5 Parameterschätzung

In den ersten beiden Kapiteln haben wir uns mit der Beschreibenden Statistik beschäftigt. Sie dient dazu, vorhandenes Datenmaterial durch Berechnung charakteristischer Kennzahlen übersichtlicher zu machen.

Wir kommen nun zu der Schließenden Statistik. Sie dient dazu, aufgrund des vorliegenden Datenmaterials auf die dem zufälligen Vorgang der Datenerhebung zugrundeliegenden Gesetzmäßigkeiten zurückzuschließen, kurz: Von einer ________________ ________________ auf die ____ zu schließen.

Dabei bezeichnen wir als ____ oder auch ______________________________ die Menge aller zu untersuchenden Elemente, die Träger eines bestimmten Merkmals sind. Als __________________________ bezeichnen wir eine beschreibbare Teilmenge daraus.

Beispiel 33

(a) *Grundgesamtheit: Alle Wähler bei einer Bundestagswahl. Eine mögliche Stichprobe daraus wäre die Teilmenge derjeniger Wähler, die vor einem Wahllokal von einem Fernsehteam befragt werden.*

(b) *Grundgesamtheit: Alle Käufer eines bestimmten Produkts. Eine mögliche Stichprobe daraus wäre die Teilmenge derjeniger Käufer, die innerhalb einer bestimmten Zeitspanne Reklamationen an den Produkthersteller richten.*

Der Statistiker interessiert sich nun sehr oft für bestimmte Statistische Kennwerte der Grundgesamtheit. Oft jedoch ist es z.B. aus Kostengründen nicht möglich, mit der ganzen Grundgesamtheit eine Statistische Erhebung durchzuführen, um diese Statistischen Kennwerte zu ermitteln. So begnügt man sich meist damit, eine Statistische Erhebung nur mit einer Stichprobe durchzuführen, um durch die Statistischen Kennwerte der Stichprobe (sog. ______________________________

____________________________________) wenigstens Näherungswerte für die gesuchten Statistischen Kennwerte der Grundgesamtheit zu erhalten.

Diese aus irgendwelchen Gründen nicht ermittelbaren Statistischen Kennwerte der Grundgesamtheit bezeichnet man als

____ .

Mögliche Parameter sind: __________________________________ oder

_____________________________ eines Merkmals der Grundgesamtheit.

Um jedoch deutlich zu machen, dass damit die nicht ermittelbaren (oder anders ausgedrückt: die gesuchten) Statistischen Kennwerte der Grundgesamtheit gemeint sind, verwendet man für die Parameter statt _____ und _____ die griechischen Buchstaben _____ und _____ (wie bei Zufallsvariablen).

Im Unterschied dazu bezeichnen wir den ermittelbaren (oder anders ausgedrückt: den gegebenen) Stichprobenmittelwert mit _____ und die Stichprobenvarianz mit _____ (dazu kommen wir später noch).

Es gibt verschiedene Möglichkeiten, eine Stichprobe auszuwählen; der Statistiker sagt: eine Stichprobe zu ziehen. Generell sollte die Stichprobe stets sehr sorgfältig gezogen werden, damit sie die Grundgesamtheit möglichst genau repräsentiert. Eine solche Stichprobe nennt man dann eine repräsentative Stichprobe.

BEISPIEL 34 *(Zum letzten Beispiel)*

(a) Je nach Zeitpunkt der Befragung und je nach Wahllokal sind beliebig viele unterschiedliche Stichproben möglich.

(b) Je nach Zeitspanne und Produkt sind auch hier beliebig viele unterschiedliche Stichproben möglich.

Die letzten beiden Beispiele machen deutlich, dass sich die Ziehung einer Stichprobe aus einer Grundgesamtheit in gewisser Weise als ein ____________________________

______________________________ betrachten lässt.

Für eine Stichprobe der Größe n können wir dann die Befragung der i-ten Person (z.B. im Kontext der letzten beiden Beispiele (a)) jeweils durch eine __________________ (für i=1,.....,n) beschreiben, wobei die _________________________ als i.i.d. angenommen werden.

Das Befragungsergebnis der i-ten Person ist dann eine Realisierung der _______________ (für i=1,.....,n). Diese Realisierung schreiben wir mit Kleinbuchstaben _____ als Unterscheidung zur Zufallsvariablen _____ (für i=1,.....,n).

Der Statistiker redet dann oft von der Stichprobe ______________________________ und meint damit den eben beschriebenen Sachverhalt.

Wie zieht nun der Statistiker Rückschlüsse von der Stichprobe auf die Grundgesamtheit, d.h. wie geht das?

Dazu verwendet er sog. ____________________ . Ein ____________________ ist eine n-dimensionale Funktion mit den Zufallsvariablen ____________________ als Platzhaler, die angibt, wie man den gesuchten Parameter der Grundgesamtheit aus den Stichprobenergebnissen näherungsweise berechnen kann.

Wenn man dann die vorliegenden Stichprobenrealiesierungen ____________________ anstelle der Platzhalter ____________________ in die Schätzfunktion einsetzt, erhält man einen konkreten Schätzwert.

Ein häufig verwendeter Schätzer für den Mittelwert μ der Grundgesamtheit ist ______________________________________ . _____ ist selbst wieder eine __________ .

Durch Einsetzen der Realisierungen einer bestimmten Stichprobe erhält man den Schätzwert ____ , das ist gerade der

____ .

Beispiel 35

Eine Sportlehrerin interessiert sich für die durchschnittliche Körpergröße 12-jähriger Jungen, um ihren Geräteaufbau bei einem Sportfest für diese Altersklasse optimal planen zu können.

Dazu misst sie die Größe von zehn 12-jährigen Jungen aus ihrer Schulklasse und berechnet aus diesen zehn Messergebnissen den (Stichproben-)Mittelwert.

Die vorliegende Grundgesamtheit ist hier die Menge aller ______________________ ______________ *. Der gesuchte Parameter der Grundgesamtheit ist die* ______________ ______________________________ *eines* ______________________________________ .

Die von der Sportlehrerin gezogene Stichprobe sind die ____________________________ ___________________ *aus ihrer Schulklasse.*

Wir können nun die Größe jeder dieser zehn Jungen durch eine ___________________ *beschreiben (für i=1,.....,10), wobei wir diese* ____________________________ *als i.i.d. annehmen.*

Der von der Lehrerin verwendete Schätzer ist _________________________________ .
Durch Einsetzen der Messergebnisse ___________________ *erhält sie den Schätzwert* ____________________________ *, das ist gerade der Stichprobenmittelwert, als näherungsweise Schätzung für den unbekannten Parameter* _____ .

Wenn wir aus derselben Grundgesamtheit nicht nur eine, sondern beliebig viele Stichproben ziehen, können wir den Stichprobenmittelwert _____ für jede einzelne Stichprobe berechnen. Diese Stichprobenmittelwerte werden mehr oder weniger stark vom Populationsparameter _____ abweichen. Je weniger sie von _____ abweichen, desto besser oder genauer schätzt ein Stichprobenmittelwert _____ den Parameter _____ .

Anders ausgedrückt: Je geringer die Streuung der Verteilung der Zufallsvariablen _____ (bei diesen vielen Stichproben) ausfällt, d.h. je geringer die Schwankung von _____ um _____ , desto besser schätzt ein einzelner Stichprobenmittelwert _____ den unbekannten Parameter _____ .

Dies alles gilt nicht nur für den Stichprobenmittelwert $\overline{x}$, sondern auch für jeden anderen Stichprobenkennwert, z.B. für die Stichprobenvarianz. Es gilt stets: Je geringer die ____________________ der sog. Stichprobenkennwerteverteilung, desto genauer schätzt ein einzelner Stichprobenkennwert den gesuchten Parameter.

Die Streuung der Stichprobenkennwerteverteilung heißt _________________________ .
Speziell heisst die Streuung von $\overline{X}$ Standardfehler des Mittelwerts und wird mit __________ bezeichnet.

Es gilt nun: Je größer die Streuung der Messwerte in der Grundgesamtheit, desto größer ist auch __________ . Und: Je größer der _________________________________ _______________________ , desto kleiner ist __________ .
(Denn: Der wachsende ____ nähert sich immer mehr der Größe der Grundgesamtheit, und demzufolge bildet auch die Stichprobe immer besser die Eigenschaften der Grundgesamtheit ab, was insbesondere für den Mittelwert gilt. Und dies bedeutet gerade, dass die Streuung von _____ immer geringer wird.)

Es gibt in der Statistik viele verschiedene Arten, Schätzer zu definieren. Das Ziel bei der Definition sollte immer sein, einen solchen Schätzer (d.h. eine solche Schätzfunktion) zu finden, dessen (zur Stichprobe gehöriger) Schätzwert den gesuchten Parameter der ________________________________ am besten schätzt, d.h. am besten annähert.

Dazu wurden in der Statistik Gütekriterien entwickelt, welche die verschiedenen Schätzer beurteilen. Die bekanntesten Gütekriterien sind ________________________ ______________ , ______________________ , ___________________ und __________ ___________________ . Wir werden hier nur die ________________________________ ______________ herausgreifen:

Ein Schätzer heißt _______________________________________ , falls der __________________________________ des Schätzers gleich dem gesuchten Parameter der Grundgesamtheit ist, falls also der Schätzer den gesuchten Parameter im Mittel auch wirklich trifft.

Beispiel 36 *(Zum letzten Beispiel)*

Der von der Lehrerin verwendete Schätzer ist _____________________________________ .

Wegen ________________________________ *ist* _________________________________ .

Damit und wegen den Rechenregeln für den Erwartungswert gilt:

womit das Kriterium für die Erwartungstreue erfüllt ist. Somit ist _____ *ein erwartungstreuer Schätzer für den Mittelwert* _____ *eines Merkmals der Grundgesamtheit.*

Ein erwartungstreuer Schätzer für die Varianz _____ eines betrachteten Merkmals in der Grundgesamtheit ist

Die Forderung der Erwartungstreue bedingt den Vorfaktor __________ im Unterschied zur Definition der Varianz _____ einer Messreihe; hier ist der Vorfaktor _____ :

Falls wir in den erwartungstreuen Schätzer _____ die Realisierungen $x_1, \ldots, x_n$ einer Stichprobe einsetzen, dann nennen wir den so erhaltenen Schätzwert

Die positive Quadratwurzel daraus:

heisst ____ .

Der zugehörige Schätzer

ist allerdings __________ erwartungstreuer Schätzer für die Streuung _____ des betrachteten Merkmals in der Grundgesamtheit.

Der gesuchte Parameter der Grundgesamtheit muss aber nicht unbedingt _____ oder _____ sein, es können auch andere ____ sein.

Generell gibt es verschiedene Methoden, wie man einen Schätzer für einen beliebig vorgegebenen Parameter der Grundgesamtheit bestimmen kann, wie z.B. die sog.

____

oder die sog. ________________________ .

Wir werden hier eine häufig verwendete Methode zur Bestimmung eines Schätzers kennenlernen, die sog. ______________________________

__________________ . Mit dieser Methode lässt sich ein Schätzer für einen unbekannten Parameter der Grundgesamtheit ermitteln, vorausgesetzt die Verteilung des untersuchten Merkmals ist bekannt.

Dass heißt, die Zufallsvariable X_i (für $i = 1, \ldots, n$) der i.i.d. Stichprobe $X_1, \ldots, X_n$ besitzt eine vom unbekannten Parameter θ der Grundgesamtheit abhängige bekannte Wahrscheinlichkeits- oder Dichtefunktion ______ .

Man wähle nun zur Stichprobenrealisierung $x_1, \ldots, x_n$ denjenigen Wert ____ als Schätzwert für den unbekannten Parameter ____ der Grundgesamtheit, unter dem die Wahrscheinlichkeit für das Eintreten dieses Ergebnisses am größten (bzw. die entsprechende Wahrscheinlichkeitsdichten) ist.

Die Funktion ____

wird Likelihood-Funktion genannt. Jedes _____ , welches die Likelihood-Funktion maximiert ist ein sogenannter Maximum-Likelihood-Schätzwert.

Beispiel 37

Ein Hersteller produziert Blitzgeräte. Er interessiert sich für die Wahrscheinlichkeit _____ , mit der ein defektes Blitzgerät produziert wird. Diese Wahrscheinlichkeit _____ ist also der gesuchte Parameter der Grundgesamtheit aller produzierten Blitzgeräte.

Zur Bestimmung des Parameters _____ entnimmt der Hersteller eine Stichprobe von _____ produzierten Blitzgeräten und stellt fest, dass davon _____ Blitzgeräte defekt sind.

Die Stichprobe kann somit durch _____ i.i.d. ___________________________ Zufallsvariablen ________________ beschrieben werden mit den Ausprägungen __________ , falls Blitzgerät intakt und __________ , falls Blitzgerät defekt (für i=1,.....,n). Es gilt:

Gesucht wird dasjenige _____ , für das die Likelihood-Funktion ________________ wird.

Dazu wenden wir zuerst auf beiden Seiten der obigen Gleichung den Logarithmus an, denn dabei bleiben die Maximalstellen unverändert. Es ergibt sich unter Beachtung der Logarithmus-Rechenregeln:

Jetzt differenzieren wir und setzen die erste Ableitung gleich Null, um eine Extremstelle zu ermitteln:

Die so gefundene Extremstelle ____________ ist tatsächlich ein Maximum, wie man anhand des negativen Vorzeichens der zweiten Ableitung überprüfen kann.

Das bedeutet: Der durch diese Methode gefundene sog. ________________________ ____ für den Parameter _____ der Grundgesamtheit ist _____ , und das ist gerade die relative Häufigkeit defekter Blitzgeräte in der Stichprobe.

6 Intervallschätzung

Bei der Parameterschätzung haben wir einen Schätzwert für einen gesuchten Parameter bestimmt. Der konkret gefundene Schätzwert sagt aber noch nichts darüber aus, wie groß seine Abweichung vom gesuchten Parameter in Wirklichkeit ist.

Bei der Intervallschätzung wird nicht ein einzelner Schätzwert ermittelt, sondern ein ganzes Intervall, in dem der gesuchte Parameter mit einer bestimmten vorgegebenen Wahrscheinlichkeit $1-\alpha$ darin liegt. Solche Intervalle heißen ____________________ ______________________ (oder ____) zum ____ .

Um eine gewisse Genauigkeit zu gewährleisten, sollten Konfidenzintervalle möglichst klein sein. Generell gilt, dass der Stichprobenumfang die Größe des Konfidenzintervalls beeinflusst, und zwar benötigt man für ein _______________ Konfidenzintervall einen _________________ Stichprobenumfang bei gleichem α.

Ein Konfidenzintervall ist eindeutig festgelegt durch seine obere und untere ____ .

Für Diese gibt es Berechnungsformeln, sog. Konfidenzintervall-Formeln, in denen sich die Zufallsvariablen $X_1, \ldots, X_n$ als Platzhalter für die Stichprobenrealisierungen $x_1, \ldots, x_n$ befinden.

Sei also $X_1, \ldots, X_n$ eine Stichprobe i.i.d. __________ -verteilter Zufallsvariablen.

Wir unterscheiden zunächst vier Fälle:

Fall 1:

Der gesuchte Parameter der Grundgesamtheit sei der ________________________ .

Die ______________________ aus der ________________________________

sei bekannt. Dann berechnet sich ein Konfidenzintervall für ____ zum Konfidenzniveau _____________ gemäß der Formel:

dabei ist ____ der Stichprobenmittelwert und __________ ist das ____________ – _____________ der Standardnormalverteilung N(0,1).

Die wichtigsten Quantile der N(0,1)-Verteilung zur Konfidenzintervall-Berechnung finden sich in folgender sog. "**z-Tabelle**": (Quantile z_p der N(0,1)-Verteilung)

p	0.55	0.6	0.7	0.75	0.8	0.85	0.9	0.95	0.975	0.99	0.995
z_p	0.126	0.253	0.524	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576

Weitere Quantile der N(0,1)-Verteilung erhält man mit der Gleichung:

____________________ , z.B.: ________________________________ .

Beispiel 38

Bei einer Untersuchung über das Verhalten von Schulkindern im Straßenverkehr interessiert sich ein Psychologe für den Erwartungswert μ der Reaktionszeit von 10-jährigen Schülern in einer bestimmten Verkehrssituation.

Aus früheren Untersuchungen weiß er, dass sich die Reaktionszeit durch eine ____________________ -verteilte Zufallsvariable beschreiben lässt. Bei 61 Messungen errechnete er einen Stichprobenmittelwert für die Reaktionszeit von $\bar{x} = 0.8$ s.

Für die Berechnung der Konfidenzintervall-Grenzen benötigen wir noch den zugehörigen Wert aus der z-Tabelle:

Damit ergibt sich insgesamt das folgende Konfidenzintervall:

Wir haben hier ein Konfidenzniveau von __________ vorliegen. Das bedeutet exakt formuliert: Bei _____________ aller gleichgroßen Stichproben liefert dieses Konfidenzintervall-Schätz-Verfahren ein Intervall, in dem der gesuchte Parameter μ auch wirklich drinliegt. Bei den restlichen ________ liefert es ein Intervall, in dem der gesuchte Parameter μ nicht drinliegt.

Veranschaulichen lässt sich dieser Sachverhalt wie folgt:

Die exakte Interpretation unseres oben ausgerechneten Konfidenzintervall lautet somit: Falls unsere Stichprobe zu den __________ "zutreffenden" Stichproben gehört, dann liegt die erwartete Reaktionszeit μ *von 10-jährigen Schülern (in dieser betimmten Verkehrssituation) zwischen _______ und _______ Sekunden.*

Fall 2:

Genauso wie Fall 1, nur diesmal sei die Varianz σ^2 unbekannt. Diese Veränderung bewirkt, dass in der Formel statt σ^2 die Stichprobenvarianz $\bar{S}^2$ vorkommt und statt N(0,1)-Quantile kommen Quantile der t-Verteilung vor. In diesem Fall berechnet sich ein Konfidenzintervall für μ zum Konfidenzniveau $1 - \alpha$ gemäß der Formel:

wobei ____________________ das ______________________________ der __________-Verteilung ist. Die Stichprobenvarianz $\bar{S}^2$ berechnet sich gemäß der Formel:

Beispiel 39 *(Fortsetzung des letzten Beispiels)*

Angenommen, die Varianz σ^2 sei unbekannt, und aus den Realisierungen der Stichprobe errechnet sich eine Stichprobenvarianz von $\bar{S}^2 = 0,0484$. Aus Tabelle D ermitteln wir: ___________________________________ *. Damit ergibt sich insgesamt für das gesuchte Konfidenzintervall:*

Dieses Konfidenzintervall ist größer als in Fall 1, weil wir hier in Fall 2 wegen der unbekannten Varianz auch weniger Information vorliegen haben.

Fall 3:

Der gesuchte Parameter der Grundgesamtheit sei die Varianz σ^2. Der Mittelwert μ aus der $N(\mu, \sigma^2)$-Verteilung sei bekannt. Dann berechnet sich ein Konfidenzintervall für σ^2 zum Konfidenzniveau $1 - \alpha$ gemäß der Formel:

Dabei ist ________________ das ________________________________ der ____________________________ und ____________ ist das ______________________ der ________________________________ .

Beispiel 40

Um die Größe eines Regenauffangbeckens besser planen zu können, interessiert sich ein Gärtner für die Varianz σ^2 der Niederschlagsmengen im regenreichsten Monat April. Als Stichprobe liegen ihm die Niederschlagsmengen [in mm] seiner Stadt vom Monat April der letzten 20 Jahre vor.

Es wird angenommen, dass die vorliegenden Messwerte Realisierungen von 20 i.i.d. ____________________ -verteilten Zufallsvariablen sind. Für die Summe errechnet sich der Wert ________________________________ .

Aus Tabelle C haben wir: ____________________________________ und ____________________________________ .

Damit ergibt sich folgendes Konfidenzintervall zum Konfidenzniveau $0,95$:

Fall 4:

Genauso wie Fall 3, nur diesmal sei der Mittelwert μ unbekannt. Diese Veränderung bewirkt, dass in der Formel statt der Summe ein Ausdruck mit der Stichprobenvarianz $\bar{S}^2$ auftaucht und die Freiheitsgrade der χ^2-Quantile im Nenner jeweils um eines herabgesetzt werden. In diesem Fall berechnet sich ein Konfidenzintervall für σ^2 zum Konfidenzniveau $1 - \alpha$ gemäß der Formel:

Dabei ist $\chi^2_{n-1;1-\frac{\alpha}{2}}$ das $1 - \frac{\alpha}{2}$-Quantil der χ^2_{n-1}-Verteilung und $\chi^2_{n-1;\frac{\alpha}{2}}$ ist das $\frac{\alpha}{2}$-Quantil der χ^2_{n-1}-Verteilung.

Beispiel 41 *(Fortsetzung des letzten Beipiels)*

Angenommen, der Mittelwert μ sei unbekannt, und aus den Realisierungen der Stichprobe errechnet sich eine Stichprobenvarianz von $\bar{S}^2 = 51,5$. Aus Tabelle C haben wir: ________________________________ und ______________________________ .

Dann ergibt sich insgesamt für das gesuchte Konfidenzintervall:

Diese Konfidenzintervall ist größer als in Fall 3, weil wir hier in Fall 4 durch den unbekannten Mittelwert auch weniger Informationen vorliegen haben.

In den letzten vier Fällen hatten wir stets normalverteilte Zufallsvariablen vorliegen. In manchen Fällen kann mit Hilfe von Grenzwertsätzen (wie z.B. dem zentralen Grenzwertsatz) Konfidenzintervalle näherungsweise bestimmt werden.

Fall 5:

Sei $X_1, \ldots, X_n$ eine Stichprobe i.i.d. verteilter Zufallsvariablen mit unbekannten Mittelwert μ und bekannter Varianz σ^2. In diesem Fall berechnet sich für große n näherungsweise ein Konfidenzintervall für μ zum Konfidenzniveau $1-\alpha$ gemäß der Formel aus dem Fall 1:

dabei ist $\bar{x}$ der Stichprobenmittelwert und $z_{1-\frac{\alpha}{2}}$ ist das $1-\frac{\alpha}{2}$-Quantil der Standardnormalverteilung $N(0,1)$.

Fall 6:

Sei $X_1, \ldots, X_n$ eine Stichprobe i.i.d. verteilter Zufallsvariablen mit unbekannten Mittelwert μ und unbekannter Varianz σ^2. In diesem Fall berechnet sich für große n näherungsweise ein Konfidenzintervall für μ zum Konfidenzniveau $1-\alpha$ gemäß der Formel:

dabei ist $\bar{x}$ der Stichprobenmittelwert, S^2 die Varianz und $z_{1-\frac{\alpha}{2}}$ ist das $1-\frac{\alpha}{2}$-Quantil der Standardnormalverteilung $N(0,1)$.

Beispiel 42

Vor einer Wahl möchte ein Meinungsforscher den Anteil p *der Wähler von Partei A unter den Wahlberechtigten ermitteln; d.h. die Grundgesamtheit ist die Menge aller Wahlberechtigten und der gesuchte Parameter* p *ist der Anteil der Wähler von Partei A.*

Dazu befragt er 35 *zufällig ausgewählte Wahlberechtigte nach ihrer Wahlabsicht.* 14 *davon wollen Partei A wählen. Der Meinungsforscher möchte nun wissen, in welchem Bereich* p *liegt zum Konfidenzniveau* 0.95.

Hier ist $X_1, \ldots, X_n$ *eine Stichprobe i.i.d.* ______________ *-verteilter Zufallsvariablen. Dabei sei die Erfolgswahrscheinlichkeit* p *der unbekannte gesuchte Parameter der Grundgesamtheit. Es gilt:*

Da die Varianz ebenfalls unbekannt ist, sind wir im Fall 6. *Sei* k *die Anzahl der Befragten mit der Wahlabsicht Partei A (bzw. die Anzahl der Treffer in der Stichprobe).*

Daraus ergibt sich folgendes Konfidenzintervall:

Hier haben wir:

Dieses Konfidenzintervall ist dem Meinungsforscher zu groß. Er will es genauer wissen und erhöht den Stichprobenumfang. Er befragt nochmal 165 *zufällig ausgewählte Wahlberechtigte nach ihrer Wahlabsicht und findet darunter* 66 *Wähler von Partei A, d.h. insgesamt ist jetzt:*

Dafür berechnet er nochmals ein Konfidenzintervall für p zum Konfidenzniveau 0.95 *und erhält*

D.h. zu 95% *oder mit einer Wahrscheinlichkeit von* 0.95 *wird Partei A in der Wahl einen Stimmenanteil p zwischen* ____________________ *und* ____________________ *erreichen können. Mit diesem Ergebnis gibt sich der Meinungsforscher zufrieden.*

7 Statistische Tests

7.1 Einführung

Wir kommen nun zum wichtigsten Kapitel in dieser Vorlesung: Statistische Tests.

In den letzten beiden Kapiteln haben wir uns damit beschäftigt, Schätzungen für einen unbekannten Parameter der Grundgesamtheit anhand den Ergebnissen einer Stichprobe zu finden.

Häufig geht es aber in der Statistik auch darum, ______________________________ zu fällen in Situationen, in denen sich eine neue und eine althergebrachte Meinung über einen bestimmten Sachverhalt gegenüberstehen und dem Statistiker die Frage gestellt wird: Welche Meinung ist besser?

Die althergebrachte Meinung nennt man in der Statistik ________________________ __________ und die neue Meinung ______________________________________ . Dabei war die althergebrachte Meinung (also die ______________________________) die bisher übliche und gültige solange, bis eine neue, dazu konkurrierende Meinung (die ______________________________________) aufgetaucht ist, die es jetzt zu überprüfen gilt.

Die Statistik hat Verfahren entwickelt, welche ______________________________ ____________________ liefern dafür, wann eine althergebrachte Meinung abzulehnen ist und wann nicht. Diese Verfahren heißen Statistische Tests.

Dabei gilt: Im Fall der ________________________ der Nullhypothese (wir sagen auch: Die Nullhypothese wird ________________________) hat sich die Alternativhypothese durchgesetzt;

Im Fall der ____________________________________ der Nullhypothese wird die Nullhypothese beibehalten (d.h. sie wird _________________________________) solange, bis eine neue Alternativhypothese auftaucht und der Entscheidungsprozess erneut beginnt.

In diesem Sinne sind die Statistischen Tests nichts anderes als ___________________ ________________________ für die _________________________________ !

Beispiel 43

Ein Lehrer behauptet, eine neue Unterrichtsmethode sei besser als die herkömmliche. Hier gilt folgende Einteilung:

____ *: Die neue Unterrichtsmethode ist besser als die alte.*

_________________________________ *: Die alte Unterrichtsmethode ist besser oder genausogut wie die neue. Alternative Formulierung: Die beiden Unterrichtsmethoden unterscheiden sich nicht oder die neue Methode ist sogar schlechter als die alte.*

Im letzten Beispiel wird durch die Worte ________________ und ________________________ eine Wertung in eine bestimmte Richtung zum Ausdruck gebracht bei der Formulierung von Nullhypothese und Alternativhypothese. Man spricht dann von einem _________________________________ .

Möchte man in einer neutralen Art und Weise einfach nur überprüfen, ob die neue Methode eine Veränderung bewirkt (egal in welche Richtung) oder nicht, so spricht man von einem ____________________ Test und verwendet folgende Formulierung:

Beispiel 44 *(Zum letzten Beispiel)*

Alternativhypothese: Die neue Unterrichtsmethode unterscheidet sich von der alten (egal in welche Richtung).

Nullhypothese: Es gibt keinen Unterschied zwischen den beiden Unterrichtsmethoden (beide sind gleichgut oder gleichschlecht).

In der Statistischen Testtheorie wird die Nullhypothese mit _____ bezeichnet und die Alternativhypothese mit _____ . Sobald man nun _____ und _____ aufgestellt hat, wird wie folgt weiterverfahren:

Es wird eine Stichprobe gezogen. Dann werden die Stichprobenrealisierungen in einen zu der vorliegenden Test-Situation passenden ____________________ eingesetzt, dessen Verteilung (unter H_0) bekannt ist, und ein ____________________ ausgerechnet.

Ein solcher Schätzer heißt Testgröße oder Prüfgröße und wird mit ____________________ bezeichnet. Der zugehörige Schätzwert heißt Wert der Testgröße oder Wert der Prüfgröße und wird mit ____________________ bezeichnet.

Schließlich wird überprüft, ob der errechnete Wert der Testgröße in einen zu der vorliegenden Test-Situation passenden ______________________________

(man sagt auch: ______________________________) fällt und daraufhin folgende Entscheidung getroffen:

Falls ____________________ in den ______________________

__________ fällt, wird H_0 verworfen, ansonsten beibehalten (d.h. nicht verworfen). D.h. im ersten Fall hat sich _____ durchgesetzt und im zweiten Fall _____ .

Wie wir aber in den letzten beiden Kapiteln gelernt haben, unterliegt eine Stichprobe immer Zufallsschwankungen, weil sie die ____________ Situation eines Sachverhalts in der Grundgesamtheit immer nur ______________________ wiedergeben kann. Das bedeutet: Der Statistiker macht bei seiner Entscheidung für oder gegen eine Nullhypothese zwangsläufig Fehler.

Dabei sind zwei verschiedene Fehler möglich:

1.) der sog. ______________ oder ______________________________ ,

nämlich: Die Nullhypothese zu verwerfen, obwohl sie richtig ist;

2.) der sog. ______________ oder ______________________________ ,

nämlich: Die Nullhypothese nicht zu verwerfen, obwohl sie falsch ist.

Beispiel 45 *(Zum vorletzten Beispiel)*

Wenn wir hier einen α-Fehler oder einen Fehler erster Art begehen, dann entscheiden wir uns fälschlicherweise ______________ *die Nullhypothese, obwohl sie* ________________ *ist; d.h. es wird fälschlicherweise angenommen, die neue Lehrmethode sei besser als die alte, was z.B. zu teuren und unnötigen Umschulungsmaßnahmen der Lehrer führen könnte.*

Wenn wir hier einen β-Fehler oder einen Fehler zweiter Art begehen, dann entscheiden wir uns fälschlicherweise _______ *die Nullhypothese, obwohl sie* ______________ *ist; d.h. es wird fälschlicherweise angenommen, die neue Lehrmethode sei nicht besser oder sogar schlechter als die alte, was dazu führen könnte, dass die Schüler weiterhin nach der alten Lehrmethode unterrichtet werden anstatt mit der besseren neuen Lehrmethode und eine Chance zu Fortschritt vertan wird.*

Die Wahrscheinlichkeit für das Auftreten eines α-Fehlers bezeichnet man als Irrtumswahrscheinlichkeit. Die Irrtumswahrscheinlichkeit ist eine bedingte Wahrscheinlichkeit, nämlich:

Die Irrtumswahrscheinlichkeit spielt in einem Statistischen Test eine große Rolle für den Ablehnungsbereich. Der ist nämlich stets so konzipiert, dass die Irrtumswahrscheinlichkeit kleiner oder gleich einem bestimmten Prozentwert α ist.

Diese obere Grenze der Irrtumswahrcheinlichkeit bezeichnet man als ______________

____________________ oder als ____ .

Die Statistiker sprechen dann davon, dass sie sich aufgrund einer vorliegenden Stichprobe für oder gegen die Ablehnung der Nullhypothese

"auf einem bestimmten ____ "

entscheiden und meinen damit, dass sie ihre Entscheidung (für oder gegen die Ablehnung der Nullhypothese) mit einer bestimmten vorgegebenen Höchst - Irrtumswahrscheinlichkeit treffen.

Natürlich sollte nicht nur die Wahrscheinlichkeit für einen α-Fehler, sondern auch die Wahrscheinlichkeit für einen β-Fehler möglichst klein gehalten werden. Beide zusammen klein zu halten, ist aber leider nicht möglich, denn sie verhalten sich gegenläufig.

Was aber möglich ist: Bei festen kleingehaltenen _____ betrachtet man ____________ ,
wobei wir hier mit _____ die Wahrscheinlichkeit des Auftretens eines β-Fehlers bezeichnen, d.h. die Wahrscheinlichkeit, eine Nullhypothese ______________ zu verwerfen, obwohl sie ______________ ist.

$1 - \beta$ ist dann die zugehörige Gegenwahrscheinlichkeit, d.h. die Wahrscheinlichkeit, eine falsche Nullhypothese auch wirklich zu verwerfen. $1-\beta$ heißt ____________________ oder "power" des Tests. Erwünscht ist eine möglichst große ________________________ (zu einem festen vorgegebenen Signifikanzniveau _____ !)

Es gilt: Die Teststärke wird umso größer, je größer der ______________________

______________________ ist.

Sowie: Die Teststärke wird umso größer, je kleiner die ____________________

des betrachteten Merkmals in der Grundgesamtheit ist.

Wir fassen zusammen:

Ein Statistischer Test auf einem Signifikanzniveau α ist ein Entscheidungsverfahren für oder gegen die Verwerfung einer aufgestellten Hypothese.

Aus den Realisierungen einer Stichprobe wird der Wert einer Testgröße ermittelt und geprüft, ob er in einen (vorher festgelegten) Ablehnungsbereich fällt oder nicht.

Falls ja, dann wird die Nullhypothese verworfen und die Alternativhypothese hat sich durchgesetzt. Falls nein, dann wird die Nullhypothese nicht verworfen, d.h. sie wird beibehalten.

Dabei ist der Ablehnungsbereich stets so konzipiert, dass eine ________________ Nullhypothese nur mit einer Wahrscheinlichkeit kleiner oder gleich dem Signifikanzniveau α verworfen wird; α wird vorher festgelegt.

Auf diesem vorher festgelegten Signifikanzniveau α sollte der Statistische Test eine möglichst große Teststärke aufweisen (d.h. eine ______________ Nullhypothese sollte mit einer möglichst großen Wahrscheinlichkeit auch wirklich verworfen werden).

7.2 T-Test und Gauß-Test

Wir kommen nun zu konkreten Statistischen Tests:

1.) <u>T-Test</u> (für eine Stichprobe)

Gegeben sei eine Stichprobe $X_1, \dots, X_n$ i.i.d. ____________________ - verteilter Zufallsvariablen mit unbekannten Erwartungswert μ und unbekannter Varianz σ^2.

Fall 1: (zweiseitiger Test)

Fall 2: (einseitiger Test)

Fall 3: (einseitiger Test)

Testgröße: (für alle drei Fälle)

wobei ____ der Stichprobenmittelwert ist

und ____ die Stichprobenvarianz ist.

Verteilung der Testgröße:

Falls wir annehmen, dass ____________ gilt, dann ist ________________________

_____________________________ .

Ablenhungsbereiche:

Ablehnung von H_0, falls

Fall 1: ____

Fall 2: ____

Fall 3: ____ ,

wobei α das (vorher festgelegte) Signifikanzniveau des Tests ist. (Die Quantile der t_{n-1}–Verteilung finden sich in Tabelle D.)

Beispiel 46

An einer Schule wird ein neuer, junger Sportlehrer eingestellt, der seine Schüler nach einer neuen Trainingsmethode unterrichten will. Ein bereits an der Schule unterrichtender älterer Sportlehrer trainiert seine Schüler schon seit Jahren nach einer alten Trainingsmethode. Beide Sportlehrer möchten ihre unterschiedlichen Trainingsmethoden miteinander vergleichen. Eine Möglichkeit des Vergleichs von neuer und alter Trainingsmethode ist die folgende:

Eine Gruppe von n Schülern wird mit der neuen Trainingsmethode unterrichtet und im Anschluss daran wird ein Leistungsnachweis durchgeführt, dessen Ergebnisse mit den Erfahrungswerten desselben Leistungstests unter der alten Trainingsmethode verglichen werden. Der Leistungstest ist ein Sprint auf einer bestimmten agbesteckten Strecke entlang des Schulgebäudes, die schon immer für Schnelllauftrainings benutzt wurde, und für die genügend Daten aus Wettläufen vergangener Jahre vorliegen.

Wir gehen davon aus, dass sich die Ergebnisse [d.h. Zeit in Sekunden] der n Schüler in diesem Schnelllauf durch i.i.d. $N(\mu, \sigma^2)$*-verteilte Zufallsvariablen* $X_1, \ldots, X_n$ *beschreiben lassen. Wir wissen nun aus Erfahrung, dass der Mittelwert aus dem vergleichbaren Leistungstest unter der alten Trainingsmethode gleich* _____ *ist. Betrachten wir zuerst die* ______________________ *Testsituation. Hier ist die Nullhypothese:* ____________________ *, d.h. die Mittelwerte der Leistungstests beider Trainingsmethoden unterscheiden sich nicht.*
Und die Alternativhypothese ist: ____________________ *, d.h. es gibt einen Unterschied zwischen den Mittelwerten der Leistungstests beider Trainingsmethoden.*

Falls nun der Stichprobenmittelwert _____ *nahe bei* _____ *liegt, so wird* _____ *durch das Stichprobenergebnis unterstützt. In diesem Fall ist die Differenz* _______________ *klein, im Idealfall gleich Null. Falls* _____ *weit weg von* _____ *liegt, d.h. falls die Differenz* _______________ *groß ist, so wird* _______ *durch das Stichprobenergebnis* ____________ *unterstützt.*

Wir berechnen nun den Wert der Testgröße durch Einsetzen der Stichproben-Realisie—rungen $x_1, \ldots, x_n$ *in* _____ *und erhalten mit*

genau diese Differenz __________________ *dividiert durch eine positive Konstante, d.h. es gilt auch hier: Kleine Werte (nahe bei Null) von* $T(x_1, \ldots, x_n)$ *unterstützen* _____ *und große Werte von* $T(x_1, \ldots, x_n)$ *unterstützen* _______________________ *.*
Wir wissen nun: Falls _______ *zutrifft, so ist die Testgröße* __________ *-verteilt mit Erwartungswert* _____ *, d.h. es liegt folgende Situation vor:*

Die Gesamtfläche unter der Dichtekurve ist 1. *Die beiden eingezeichneten Quantile der* __________ *-Verteilung schneiden von der Gesamtfläche jeweils* _____ *Flächen-einheiten ab, d.h. in der Mitte bleibt ein Flächeninhalt von* _____________ *Flächen-einheiten übrig.*

Das bedeutet: Mit Wahrscheinlichkeit __________ *liegt der Wert der Testgröße auf dem mittleren Bereich der x-Achse und mit Wahrscheinlichkeit* __________________ *auf einen der beiden Randbereichen; immer unter der Annahme, dass die Nullhypothese zutrifft.*

Der Ablehnungsbereich ist nun gerade so konzipiert, dass H_0 verworfen wird, falls der Wert der Testgröße in einen der beiden Randbereiche fällt. Und das bedeutet: Unter der Annahme, ____________________________ *, wird H_0 höchstens mit einer Wahrscheinlichkeit α abgelehnt, was wiederum bedeutet: Das Signifikanzniveau α des Tests wird eingehalten.*

Und genau auf diesem Prinzip basieren alle Ablehnungsbereiche eines Statistischen Tests! Sie sind stets so konzipiert, dass das vorgegebene Signifikanzniveau α eingehalten wird.

Unser Beispiel konkret mit Zahlen:

$n = 25$ *Schüler erreichen nach der neuen Trainingsmethode ein mittleres Sprintergebnis von* $\bar{x} = 16.3$ *(in Sekunden). Das mittlere Sprintergebnis unter der alten Trainingsmethode sei* $\mu_0 = 15.0$. *Als Stichprobenvarianz berechne sich* $\bar{S}^2 = 64$. *Damit erhalten wir folgenden Wert der Testgröße:*

Aus Tabelle D erhalten wir für ein Signifikanzniveau von $\alpha = 0.05$*:*

Wegen ______________________________________ *fällt der Wert der Testgröße* ____________ *in den Ablehnungsbereich, und somit kann die Nullhypothese auf diesem Signifikanzniveau* ____________ *verworfen werden. D.h. der Test zeigt keine Unterschiede in den mittleren Testergebnissen der beiden Trainingsmethoden.*

Das liegt daran, dass das mittlere Testergebnis $\bar{x}$ der neuen Trainingsmethode zu nahe am mittleren Testergebnis μ_0 der alten Trainingsmethode liegt, d.h. die Differenz $\bar{x} - \mu_0$ ist zu klein und daher fällt $T(x_1, \ldots, x_n)$ nicht in die schraffierten (Ablehnungs-)Randbereiche der Dichtefunktion, sondern bleibt im mittleren Bereich. Das Testergebnis unterstützt damit nicht die Ablehnung von _____ , d.h. die beiden Trainingsmethoden erbringen keine wesentlich abweichenden mittleren Testergebnisse, die eine Verwerfung von _____ rechtfertigen würden.

Wäre das mittlere Testergebnis μ_0 der alten Trainingsmethode gleich 12.8*, so erhielten wir als Wert der Testgröße:*

Wegen ___________________________________ ist in diesem Fall die Abweichung von $\bar{x}$ und μ_0 groß genug, so dass H_0 (auf demselben Signifikanzniveau) verworfen werden kann. D.h. in diesem Fall weist der Test Unterschiede in den mittleren Leistungstestergebnissen der beiden Trainingsmethoden nach.

Es gibt noch eine andere Möglichkeit, wie alte und neue Trainingsmethode aus dem letzten Beispiel miteinander verglichen werden können:
Eine Gruppe von _____ Schülern wird ein Schuljahr lang nach der alten Trainingsmethode unterrichtet und parallel dazu eine Gruppe von _____ Schülern im selben Schuljahr nach der neuen Trainingsmethode. Zum Schuljahresabschluss wird in beiden Gruppen der Sprint durchgeführt und die Ergebnisse miteinander verglichen.

Das heißt: Hier werden zwei voneinander unabhängige Stichproben gezogen und nicht nur eine wie in unseren vorhergehenden Betrachtungen. Diese neue Stuation benötigt folgenden neuen Statistischen Test: Den t-Test für zwei unabhängige Stichproben.

2.) T-Test (für zwei unabhängige Stichproben)

Gegeben seien zwei voneinander unabhängige Stichproben $X_1, \dots X_{n_1}$ i.i.d. __________ - verteilter Zufallsvariablen und $Y_1, \dots Y_{n_2}$ i.i.d. ____________________ - verteilter Zufallsvariablen mit unbekannten Erwartungswerten μ_1, μ_2 und unbekannter Varianz σ^2. (Man beachte: σ^2 ist für beide Stichproben derselbe Wert!)

Fall 1: (zweiseitiger Test)	Fall 2: (einseitiger Test)	Fall 3: (einseitiger Test)
$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 \leq \mu_2$	$H_0: \mu_1 \geq \mu_2$
$H_1: \mu_1 \neq \mu_2$	$H_1: \mu_1 > \mu_2$	$H_1: \mu_1 < \mu_2$

Testgröße: (für alle drei Fälle)

mit

Stichprobenvarianz der ersten bzw. zweiten Stichprobe sowie _____ Stichprobenmittelwert der ersten und _____ Stichprobenmittelwert der zweiten Stichprobe.

Verteilung der Testgröße:

Falls wir annehmen, dass ____________ gilt, dann ist ______________________________

_____________________________________ .

Ablenhungsbereiche:

Ablehnung von H_0, falls

Fall 1: ____

Fall 2: ____

Fall 3: ____

wobei α das (vorher festgelegte) Signifikanzniveau des Tests ist. (Die Quantile der $t_{n_1+n_2-2}$–Verteilung finden sich in Tabelle D.)

$x_1, \ldots, x_n$ sind die Realisierungen der ersten Stichprobe und

$y_1, \ldots, y_n$ sind die Realisierungen der zweiten Stichprobe.

Beispiel 47 *(Fortsetzung vom letzten Beispiel)*

Wir betrachten die Zwei-Stichproben-Situation: _____ Schüler werden mit der alten Trainingsmethode unterrichtet und _____ Schüler mit der neuen. Dann wird in beiden Gruppen ein Leistungstest durchgeführt. Wir gehen davon aus, dass sich die Testergebnisse der Schüler aus der ersten Gruppe durch i.i.d. $N(\mu_1, \sigma^2)$-verteilte Zufallsvariablen $X_1, \ldots, X_n$ beschreiben lassen und die Testergebnisse der Schüler aus der zweiten Gruppe durch $N(\mu_2, \sigma^2)$-verteilte Zufallsvariablen $Y_1, \ldots Y_n$.

Wir betrachten die zweiseitige Testsituation mit folgender Nullhypothese:

____________________ *, d.h. die mittleren Leistungstestergebnisse sind gleich; und mit folgender Alternativhypothese:* ____________________ *, d.h. die mittleren Leistungstestergebnisse sind unterschiedlich.*

Seien konkret in der ersten Gruppe $n_1 = 20$ *Schüler und in der zweiten Gruppe* $n_2 = 22$ *Schüler. In der ersten Gruppe sei* $\bar{x} = 15.7$ *und in der zweiten Gruppe sei* $\bar{y} = 16.9$ *mit Stichprobenvarianzen* $\bar{s}_x^2 = 9.3$ *und* $\bar{s}_y^2 = 8.1$*. Dann errechnet sich für die Testgröße folgender Wert:*

Aus Tabelle D lesen wir ab für $\alpha = 0.05$*:*

Es gilt nun: ______________________________ *, d.h. der Wert der Testgröße fällt* __________ *in den Ablehnungsbereich, und somit kann die Nullhypothese auf diesem Signifikanzniveau* __________ *verworfen werden, was bedeutet, dass kein Unterschied in den mittleren Leistungstestergebnissen der beiden Trainingsmethoden nachgewiesen werden kann.*

Dies liegt daran, dass die mittleren Leistungstestergebnisse _____ *und* _____ *zu dicht beieinander liegen, d.h. die Differenz* ____________ *ist zu klein, liegt nahe bei Null und kommt daher nicht in die (schraffierten) Randbereiche der zugehörigen Dichtefunktion, wie wir das im letzten Beispiel ausführlich beschrieben haben. Da das Testergebnis keinen Anlass dazu gibt, H_0 zu verwerfen, wird sie beibehalten.*

Wäre das mittlere Testergebnis in der zweiten Gruppe schlechter, etwa $\bar{y} = 18.1$, dann hätten wir folgenden Wert für die Testgröße:

Wegen ____ *wird nun* _____ *verworfen, da der errechnete Wert der Testgröße in den Ablehnungsbereich fällt; d.h. die mittleren Testergebnisse weichen weit genug voneinander ab, so dass* _____ *nicht mehr aufrechterhalten werden kann, und somit kann ein Unterschied in den mittleren Leistungstestergebnissen beider Trainingsmethoden nachgewiesen werden.*

Sowohl beim Ein-Stichproben-t-Test als auch beim Zwei-Stichproben-t-Test sind wir von normalverteilten Zufallsvriablen mit ______________________________ Varianz σ^2 ausgegangen. Falls uns nun die Varianz σ^2 ______________________ ist, dann können wir diese Zusatzinformation nutzen, um statt dem t-Test einen anderen Statistischen Test, den Gauß-Test, zu verwenden, der eine bessere ____________________________ aufweist als der t-Test.

3.) <u>Gauß-Test</u> (für eine Stichprobe)

Gegeben sei eine Stichprobe $X_1, \ldots X_n$ i.i.d. ______________________ - verteilter Zufallsvariablen mit unbekannten Erwartungswert μ und bekannter Varianz σ_0^2.

Fall 1: (zweiseitiger Test)	Fall 2: (einseitiger Test)	Fall 3: (einseitiger Test)
$H_0: \ \mu = \mu_0$	$H_0: \ \mu \leq \mu_0$	$H_0: \ \mu \geq \mu_0$
$H_1: \ \mu \neq \mu_0$	$H_1: \ \mu > \mu_0$	$H_1: \ \mu < \mu_0$

Testgröße: (für alle drei Fälle)

wobei $\bar{X}$ der Stichprobenmittelwert ist.

Verteilung der Testgröße:

Falls wir annehmen, dass ______________ gilt, dann ist ____________________________

_________________________________ .

Ablenhungsbereiche:

Ablehnung von H_0, falls

Fall 1: ______________________________

Fall 2: ______________________________

Fall 3: ______________________________ ,

wobei α das (vorher festgelegte) Signifikanzniveau des Tests ist. Die Quantile der Standardnormal-Verteilung (N(0,1)-Verteilung) finden sich in der z-Tabelle.

Beispiel 48 *(Fortsetzung vom letzten Beispiel)*

Wir betrachten die Ein-Stichproben-Situation. Eine Gruppe von $n = 25$ Schülern werde mit der neuen Trainingsmethode gelehrt und es wird ein mittleres Leistungstestergebnis von $\bar{x} = 16.5$ erreicht.

Ein Erfahrungswert des mittleren Leistungstestergebnisses unter der alten Trainingsmethode sei $\mu_0 = 15.2$. Ferner sei die Varianz $\sigma_0^2 = 8.7$ der Leistungstestergebnisse bekannt.

Wir wollen zweiseitig testen, d.h. die Nullhypothese ist ______________ , und die Alternativhypothese ist ______________ . Als Wert der Testgröße ergibt sich:

Für $\alpha = 0.05$ *lesen wir aus der z-Tabelle das* __________ *-Quantil der* $N(0,1)$*-Verteilung ab:*

Da nun wegen

der Wert der Testgröße in den Ablehnungsbereich fällt, wird die Nullhypothese auf diesem Signifikanzniveau verworfen. Das heißt: Die Nullhypothese, dass sich die mittleren Leistungstestergebnisse von alter und neuer Trainingsmethode nicht unterscheiden, wird abgelehnt.

Falls wir als Signifikanzniveau $\alpha = 0.02$ *wählen, erhalten wir folgendes Quantil aus der z-Tabelle:*

In diesem Fall kann wegen

die Nullhypothese nicht verworfen werden, d.h. auf diesem strengeren Signifikanzniveau fällt der Wert der Testgröße nicht in den Ablehnungsbereich, was wiederum bedeutet: Die Stichprobenergebnisse rechtfertigen auf diesem Signifikanzniveau keine Ablehnung der Nullhypothese.

Wir kommen nun zum Gauß-Test für den Zwei-Stichproben-Fall. Auch hier ist zum t-Test wieder der Unterschied, dass uns diesmal die Varianzen (beide) bekannt sind. (Und diesmal dürfen beide Varianzen auch unterschiedlich sein, im Gegensatz zum 2-Stichproben-t-Test!)

4.) <u>Gauß-Test</u> (für zwei unabhängige Stichproben)

Gegeben seien zwei voneinander unabhängige Stichproben $X_1, \ldots, X_{n_1}$ i.i.d. ________- verteilter Zufallsvariablen und $Y_1, \ldots, Y_{n_2}$ i.i.d. ________________ - verteilter Zufallsvariablen mit unbekannten Erwartungswerten μ_1, μ_2 und bekannten Varianzen σ_1^2 und σ_2^2.

<u>Fall 1:</u> (zweiseitiger Test)	<u>Fall 2:</u> (einseitiger Test)	<u>Fall 3:</u> (einseitiger Test)
$H_0 : \mu_1 = \mu_2$	$H_0 : \mu_1 \leq \mu_2$	$H_0 : \mu_1 \geq \mu_2$
$H_1 : \mu_1 \neq \mu_2$	$H_1 : \mu_1 > \mu_2$	$H_1 : \mu_1 < \mu_2$

<u>Testgröße:</u> (für alle drei Fälle)

mit $\bar{X}$ Stichprobenmittelwert der ersten und $\bar{Y}$ Stichprobenmittelwert der zweiten Stichprobe.

<u>Verteilung der Testgröße:</u>

Falls wir annehmen, dass __________ gilt, dann ist ________________________

______________________________ .

Ablenhungsbereiche:

Ablehnung von H_0, falls

Fall 1: ______________________________

Fall 2: ______________________________

Fall 3: ______________________________

wobei α das (vorher festgelegte) Signifikanzniveau des Tests ist. Die Quantile der N(0,1)-Verteilung finden sich in der z-Tabelle. $x_1, \ldots, x_n$ sind die Realisierungen der ersten Stichprobe und $y_1, \ldots, y_n$ sind die Realisierungen der zweiten Stichprobe.

Beispiel 49 *(Fortsetzung vom letzten Beispiel)*

Wir betrachten die Zwei-Stichproben-Situation: $n_1 = 20$, $n_2 = 22$ *mit* $\bar{x} = 15.7$, $\bar{y} = 16.9$ *und es seien zusätzlich bekannt:* $\sigma_1^2 = 9.0$ *und* $\sigma_2^2 = 8.0$. *Wir wollen zweiseitig testen (d.h.* ______________ , ______________ *) und berechnen dazu folgenden Wert der Testgröße:*

Aus der z-Tabelle finden wir für $\alpha = 0.05$*:*

Wegen ____________________ *kann auf diesem Signifikanzniveau die Nullhypothese* __________ *verworfen werden, weil der Wert der Testgröße* __________ *in den Ablehnungsbereich fällt.*

Falls das mittlere zweite Leistungstestergebnis $\bar{y}$ noch schlechter ausfallen würde, etwa $\bar{y} = 17.6$, dann berechnet sich folgender Wert der Testgröße:

Wegen ________________________ *kann hier (auf demselben Signifikanzniveau) die Nullhypothese verworfen werden.*

In den bisher behandelten Zwei-Stichproben-Fällen (sowohl beim t-Test als auch beim Gauß-Test) werden stets zwei voneinander ______________________________ Stichproben vorausgesetzt. In den zugehörigen Beispielen wurden dafür alte und neue Trainingsmethode an zwei verschiedenen Schülergruppen getestet.

Wir wollen jetzt noch den Zwei-Stichproben-t-Test mit zwei voneinander ________ ______________________ (man sagt auch: ______________________________) Stichproben betrachten. Abhängige Stichproben liegen immer dann vor, wenn zwei Merkmale an ein und derselben Gruppe von Untersuchungseinheiten erhoben werden, wenn z.B. ein und dieselbe Schülergruppe erst mit der alten und dann mit der neuen Trainingsmethode gelehrt wird und beidesmal jeweils im Anschluss daran der Leistungstest durchgeführt wird.

Im Gegensatz zum Ein-Stichproben-t-Test liegt uns also hier nicht ein Erfahrungswert _____ des mittleren Leistungstestergebnisses der alten Trainingsmethode aus der Vergangenheit vor, sondern ein konkret erhobenes mittleres Leistungstestergebnis _____ der alten Trainingsmethode.

Und dafür gibt es folgenden Statistischen Test:

5.) <u>T-Test</u> (für zwei abhängige Stichproben)

Gegeben seien zwei voneinander abhängige Stichproben $X_1, \ldots, X_n$ i.i.d. $N(\mu_1, \sigma_1^2)$-verteilter Zufallsvariablen und $Y_1, \ldots, Y_n$ i.i.d. $N(\mu_2, \sigma_2^2)$-verteilter Zufallsvariablen mit unbekannten Erwartungswerten μ_1, μ_2 und unbekannten Varianzen σ_1^2 und σ_2^2. Ferner seien die Differenzen ____________________ für $i = 1, \ldots, n$) alles normalverteilte Zufallsvariablen.

(Merkhilfe: $X_1, \ldots, X_n$ sind die Stichprobenergebnisse "vorher" und $Y_1, \ldots Y_n$ sind die Stichprobenergebnisse "nachher".)

Fall 1: (zweiseitiger Test)	Fall 2: (einseitiger Test)	Fall 3: (einseitiger Test)
$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 \leq \mu_2$	$H_0: \mu_1 \geq \mu_2$
$H_1: \mu_1 \neq \mu_2$	$H_1: \mu_1 > \mu_2$	$H_1: \mu_1 < \mu_2$

<u>Testgröße:</u> (für alle drei Fälle)

wobei $\bar{X}$ und $\bar{Y}$ die beiden Stichprobenmittelwerte sind und

Verteilung der Testgröße:

Falls wir annehmen, dass $\mu_1 = \mu_2$, dann ist $T(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$ $t_{(n-1)}$-verteilt.

Ablenhungsbereiche: Ablehnung von H_0, falls

Fall 1: $|T(X_1, \ldots, X_n, Y_1, \ldots, Y_n)|$

Fall 2: $T(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$

Fall 3: $T(X_1, \ldots, X_n, Y_1, \ldots, Y_n)$

wobei α das (vorher festgelegte) Signifikanzniveau des Tests ist.

Beispiel 50 *(Fortsetzung von den letzten Beispielen)*

Wir betrachten die Zwei-Stichproben-Situation mit zwei abhängigen Stichproben: $n = 24$ *Schüler werden zuerst mit der alten und dann mit der neuen Trainingsmethode gelehrt. Das mittlere Leistungstestergebnis nach der alten Methode sei* $\bar{x} = 14.1$ *und nach der neuen Methode sei* $\bar{y} = 17.2$*. Ferner errechne sich aus den Realisierungen beider Stichproben der Wert* $\bar{D}^2 = 49$*. Damit ergibt sich für die Testgröße:*

Auf einem Signifikanzniveau von $\alpha = 0.05$ *lesen wir in Tabelle D ab:*

Wegen ________________________________ *kann hier die Nullhypothese verworfen werden, d.h. die Stichprobenergebnisse rechtfertigen die Alternativhypothese, dass es einen Unterschied zwischen den beiden Trainingsmethoden gibt.*