werden.

Aufgabe 3:

- a) Die relative Häufigkeit p kann allgemein Werte zwischen Null und Eins annehmen: $0 \leq p \leq 1$. Bei dieser Stichprobe kann zudem p nur 60stel-Werte annehmen, insgesamt sind das 61 Werte: $p = \{0/60, 1/60, 2/60, \dots, 60/60\}$
 b) Die Wahrscheinlichkeitsverteilung von p ist eine $N(0,2;0,05)$ -Verteilung.



- d) Die Voraussetzung $n > \frac{9}{\pi \cdot (1-\pi)} = \frac{9}{0,2 \cdot (1-0,2)} = 56,25$ ist mit $n=60$ erfüllt.

Aufgabe 4:

Berechnung eines Konfidenzintervalls für μ mit bekanntem σ : $\mu = \bar{x} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

Der wahre IQ dieser Person ist mit einer Sicherheitswahrscheinlichkeit von 95% in einem Bereich zwischen 105,2 und 124,8 zu vermuten: $105,2 \leq \mu \leq 124,8$.

Aufgabe 5:

Der Fehler 1. Art (α -Fehler) ist der Fehler, dass eine Nullhypothese abgelehnt wird, obwohl sie richtig ist. Der Fehler entsteht durch das Setzen des Signifikanzniveaus α und entspricht ihm in seiner Fehlerwahrscheinlichkeit α . Je größer das Signifikanzniveau, desto größer der Fehler 1. Art.

Der Fehler 2. Art (β -Fehler) ist der Fehler, dass eine Nullhypothese beibehalten wird, obwohl sie falsch ist. Die zugehörige Fehlerwahrscheinlichkeit β lässt sich nicht rechnerisch ermitteln und hängt ab von:

- der Fehlerwahrscheinlichkeit α : je größer α , desto kleiner β
- dem Stichprobenumfang n : je größer n , desto kleiner β
- der Streuung σ der Messwerte: je kleiner σ , desto kleiner β
- dem Unterschied zwischen hypothetischen und tatsächlichem Mittelwert: je größer der U., desto kleiner β
- vom verwendeten Testverfahren

Aufgabe 6:

- a) Mit zunehmenden Freiheitsgrad wird die anfangs breite und flache t-Verteilung immer schmaler und höher, da die t-Werte von der Standardabweichung s der Stichprobe abhängen und s mit zunehmenden Stichprobenumfang immer mehr dem σ der entsprechenden Normalverteilung entspricht. Wächst der Stichprobenumfang auf Unendlich an und entspricht damit der hypothetischen Grundgesamtheit ($n=\infty$), entspricht die t-Verteilung der Standardnormalverteilung. Dies hat folgende Konsequenzen: Die kritischen t-Grenzen zu einem bestimmten Signifikanzniveau rücken mit zunehmenden Freiheitsgrad df zusammen, da die t-Verteilung

immer schmaler wird, bis sie mit $df=\infty$ den kritischen z-Grenzen der Standardnormalverteilung entsprechen. Die positiven kritischen t-Grenzen werden dabei immer kleiner, bis sie mit $df=\infty$ den positiven kritischen z-Grenzen entsprechen.

- b) χ^2 -Werte werden mit zunehmenden Freiheitsgrad immer größer. Durch das Quadrat in der Berechnungsformel und dem Umstand, dass Häufigkeiten nicht negativ sein können, liegen χ^2 -Werte immer im positiven Bereich.

7.4.2. Klausur 1990

Aufgabe 1:

Berechnung des Konfidenzintervalls für μ mit σ unbekannt: $\mu = \bar{x} \pm t_{1-\frac{\alpha}{2}; n-1} \cdot \frac{s}{\sqrt{n}}$

Das Durchschnittsalter der gesamten Bevölkerungsgruppe liegt mit einer Sicherheitswahrscheinlichkeit von 99% zwischen 34,5 und 39,5 Jahren: $34,5 \leq \mu \leq 39,5$

Aufgabe 2:

Es muss ein eigenständiges Beispiel in der Art der Hausaufgabe 61 oder 62 gebracht werden.

Aufgabe 3:

Die Wahrscheinlichkeit, dass *eine* Person über 60 Punkte erreicht, ist:

$$W(A) = W\{x > 60\} = 1 - \Phi\left(\frac{60 - 50}{10}\right) = 0,159$$

- a) Die Wahrscheinlichkeit, dass *zwei* Personen *gleichzeitig* einen Testwert von über 60 Punkten erreichen, beträgt 2,5%: $W(A \cap A) = W(A) \cdot W(A) = 0,025$.
- b) Die Wahrscheinlichkeit, dass *mindestens eine* von beiden Personen einen Testwert von über 60 Punkten erreicht, lautet 29,3%:
 $W(A \cup A) = 1 - W(\bar{A}) \cdot W(\bar{A}) = 1 - 0,841 \cdot 0,841 = 0,293$
- c) Es liegt eine Stichprobenverteilung von Stichproben mit dem Umfang zwei Personen vor: $N(50; 7,1)$

$$\mu_{\bar{x}} = \mu_x = 50 \quad \sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} = 7,1$$

Die Wahrscheinlichkeit, dass eine Stichprobe mit zwei Personen einen Durchschnittswert von über 60 Punkten erreicht, beträgt 8,1%:

$$W\{\bar{x} > 60\} = 1 - \Phi\left(\frac{60 - 50}{7,1}\right) = 0,081$$

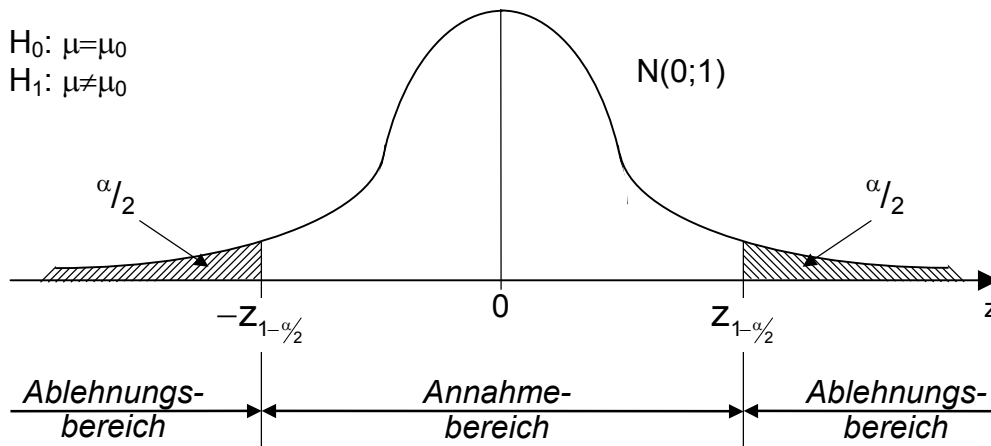
Aufgabe 4:

1. $H_0: \rho=0 \quad H_1: \rho \neq 0$
2. $\alpha = 0,05$
3. Formel (3), STAT -9-
4. $r=0,2; n=70$
5. $t=1,68$
6. $t_{0,975;68} = 1,99$
7. $1,68 < 1,99 \Rightarrow H_0$ annehmen, H_1 ablehnen

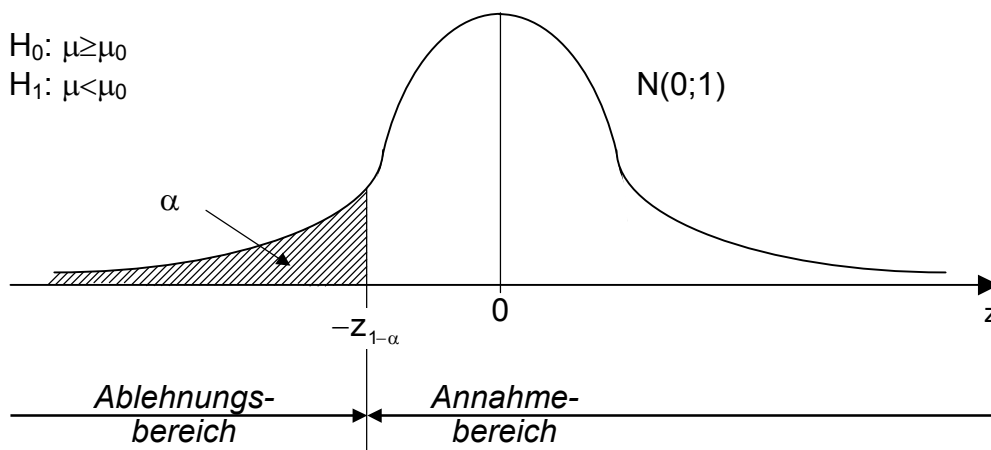
Die Behauptung, dass zwischen den zwei Merkmalen ein signifikanter Zusammenhang besteht, muss auf dem 5%-Signifikanzniveau abgelehnt werden.

Aufgabe 5:

Zweiseitiger Test:



Einseitiger Test:



Aufgabe 6:

Die Präzision eines Konfidenzintervalls hängt ab von:

- der Sicherheitswahrscheinlichkeit $1-\alpha$ (dem Signifikanzniveau α)
- dem Stichprobenumfang n
- der Streuung σ der normalverteilten Grundgesamtheit der Messwerte

Die Präzision eines Konfidenzintervalls kann erhöht werden durch:

- Verkleinerung der Sicherheitswahrscheinlichkeit $1-\alpha$ (Vergrößerung des Signifikanzniveaus α)
- Vergrößerung des Stichprobenumfangs n
- durch Auswahl eines geeigneten Verfahrens, da σ verfahrensabhängig einen Einfluss hat

7.4.3. Klausur 1991

Aufgabe 1:

Konfidenzintervalls eines Anteilwerts π : $\pi = p \pm z_{1-\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$

Die Voraussetzung $n > \frac{9}{p \cdot (1-p)} = \frac{9}{0,75 \cdot (1-0,75)} = 48$ ist mit $n=80$ erfüllt.

Der Anteil der in der Grundgesamtheit mit ihrer Studiensituation zufriedenen Studenten kann in einem Bereich von 65,5% und 84,5% vermutet werden: $0,655 \leq \pi \leq 0,845$.

Aufgabe 2:

1. $H_0: \mu_1 - \mu_2 = 0$ $H_1: \mu_1 - \mu_2 \neq 0$
2. $\alpha = 0,01$
3. Klassischer t-Test

Prüfung auf Gleichheit der Varianzen (Vor. zum klass. t-Test):

1. $H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$
2. $\alpha = 0,10$
5. $F = 1,36$
6. $F_{0,95;59;59} = 1,58$
7. $1,36 < 1,58 \Rightarrow H_0$ beibehalten \Rightarrow Varianzen sind homogen
 \Rightarrow Voraussetzung für den klassischen t-Test ist erfüllt

4. $n_1 = 60, \bar{x}_1 = 38, s_1 = 6; n_2 = 60, \bar{x}_2 = 43, s_2 = 7$
5. $t = -4,20$
6. $t_{0,995;118} = 2,58$
7. $|-4,20| > 2,58 \Rightarrow H_0$ verwerfen, H_1 annehmen

Es konnte auf einem Signifikanzniveau von 1% bestätigt werden, dass zwischen beiden Gruppen ein signifikanter Unterschied besteht.

Aufgabe 3:

x_k	π_k	$x_k \cdot \pi_k$	$x_k - \mu$	$(x_k - \mu)^2$	$(x_k - \mu)^2 \cdot \pi_k$
0	0,3	0	-1,1	1,21	0,363
1	0,4	0,4	-0,1	0,01	0,004
2	0,2	0,4	0,9	0,81	0,162
3	0,1	0,3	1,9	3,61	0,361
Σ		1,1			0,89

a) Der Erwartungswert liegt bei 1,1 Unfällen: $\mu = \Sigma x_k \cdot \pi_k = 1,1$

b) Standardabweichung 0,94 Unfälle: $\sigma^2 = \Sigma (x_k - \mu)^2 \cdot \pi_k = 0,89$ $\sigma = \sqrt{\sigma^2} = 0,94$

Aufgabe 4:

Es muss ein eigenständiges Beispiel in der Art von Hausaufgabe 60 (mit $r=0,4$) gebracht werden.

Aufgabe 5: ???

Aufgabe 6:

- a) siehe Anhang von Herrn HUBER, jeweils oberhalb der z- und t-Wert-Tabellen
- b) t-Verteilungen sind breiter und flacher, da die Streuung der t-Verteilung gegenüber der Standardnormalverteilung - durch die Verwendung der Standardabweichung der Stichprobe zur Schätzung der Streuung der Standardnormalverteilung - vergrößert ist, die Gesamtfläche aber 1 bleiben muss.

- c) Mit zunehmenden Freiheitsgrad wird die Streuung kleiner, womit die Dichtekurve höher und schmaler wird, bis sie mit $df = \infty$ der Standardnormalverteilung entspricht.
- d) Eine t-verteilte Zufallsgröße ist ein stetiges Merkmal, dessen Werte einer Grundgesamtheit, deren σ unbekannt ist, normalverteilt sind: z.B. die Körpergröße der Pygmäen (biologische Merkmale sind zumeist normalverteilt, die Streuung der Körpergröße der Pygmäen dürfte unbekannt sein).

7.4.4. Klausur 1992

Aufgabe 1:

Konfidenzintervalls eines Anteilwerts π : $\pi = p \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}$

Die Voraussetzung $n > 70,6$ ist mit $n = 80$ erfüllt. In der zugehörigen Grundgesamtheit ist der Anteil der richtigen Diagnosen zwischen 77% und 93% zu vermuten.

Aufgabe 2:

1. $H_0: \mu_1 - \mu_2 \geq 0$ $H_1: \mu_1 - \mu_2 < 0$
2. $\alpha = 0,05$
3. z-Test für abhängige Stichproben (Formel (4)-(I), STAT -9-)
4. $\bar{d} = \bar{x}_1 - \bar{x}_2 = 6 - 5,5 = 0,5$, $s_d = 1,5$, $n = 40$
5. $t = 2,11$
6. $t_{0,95;39} = -1,68$
7. $2,11 > -1,68 \Rightarrow H_0$ beibehalten, H_1 ablehnen

Es konnte auf dem 5%-Signifikanzniveau die Hypothese nicht bestätigt werden, dass durch die Erweiterung des Fernsehprogramms die täglichen Einschaltzeiten signifikant verlängert wurden.

Anmerkung: Aufgrund von $\bar{X}_1 > \bar{X}_2$ hätte man sich die Rechnerei auch sparen können: Die positive Differenz $\bar{X}_1 - \bar{X}_2$ hätte nie und nimmer die negative kritische Grenze unterschritten!

Aufgabe 3:

- | | |
|-----------------------------------------------------------------|---------------------|
| $A = \{\mu \text{ liegt im Konfidenzintervall A}\}$ | $W(A) = 0,95$ |
| $\bar{A} = \{\mu \text{ liegt nicht im Konfidenzintervall A}\}$ | $W(\bar{A}) = 0,05$ |
| $B = \{\mu \text{ liegt im Konfidenzintervall B}\}$ | $W(B) = 0,95$ |
| $\bar{B} = \{\mu \text{ liegt nicht im Konfidenzintervall B}\}$ | $W(\bar{B}) = 0,05$ |

- a) Alle beide Intervalle enthalten μ , d.h. Konfidenzintervall A *und* Konfidenzintervall B enthält μ : $A \cap B$ (A, B unabhängig)
 $W(A \cap B) = W(A) \cdot W(B) = 0,95 \cdot 0,95 = 0,9025$
 Die Wahrscheinlichkeit, dass beide Intervalle μ enthalten, beträgt 90,25%
- b) Keines der Intervalle enthält μ , d.h. Konfidenzintervall A *und* Konfidenzintervall B enthält μ nicht: $\bar{A} \cap \bar{B}$ (\bar{A} , \bar{B} unabhängig)
 $W(\bar{A} \cap \bar{B}) = W(\bar{A}) \cdot W(\bar{B}) = 0,05 \cdot 0,05 = 0,0025$
 Die Wahrscheinlichkeit, dass keines der Intervalle μ enthält, lautet 0,25%
- c) Mindestens eins der Intervalle enthält μ nicht, d.h. Konfidenzintervall A *oder* Konfidenzintervall B enthält μ nicht: $\bar{A} \cup \bar{B}$ (\bar{A} , \bar{B} unabhängig)
 $W(\bar{A} \cup \bar{B}) = 1 - W(A) \cdot W(B) = 1 - 0,95 \cdot 0,95 = 0,0975$
 Die Wahrscheinlichkeit, dass mindestens eins von zwei Intervallen μ enthält, ist 9,75%.

Aufgabe 4:

Es muss ein eigenständiges Beispiel in der Art von Hausaufgabe 72 oder 76 gebracht werden.

Aufgabe 5: s. Klausur 1989, Aufgabe 5

Aufgabe 6:

Die Grundgesamtheit ist die Gesamtheit, für die wir eine Aussage erstellen wollen. Die Stichprobe ist die Gesamtheit, von der Daten erhoben werden, und ist ein Teil der Grundgesamtheit. In der Hoffnung, dass die Stichprobe ein Abbild der Grundgesamtheit darstellt, wird von der Stichprobe durch Induktionsschluss auf die Grundgesamtheit geschlossen. Eine Stichprobenverteilung ist eine Wahrscheinlichkeitsverteilung von Stichprobenkennwerten, z.B. des arithmetischen Mittels. Die Stichprobenverteilung des arithmetischen Mittels ist eine Normalverteilung, dessen Erwartungswert $\mu_{\bar{x}}$ dem Erwartungswert μ_x der Normalverteilung des zugrundeliegenden Merkmals entspricht und dessen Standardabweichung $\sigma_{\bar{x}}$ durch folgende Gleichung aus der Standardabweichung σ_x der Normalverteilung des zugrundeliegenden Merkmals berechnet wird: $\sigma_{\bar{x}} = \sigma_x/\sqrt{n}$.

7.4.5. Klausur 1993

Aufgabe 1:

1. $H_0: \pi \geq 0,5$ $H_1: \pi < 0,5$

2. $\alpha = 0,05$

3. *Auswahl des Verfahrens:*

- Exakter Binomialtest
Die Voraussetzung „dichotomes Merkmal“ ist gegeben.
- Approximativer Binomialtest
Die Voraussetzung $n > 36$ ist nicht gegeben.
- χ^2 -Anpassungstest
Dieser kann nicht gerechnet werden, da gerichtete Hypothesen vorliegen.
⇒ Es kann nur der exakte Binomialtest gerechnet werden.

4. $n = 18$

5. $f = 6$

6. $P = W\{f \leq 6\} = 0,119$

7. $0,119 > 0,05 \Rightarrow H_0$ annehmen

Es wurde auf dem 5%-Signifikanzniveau nicht bestätigt, dass mit einer bestimmten Heilmethode signifikant weniger als die Hälfte geheilt werden.

Aufgabe 2:

Eigenständiges Beispiel in der Art von Hausaufgabe 61 oder 62.

Aufgabe 3:

1. $H_0: \sum(\mu_j - \mu)^2 = 0$ $H_1: \sum(\mu_j - \mu)^2 > 0$

2. $\alpha = 0,01$

3. Einfaktorielle Varianzanalyse für unabhängige Stichproben

4. $n_1=5, \bar{x}_1=6, SAQ_1=28; n_2=5, \bar{x}_2=14, SAQ_2=58; n_3=5, \bar{x}_3=10, SAQ_3=40$

$N=15, J=3; \bar{x}_G=10; SAQ_Z=160; SAQ_I=126; SAQ_G=286$

5. Tafel der Varianzanalyse:

	SAQ	df	MQ	F	$F_{0,99;2;12}$
Zwischen	160	2	80	7,62	6,93
Innerhalb	126	12	10,5		
Gesamt	286				

7. $7,62 > 6,93 \Rightarrow H_0$ wird abgelehnt, H_1 wird angenommen

Es konnte auf einem Signifikanzniveau von 1% bestätigt werden, dass sich die drei Diäten signifikant voneinander unterscheiden.

Aufgabe 4:

1. $H_0: \rho \leq 0$ $H_1: \rho > 0$

2. $\alpha = 0,05$

3. Formel (3), STAT -9-

4. $r=0,2$; $n=80$

5. $t=1,80$

6. $t_{0,95;78} = 1,67$

7. $1,80 > 1,67 \Rightarrow H_0$ verwerfen, H_1 annehmen

Die Behauptung, der Zusammenhang zwischen den zwei Merkmalen sei positiv, konnte auf dem 5%-Signifikanzniveau bestätigt werden.

Aufgabe 5: s. Klausur 1990, Aufgabe 6

Aufgabe 6:

a) – Die den Stichproben zugehörigen Grundgesamtheiten sind normalverteilt.

– Diese Grundgesamtheiten sind auch varianzhomogen.

– (Die abhängige Variable ist metrisch, die unabhängige Variable nominal.)

b) – Die Umfänge der Stichproben sind hinreichend groß.

– Die Umfänge der Stichproben sind alle gleich groß.

c) Keine Wechselwirkung liegt vor, wenn der Effekt eines Faktors auf allen Stufen des anderen signifikant gleich groß ist. Eine Wechselwirkung liegt vor, wenn der Effekt eines der beiden Faktoren auf mindestens zwei Stufen des anderen Faktors signifikant unterschiedlich ist.

7.4.6. Klausur 1994

Aufgabe 1:

1. $H_0: \mu_1 - \mu_2 \geq 0$ $H_1: \mu_1 - \mu_2 < 0$

2. $\alpha = 0,01$

3. WELCH-Test

4. $n_1 = 50, \bar{x}_1 = 38, s_1 = 6; n_2 = 50, \bar{x}_2 = 41, s_2 = 9$

5. $t = -1,96$

6. $t_{0,99;\infty} = -2,33$ ($n_1, n_2 \geq 50$ und $H_1: <0!$)

7. $-1,96 > -2,33 \Rightarrow H_0$ annehmen, H_1 ablehnen

Auf einem Signifikanzniveau von 1% wird die Hypothese verworfen, dass die Lehrmethode 2 bei Kindern zu höheren Leistungen führt.

Aufgabe 2:

1. $H_0: A, B$ unabhängig $H_1: A, B$ abhängig

2. $\alpha = 0,01$

3. χ^2 -Unabhängigkeitstest
4. $J=2, K=2; 102 [80], 98 [120], 58 [80], 142 [120]$
5. $\chi^2 = 20,17$
6. $\chi_{0,95;1}^2 = 6,6$
7. $20,17 > 6,6 \Rightarrow H_1$ annehmen

Es konnte auf einem Signifikanzniveau von 1% festgestellt werden, dass es einen signifikanten Zusammenhang zwischen den Merkmalen Bildungsniveau und Urlaubsziel gibt: Urlauber mit niedrigem Bildungsniveau bevorzugen das Ausland, mit hohem das Inland.

Aufgabe 3:

1. $H_0: \sum(\mu_j - \mu)^2 = 0 \quad H_1: \sum(\mu_j - \mu)^2 > 0$
2. $\alpha = 0,05$
3. Einfaktorielle Varianzanalyse für abhängige Stichproben
4. $n=4; \bar{X}_{\bullet 1}=8,5; \bar{X}_{\bullet 2}=9,5; \bar{X}_{\bullet 3}=15; J=3; \bar{X}_{1\bullet}=10; \bar{X}_{2\bullet}=13; \bar{X}_{3\bullet}=12; \bar{X}_{4\bullet}=9; \bar{X}=11$
5. Tafel der Varianzanalyse:

	SAQ	df	MQ	F
Treatment	98	2	49	13,36
Blöcke	30	3		
Rest	22	6	3,667	
Gesamt	150			

6. $F_{0,95;2;6} = 5,14 \quad (\epsilon!)$
 $F_{0,95;1;3} = 10,1$
7. $13,36 > (5,14 \text{ bis } 10,1) \Rightarrow H_0$ wird abgelehnt, H_1 wird beibehalten
 Zwischen den drei Testzeitpunkten konnte mit einer statistischen Sicherheit von 95% ein signifikanter Leistungsunterschied festgestellt werden.

Aufgabe 4:

Berechnung eines Konfidenzintervalls für μ mit bekanntem σ : $\mu = \bar{x} \pm z_{1-\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

Der wahre IQ dieser Person liegt bei einer Sicherheitswahrscheinlichkeit von 95% in einem Bereich zwischen 95,2 und 114,8: $95,2 \leq \mu \leq 114,8$

Aufgabe 5: ???

Aufgabe 6:

- a) siehe Anhang von Herrn HUBER, jeweils oberhalb der z- und t-Wert-Tabellen
- b) t-Verteilungen sind breiter und flacher, da die Streuung der t-Verteilung gegenüber der Standardnormalverteilung - durch die Verwendung der Standardabweichung der Stichprobe zur Schätzung der Streuung der Standardnormalverteilung - vergrößert ist, die Gesamtfläche aber 1 bleiben muss.
- c) Das Quadrat einer normalverteilten z-Größe für eine zweiseitige Fragestellung ist χ^2 -verteilt mit dem Freiheitsgrad 1 für eine einseitige Fragestellung:

$$\left(z_{1-\alpha/2}\right)^2 = \chi_{1-\alpha;1}^2$$

- d) Das Quadrat einer t-Größe für eine zweiseitige Fragestellung ist eine F-Größe für eine einseitige Fragestellung mit dem Zählerfreiheitsgrad 1. Der Nennerfreiheitsgrad ist dabei der Freiheitsgrad der t-Größe. Für das Quadrat einer t-Größe mit dem Freiheitsgrad $df=10$ gilt:

$$\left(t_{1-\alpha/2;10}\right)^2 = F_{1-\alpha;1;10}$$

7.4.7. Klausur 1996

Aufgabe 1:

1. $H_0: \mu_1 - \mu_2 = 0$ $H_1: \mu_1 - \mu_2 \neq 0$
 2. $\alpha = 0,01$
 3. Vergleich der Mittelwerte zweier abhängiger Stichproben
 4. $\bar{d} = -2$; $s_d = 4$; $n = 25$
 5. $t = -2,5$
 6. $t_{0,995;24} = 2,80$
 7. $|-2,5| < 2,80 \Rightarrow H_0$ annehmen, H_1 ablehnen
- Die durchschnittliche Gewichtsabnahme von 2 kg muss auf dem 1%-Niveau als nicht signifikant beurteilt werden.

Aufgabe 2:

1. $H_0: \pi = 0,5$ $H_1: \pi \neq 0,5$ ($\pi =$ Anteil In-der-Stadt-leben-Wollende in Bev.)
 2. $\alpha = 0,05$
 3. Approximativer Binomialtest; Vor.: $n > 36$ ist erfüllt!
 4. $f = 52$; $n = 80$; $p = 0,65$; $\pi_0 = 0,5$ ($p =$ Anteil In-der-Stadt-leben-Wollende in St.)
 5. $z = 2,68$
 6. $z_{0,95} = 1,64$
 7. $2,68 > 1,64 \Rightarrow H_1$ beibehalten, H_0 zurückweisen
- Auf dem 5%-Niveau konnte die Behauptung angenommen werden, dass auch in der Grundgesamtheit mehr Leute in der Stadt als auf dem Land leben wollen.

Aufgabe 3:

	SAQ	df	MQ	F	$F_{0,95}$	Ergebnis
Faktor A	20	1	20	8	4,49	s.
Faktor B	45	1	45	18	4,49	s.
Wechselwirkung AxB	5	1	5	2	4,49	n.s.
Innerhalb	110	16	2,5			

Auf dem Signifikanzniveau von 5% konnten signifikante Haupteffekte für beide Faktoren, aber keine signifikante Wechselwirkung festgestellt werden.

Aufgabe 4:

$$SAQ_Z = SAQ_A + SAQ_B + SAQ_{AxB} = 70; SAQ_G = SAQ_I + SAQ_Z = 110$$

$$\eta_Z^2 = 0,636; \eta_A^2 = 0,0182; \eta_B^2 = 0,409; \eta_{AxB}^2 = 0,045$$

Die Gesamtvarianz kann zu 63,6% von den Faktoren – genauer: zu 18,2% von Faktor A, zu 40,9% von Faktor B und zu 4,5% von der nicht signifikanten Wechselwirkung zwischen beiden Faktoren – erklärt werden.

Aufgabe 5:

a) s. Klausur 91, Aufg. 6a); b) s. Klausur 89, Aufg. 6a); s. Klausur 89, Aufg. 6b)

Aufgabe 6: s. Klausur 90, Aufgabe 6

7.4.8. Klausur 1997

Aufgabe 1:

1. $H_0: \sigma_1^2 = \sigma_2^2$ $H_1: \sigma_1^2 \neq \sigma_2^2$
 2. $\alpha = 0,10$
 3. F-Test zur Prüfung der Varianzhomogenität zweier Stichproben
 4. $S_1 = 8, n_1 = 25, s_2 = 6, n_2 = 30$
 5. $F = 2$
 6. $F_{0,90;24;29} = 1,90$
 7. $2 > 1,90 \Rightarrow H_1$ beibehalten, H_0 zurückweisen
- Auf dem 10%-Niveau der Verlässlichkeit unterscheiden sich die Varianzen beider Gruppen signifikant. Voraussetzung: Normalverteilung beider Grundgesamtheiten.

Aufgabe 2:

1. $H_0: \pi \geq 0,5$ $H_1: \pi < 0,5$ ($\pi =$ Anteil Raucher in der Bevölkerung)
 2. $\alpha = 0,05$
 3. Vor. für approximativen Binomialtest mit $n = 24 > 36$ nicht erfüllt, daher:
→ Exakter Binomialtest
 4. $f = 9; n = 24$
 5. $f' = 9$, da $f < n/2$
 6. $P = 2 \cdot W(f \leq 9) = 2 \cdot 0,154 = 0,308$
 7. $0,308 > 0,05 \Rightarrow H_0$ beibehalten, H_1 zurückweisen
- Mit einer Irrtumswahrscheinlichkeit von 5% konnte nicht bestätigt werden, dass es in einer bestimmten Bevölkerungsgruppe weniger Raucher als Nichtraucher gibt.

Aufgabe 3:

1. $H_0: \sum(\mu_j - \mu)^2 = 0$ $H_1: \sum(\mu_j - \mu)^2 > 0$
2. $\alpha = 0,05$
3. Einfaktorielle Varianzanalyse für abhängige Stichproben
4. $n=3; \bar{X}_{\cdot 1}=10; \bar{X}_{\cdot 2}=12; \bar{X}_{\cdot 3}=17; J=3; \bar{X}_{1\cdot}=12; \bar{X}_{2\cdot}=14; \bar{X}_{3\cdot}=13; \bar{X}=13$
5. Tafel der Varianzanalyse:

	SAQ	df	MQ	F
Treatment	78	2	39	26
Blöcke	6	2		
Rest	6	4	1,5	
Gesamt	90			

6. $F_{0,95;2;4} = 6,94$ (ε!)
 $F_{0,95;1;2} = 18,5$
 7. $26 > (6,94 \text{ bis } 18,5) \Rightarrow H_0$ wird abgelehnt, H_1 wird beibehalten
- Zwischen den 3 Testzeitpunkten konnte auf einem Signifikanzniveau von 5% ein signifikanter Leistungsunterschied zwischen den drei Zeitpunkten festgestellt werden.

Aufgabe 4:

Berechnung des Konfidenzintervalls für μ mit σ unbekannt: $\mu = \bar{x} \pm t_{1-\frac{\alpha}{2};n-1} \cdot \frac{s}{\sqrt{n}}$

Die durchschnittliche Wohngröße der Haushalte dieses Landes liegt mit einer Sicherheitswahrscheinlichkeit von 95% zw. 77,42qm und 82,58qm: $77,42 \leq \mu \leq 82,58$

Aufgabe 5: ???

Aufgabe 6:

$$\begin{aligned} \text{a) } W\{85 \leq x \leq 115\} &= \Phi\left(\frac{115 - 100}{15}\right) - \Phi\left(\frac{85 - 100}{15}\right) \\ &= 0,841 - (1 - 0,841) = 0,682 \end{aligned}$$

Die Wahrscheinlichkeit, dass eine zufällig ausgewählte Person einen IQ zwischen 85 und 115 hat, ist 68,2%.

$$\begin{aligned} \text{b) } W\{85 \leq \bar{x} \leq 115\} &= \Phi\left(\frac{115 - 100}{15/\sqrt{9}}\right) - \Phi\left(\frac{85 - 100}{15/\sqrt{9}}\right) \\ &= 0,999 - (1 - 0,999) = 0,998 \end{aligned}$$

Die Wahrscheinlichkeit, dass der Durchschnitts-IQ von 9 zufällig ausgewählten Personen zwischen 85 und 115 liegt, beträgt 99,8%.

Schritte	Zweiseitige Fragestellung		Einseitige Fragestellung	
1. Hypothesen	<u>Nullhypothese:</u> (1) $H_0 : \mu = \mu_0$ (2) $H_0 : \pi = \pi_0$ (3) $H_0 : \rho = 0$ (4) $H_0 : \mu_1 - \mu_2 = 0$	<u>Alternativhypothese:</u> (1) $H_1 : \mu \neq \mu_0$ (2) $H_1 : \pi \neq \pi_0$ (3) $H_1 : \rho \neq 0$ (4) $H_1 : \mu_1 - \mu_2 \neq 0$	<u>Nullhypothese:</u> rechtsseitig: $H_0 : \mu \leq \dots$ linksseitig: $H_0 : \mu \geq \dots$!!! Inhaltliche Hypothese !!! <u>Alternativhypothese:</u> rechtsseitig: $H_1 : \mu > \dots$ linksseitig: $H_1 : \mu < \dots$
2. Signifikanzniveau	$\alpha = 0,01 / 0,05$ (bei 1% / 5%)		$\alpha = 0,01 / 0,05$ (bei 1% / 5%)	
3. Auswahl d. Verfahrens	z-Prüfgröße: wähle Formel	t-Prüfgröße: wähle Formel	z-Prüfgröße: wähle Formel	t-Prüfgröße: wähle Formel
4. Stichprobe	Werte zum Einsetzen in die Prüfgröße sammeln	Werte zum Einsetzen in die Prüfgröße sammeln	Werte zum Einsetzen in die Prüfgröße sammeln	Werte zum Einsetzen in die Prüfgröße sammeln
5. Prüfgröße	z= Werte von 4 in 3 einsetzen	t= Werte von 4 in 3 einsetzen	z= Werte von 4 in 3 einsetzen	t= Werte von 4 in 3 einsetzen
6a. Kritische Grenze	$z_{1-\alpha/2}$ t-Tabelle unter ∞ nachschauen	$t_{1-\alpha/2; df}$ t-Tabelle unter df nachschauen	$z_{1-\alpha}$ t-Tabelle unter ∞ nachschauen Vorzeichen entsprechend H_1 !	$t_{1-\alpha; df}$ t-Tabelle unter df nachschauen Vorzeichen entsprechend H_1 !
7a. Entscheidung	$ z < z_{1-\alpha/2} : H_0$ beibehalten $ z \geq z_{1-\alpha/2} : H_1$ beibehalten	$ t < t_{1-\alpha/2; df} : H_0$ beibehalten $ t \geq t_{1-\alpha/2; df} : H_1$ beibehalten	Es wird die Hypothese beibehalten, dessen Vergleichszeichen bestätigt wird	Es wird die Hypothese beibehalten, dessen Vergleichszeichen bestätigt wird
6b. Überschreitungswahrscheinlichkeit P	$P = 2 \cdot (1 - \Phi(z))$		$P = 1 - \Phi(z)$	
7b. Entscheidung	$P > \alpha \Rightarrow H_0$ beibehalten $P \leq \alpha \Rightarrow H_1$ beibehalten		$P > \alpha \Rightarrow H_0$ beibehalten $P \leq \alpha \Rightarrow H_1$ beibehalten	

INDEX

—#—

- π *Siehe* Korrekturfaktor *Siehe*
Korrelationskoeffizient *Siehe* Signifikanzniveau
Siehe Wahrscheinlichkeit
- η^2 *Siehe* Chi-Quadrat *Siehe* Eta-Quadrat
- χ^2 -Anpassungstest 85
- χ^2 -Unabhängigkeitstest 83
- χ^2 -Werte 83, 85
- α -Adjustierung 47
- α -Fehler *Siehe* Fehlerarten *Siehe* Fehlerarten
- α -Fehlerkumulierung 47

—A—

- abhängige Stichproben 37
- abhängige Variable 35
- Ablehnungsbereich 13
- Alternativhypothese 11, 99
- Annahmebedingungen für die Nullhypothese
einseitiger Test (P-Wert) 26
linksseitiger Test (kritische Grenze) 26
rechtsseitiger Test (kritische Grenze) 25
Vereinfachungen 27
zweiseitiger Test (kritische Grenzen) 16
zweiseitiger Test (P-Wert) 22
- Annahmebereich 13
- approximativer Binomialtest 79
- Approximativer Binomialtest 28
- AV *Siehe* abhängige Variable

—B—

- Bedingung *Siehe* Faktor
- Bildungsgesetz 75
- Binomialkoeffizient 75
- Binomialverteilung 10, 75
- Block 52
- Blockmittel 53, 55

—C—

- Chi-Quadrat 83
- Cramer's V 83

—D—

- df *Siehe* Freiheitsgrad
- dichotom 9
- Dichotomisierung 10
- Durchschnittsverteilung 6, 7

—E—

- Effekt 35
- Effektstärke 51, 63
- einfache Haupteffekte 67, 71
- Einflußgröße *Siehe* Faktor
- Entscheidung 16, 19
- Erfolgszahl 78
- erwartete Häufigkeiten 83, 85
- Eta-Quadrat
bei einfaktorieller Varianzanalyse 51

- bei zweifaktorieller Varianzanalyse 63
- exakter Binomialtest 79
- Experimentalgruppe 35, 37

—F—

- Faktor 35, 46, 57
- Faktorstufe 53
- Faktorstufenkombination 58
- Faktorstufenmittel 53, 55
- falsifiziert 11, 16
- Fehlentscheidung 12
- Fehlerarten 17
- Freiheitsgrad 15
- Freiheitsgrad 9
- F-Test 44
- F-Verteilung 44
- F-Werte 44, 74

—G—

- gerichtete Hypothesen 23
- Gesamteffekt 60
- Gesamtmittel 49, 55, 59

—H—

- h.s. *Siehe* hochsignifikant
- H_0 *Siehe* Nullhypothese
- H_1 *Siehe* Alternativhypothese
- Häufigkeiten 9
- Haupteffekt Faktor A 60
- Haupteffekt Faktor B 61
- hochsignifikant 13, 22

—I—

- inhaltliche Hypothese 12, 99
- Interaktion 57, 67, 71
disordinale 70
Interaktionsdiagramm 67
Interaktionstypen 68
Parallelität der Linien 68
semiordinale (hybride) 70
Wechselwirkung AxB 62
- intraindividuelle Bedingungsvariation *Siehe*
Meßwiederholung
- IQ 12
- Irrtumswahrscheinlichkeit *Siehe* Signifikanzniveau

—K—

- klassischer t-Test 42, 47
- Konfidenzintervalle 29
Konfidenzkoeffizient *Siehe*
Sicherheitswahrscheinlichkeit
- Konfidenzniveau *Siehe*
Sicherheitswahrscheinlichkeit
- mathematische Herleitung 33
- Präzision 33
- Kontrolle von Störeinflüssen 37
- Kontrollgruppe 35, 37
- Korrekturfaktor 54
- Korrelationskoeffizient 34, 83
- kritische Grenzen 13, 15

kritischer Bereich 55

—M—

Meßwertpaardifferenzen 38
 Meßwiederholung 37, 52
 Mittelwertdifferenzen 36, 39
 Mittelwertdifferenzenverteilung 40
 Mittelwertvergleich 35
 mittlere Quadrate 48
 multipler Mittelwertvergleich 51

—N—

n.s. *Siehe* nicht signifikant
 Nennerfreiheitsgrad 45
 nicht signifikant 22
 nichtparametrisch 75
 Niveau der Verlässlichkeit *Siehe* Signifikanzniveau
 Nullhypothese 11, 99

—P—

p *Siehe* relative Häufigkeit
 parallelisierte Stichproben *Siehe* abhängige
 Stichproben
 Parallelisierung 52
 Parallelität der Linien 68
 Populationsparameter *Synonym zu* Parameter
 einer hypothetischen Grundgesamtheit
 Prüfgröße 14, 15
 Prüfverfahren *Siehe* Signifikanztest

—Q—

Quadrate
 von t-Werten 74
 von z-Werten 85

—R—

Rechteckverteilung 6
 relative Häufigkeit 9

—S—

s. *Siehe* signifikant
 s.s. *Siehe* sehr signifikant
 SAQ 49
 Satz der Streuungszerlegung
 einfaktoruell bei abhängigen Stichproben 53
 einfaktoruell bei unabhängigen Stichproben 48
 zweifaktoruell 62
 SCHEFFÉ-Test 51
 sehr signifikant 12, 22
 Sicherheitswahrscheinlichkeit 12, 30
 signifikant 11, 12, 22
 Signifikanzniveau 12, 14, 22
 Signifikanztest 11
 einseitig 23
 linksseitig 25
 rechtsseitig 24
 Spaltenmittel 59, 61
 spezifische Hypothesen *Siehe* gerichtete
 Hypothesen
 Standardfehler 8
 statistische Hypothesen 11

statistische Sicherheit *Siehe*
 Sicherheitswahrscheinlichkeit
 statistischer Test *Siehe* Signifikanztest
 Stichprobe 15
 Stichprobenverteilung 6, 13
 Störeinflüsse *Siehe* Störfaktor
 Störfaktor 57

—T—

Tafel der Varianzanalyse
 einfaktoruell bei abhängigen Stichproben 55
 einfaktoruell bei unabhängigen Stichproben 49
 zweifaktoruell 63
 Testverfahren *Siehe* Signifikanztest
 Treatment 53
 Trend 69
 t-Verteilung 8
 t-Werte 9, 15, 25

—Ü—

Überschreitungswahrscheinlichkeit
 einseitiger Test 26
 exakter Binomialtest 80, 81
 zweiseitiger Test 21
 überzufällig *Siehe* signifikant

—U—

unabhängige Stichproben 37
 unabhängige Variable *Siehe* Faktor
 ungerichtete Hypothesen 23
 unspezifische Hypothesen *Siehe* ungerichtete
 Hypothesen
 unverbundene Stichproben *Siehe* unabhängige
 Stichproben
 UV *Siehe* unabhängige Variable

—V—

Varianzanalyse 46
 einfaktoruell 46
 einfaktoruell für abhängige Stichproben 52
 einfaktoruell für unabhängige Stichproben 48
 mehrfaktoruell 46, 57
 Vergleich mit klassischem t-Test 74
 zweifaktoruell 58
 Varianzaufklärung 51, 63
 Varianzhomogenität 43, 44, 74
 Verallgemeinerung 16
 verbundene Stichproben *Siehe* abhängige
 Stichproben
 verifiziert 16
 verteilungsfrei 75
 Voraussetzungen
 bei Chi-Quadrat-Tests 84, 85
 bei Konfidenzintervallen 31
 bei Varianzanalysen 74
 beim approximativen Binomialtest 28, 79
 beim klassischen t-Test 43
 beim SCHEFFÉ-Test 51
 beim t-Test für abhängige Stichproben 38

—W—

Wahrscheinlichkeit 9
 Wahrscheinlichkeitsrechnung 75

Wechselwirkung *Siehe* Interaktion
Wechselwirkung AxB 62
WELCH-Test 41
Wirkung *Siehe* Effekt

—Z—

Zählerfreiheitsgrad 45

Zeilenmittel 58, 60
Zelle *Siehe* Faktorstufenkombination
Zellmittel 58, 60, 67
zentraler Grenzwertsatz 6
Zufallsstreuung 8, 48, 54, 57, 60, 61, 62, 83, 85
Zusammenhang *Siehe* Korrelationskoeffizient
z-Werte 15, 25