

Statistische Schätz- und Testverfahren

Mike Hüftle

31. Juli 2006

Inhaltsverzeichnis

1	Einleitung	2
1.1	Statistische Schätzungen	2
1.2	Statistische Tests	3
2	Statistische Schätzungen	4
2.1	Eigenschaften einer Schätzung	4
2.2	Punktschätzung	5
2.2.1	Nebenpfad: Methode der kleinsten Quadrate	5
2.3	Maximum-Likelihood-Methode	6
2.3.1	Nebenpfad: Maximum-Likelihood-Schätzer	6
2.4	Konfidenzschätzung	8
2.4.1	Nebenpfad: Konfidenzschätzung bei Normalverteilungsan- nahme	9
2.4.2	Nebenpfad: Bestimmung des Stichprobenumfanges	9
2.5	Toleranzschätzungen	11
3	Statistische Tests	13
3.1	Einleitung	13
3.2	Tests bei normalverteilter Grundgesamtheit	14
3.3	Tests bei normalverteilter Grundgesamtheit	15
3.4	Nichtparametrische Tests	16
3.5	Vorzeichentest	17
4	Anpassungstests	18
4.1	18
4.2	Chi-Quadrat-Anpassungstest	19
4.2.1	Nebenpfad: Bestimmung der beobachteten und theoretischen Verteilung	20
5	Ausreißer-Tests	21
5.1	21
5.2	Faustregeln	22
5.3	Ausreißer-Tests nach Dixon und nach Grubbs	23

5.3.1	Nebenpfad: Tabellierte Werte für den Test nach Dixon . . .	24
5.4	Ausreißer-Test nach Walsh	25
5.5	Weitere Ausreißer-Tests	26
6	Variablenauswahl	27
6.1	Problemstellung	27
6.2	Lösungsansätze	28
6.2.1	Nebenpfad: Mallows Cp-Statistik	28
7	Literatur	30
7.1	Literatur zur Zeitreihenanalyse	30

1 Einleitung

1.1 Statistische Schätzungen

Aussagen über die Verteilungsfunktion Statistische Schätzmethoden dienen dazu, Aussagen über die unbekannte Verteilungsfunktion der Grundgesamtheit oder deren Parameter anhand von Stichproben aus dieser Grundgesamtheit machen zu können.

Punktschätzung Diese Aussagen können in Form eines **Schätzwertes für den oder die unbekannten Parameter** in der Grundgesamtheit gemacht werden (Punktschätzung).

Konfidenzschätzung Um eine Aussage über die Genauigkeit und Sicherheit eines solchen Schätzwertes machen zu können wird ein **Schätzintervall** berechnet (Konfidenzschätzung).

Toleranzschätzung Toleranzschätzungen geben Intervalle für den oder die Parameter der Verteilungsfunktion an, so dass die Intervalle einen vorgegebenen Anteil der Grundgesamtheit **mindestens beinhalten**.

Geschätzte Funktionen und Parameter werden üblicherweise durch ein **Dach gekennzeichnet**, beispielsweise $\hat{\Theta}$ oder $\hat{\alpha}$.

1.2 Statistische Tests

Verteilungsannahmen. Statistische Tests dienen dazu, anhand von Stichproben **Annahmen (Hypothesen) über Verteilungen** in der Grundgesamtheit der Daten **zu überprüfen**.

Insbesondere kann überprüft werden, ob bestimmte Sachverhalte in den Stichprobendaten zufällig sind oder nicht.

2 Statistische Schätzungen

2.1 Eigenschaften einer Schätzung

- Eigenschaften einer guten Schätzung
- Eine gute Schätzung $\hat{\Theta}$ soll die folgenden **vier Eigenschaften** aufweisen:
- Die Schätzung soll **unverzerrt** sein, d.h. der Erwartungswert von $\hat{\Theta}$ soll gleich dem zu schätzenden Parameter sein.
 - Die Schätzung soll **konsistent** sein, d.h. $\hat{\Theta}$ soll mit wachsendem Stichprobenumfang gegen den wahren Parameter Θ konvergieren.
 - Eine effiziente Schätzung schätzt mit einer **geringen Varianz**. Gilt für zwei erwartungstreue und konsistente Schätzungen $\hat{\Theta}_1$ und $\hat{\Theta}_2$ desselben Parameters Θ die Beziehung $Var(\hat{\Theta}_1) < Var(\hat{\Theta}_2)$, so heißt $\hat{\Theta}_1$ relativ effizienter als $\hat{\Theta}_2$. Die effizientere Schätzung wird bevorzugt.
 - Eine Schätzung soll **erschöpfend** oder **suffizient** sein. Das bedeutet, dass durch sie alle in der Stichprobe enthaltenen Informationen über den unbekannt Parameter Θ ausgenutzt werden und keine andere Schätzung zusätzliche Erkenntnisse über Θ bringt.

Beispiel Beispielsweise ist die relative Häufigkeit eines Ereignisses A aus einer dichotomen Stichprobe eine erwartungstreue, konsistente und effektive Schätzung für die Eintrittswahrscheinlichkeit p des Ereignisses A . Die Wahrscheinlichkeit p ist hier der unbekannt Parameter Θ .

2.2 Punktschätzung

Problem der Punktschätzung Das Problem der Punktschätzung besteht darin, für die unbekannte Verteilungsfunktion der Grundgesamtheit oder für ihre Parameter Θ anhand von Stichproben Schätzungen zu bestimmen und eventuell aus mehreren Schätzungen die geeignetste auszuwählen.

Schätzfunktion Eine Schätzfunktion, Punktschätzung oder auch kurz Schätzung genannt ist eine **Zufallsgröße** und besitzt somit eine **Verteilung**, die von der Verteilungsfunktion der Grundgesamtheit abhängt und aus dieser hergeleitet werden kann.

Methode der kleinsten Quadrate Die auf C.F. Gauss zurückgehende **II** Methode der kleinsten Quadrate ist der wohl am häufigsten angewendete Punktschätzer und kann im Gegensatz zur Maximum-Likelihood-Methode auch dann angewendet werden, wenn der **Verteilungstyp der Grundgesamtheit nicht bekannt** ist.

2.2.1 Nebenpfad: Methode der kleinsten Quadrate

Minimierung der quadratischen Abweichungen Die Methode der kleinsten Quadrate **minimiert die quadratischen Abweichungen** zwischen den Stichprobenwerten x_i und dem unbekanntem Parameter Θ . Hierzu muss für Θ die Forderung

$$\sum_{i=1}^n (x_i - \hat{\Theta})^2 = \min \sum_{i=1}^n (x_i - \Theta)^2 \quad (1)$$

gelten. Die Abweichung bzw. der Abstand zwischen der Schätzung $\hat{\Theta}$ und den gemessenen Werten x_i kann als eine Schätzung für die in der Realität auftretende **Störgröße** interpretiert werden, welche durch die Schätzung minimiert werden soll.

2.3 Maximum-Likelihood-Methode

Methodenbeschreibung Die von R.A. Fisher entwickelte Maximum-Likelihood-Schätzmethode bestimmt den geschätzten Parameter, welcher die **Wahrscheinlichkeit (likelihood) des Auftretens** der in einer Stichprobe beobachteten Messungen maximiert. Eine detailliertere Methodenbeschreibung finden Sie [\[1\]](#) hier.

Eigenschaften der ML-Schätzung Existiert für einen Parameter ein **erschöpfender Schätzwert**, dann entspricht dieser dem nach der Maximum-Likelihood-Methode bestimmten Parameter. Hieraus folgt, dass erschöpfende statistische Kennwerte, wie beispielsweise das arithmetische Mittel oder die Varianz, gleichzeitig Maximum-Likelihood-Schätzungen der Parameter μ und σ^2 sind. Jedoch sind die Maximum-Likelihood-Schätzungen nicht gleichzeitig **erwartungstreue Schätzungen**.

Anwendung der ML-Schätzung Die Maximum-Likelihood-Methode führt in den meisten praktischen Fällen zum Ziel. Sie setzt jedoch voraus, dass der **Verteilungstyp** F der Zufallsgröße X in der Grundgesamtheit **bekannt** ist. Bei bekanntem Verteilungstyp liefern die Methode der kleinsten Quadrate und die Maximum-Likelihood-Methode **asymptotisch dieselben Schätzungen**.

Weitere Punktschätzer Weitere Schätzmethoden sind die **Minimum-Chi-Quadrat-Methode** und die **Momentenmethode**.

2.3.1 Nebenpfad: Maximum-Likelihood-Schätzer

Methodenbeschreibung Ausgangspunkt dieser Schätzmethode ist eine Stichprobe vom Umfang n aus einer nach F verteilten Grundgesamtheit. F hänge von einem unbekanntem Parameter Θ ab.

Um die maximale Auftretenswahrscheinlichkeit der Messungen in einer Stichprobe in Abhängigkeit vom unbekanntem Parameter Θ zu ermitteln wird eine Wahrscheinlichkeitsfunktion, die **Likelihoodfunktion**, definiert, welche nach Θ differenziert wird. Ist die Zufallsgröße X diskret verteilt, so ergibt sich die Likelihoodfunktion zu

$$L(x_1, \dots, x_n; \Theta) = P(X = x_1) \cdot \dots \cdot P(X = x_n) = \prod_{i=1}^n P(X = x_i) \quad (2)$$

und bei stetiger Verteilung zu

$$L(x_1, \dots, x_n; \Theta) = \prod_{i=1}^n f_i(x) \quad (3)$$

Wird die Ableitung dieser Gleichung Null gesetzt, so ergibt sich die **Bestimmungsgleichung für den unbekannt Parameter Θ** zu

$$\frac{d \ln L}{d \Theta} = 0 \quad (4)$$

2.4 Konfidenzschätzung

Parameterintervalle Da Punktschätzungen **zufallsabhängig** sind und damit von Stichprobe zu Stichprobe schwanken können, ist es oft sinnvoll Intervalle zu berechnen, in denen die unbekannt Parameter mit großer Wahrscheinlichkeit liegen.

Konfidenzgrenzen Konfidenzschätzungen (Intervallschätzungen, Bereichsschätzungen) berechnen aus der Stichprobe Intervalle, in denen der **unbekannte Parameter der Verteilung** der Grundgesamtheit mit einer Wahrscheinlichkeit ϵ zu erwarten ist. Solche Intervalle bezeichnet man als Konfidenz- oder Vertrauensintervalle $J = (G_u, G_o)$ mit der unteren bzw. oberen Konfidenzgrenze G_u bzw. G_o . g_u und g_o sind die tatsächlichen Realisierungen der Konfidenzgrenzen.

Konfidenzkoeffizient wird Konfidenzkoeffizient, Konfidenzniveau oder Überdeckungswahrscheinlichkeit und Signifikanzniveau $\alpha = 1 - \epsilon$ bezeichnet die **Irrtumswahrscheinlichkeit** oder das **Signifikanzniveau**. Da die Konfidenzgrenzen von der Stichprobe abhängen und somit Zufallsgrößen sind, ist auch das Konfidenzintervall J ein zufälliges Intervall.

Genauigkeit der Schätzung Die Länge des Konfidenzintervalls $L = g_o - g_u$ ist ein Maß für die Genauigkeit der Schätzung. Mit **wachsendem Stichprobenumfang** wird das Konfidenzintervall kleiner, die Schätzung also genauer. Demgegenüber wird mit **wachsender Überdeckungswahrscheinlichkeit** ϵ die Schätzung ungenauer.

Anwendung Welche Konfidenzschätzungen für ein bestimmtes Problem anzuwenden sind hängt davon ab, **welche Parameter der Verteilungsfunktion** bekannt sind bzw. ob eine **Annahme über den Verteilungstyp** getroffen werden kann. Beispielhaft werden die **I** Konfidenzschätzungen für die Parameter bei normalverteilter Grundgesamtheit angegeben. Für weitere Konfidenzschätzungen wird auf die angegebene Literatur verwiesen. Nähere Erläuterungen zur Anwendung von Konfidenzschätzungen bei der Bestimmung des Stichprobenumfanges finden Sie **I** hier.

2.4.1 Nebenpfad: Konfidenzschätzung bei Normalverteilungsannahme

Konfidenzschätzung des Verteilungsgesetz der Grundgesamtheit die **Normalverteilung mit unbekanntem Parameter** μ und (aus Erfahrungswerten) **bekanntem** σ^2 , so kann das Konfidenzintervall für den Mittelwert μ wie folgt berechnet werden: Der Mittelwert wird als arithmetisches Mittel \bar{X} geschätzt. Das Konfidenzintervall für μ lautet:

$$\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad (5)$$

mit dem arithmetischen Mittel \bar{X} und der Stichprobengröße n .

Für **vorgegebene Werte** von α liegen die Quantile der standardisierten Normalverteilung $z_{1-\frac{\alpha}{2}}$ tabelliert vor.

Konfidenzschätzung der Schätzung des **Konfidenzintervalls für den Mittelwert** μ der normalverteilten Grundgesamtheit bei **unbekanntem** σ^2 wird zunächst die Standardabweichung geschätzt und dann das Konfidenzintervall bestimmt zu μ /unbekanntem σ^2

$$\bar{X} - t_{m; 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{m; 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \quad (6)$$

wobei S die empirische Standardabweichung als Schätzung für σ^2 ist und $t_{m; 1-\frac{\alpha}{2}}$ das Quantil der t -Verteilung mit m Freiheitsgraden

2.4.2 Nebenpfad: Bestimmung des Stichprobenumfanges

Stichprobenumfang Konfidenzintervalle werden auch benutzt, um den **notwendige Stichprobenumfang** zur Schätzung eines Parameters zu bestimmen. Gibt der Benutzer ein bestimmtes Konfidenzniveau vor, welches die Genauigkeit der Schätzung definiert, so kann hieraus der notwendige Stichprobenumfang zur Erreichung dieser Genauigkeit berechnet werden.

Berechnung des Stichprobenumfanges Generell gilt, dass mit kleiner werdendem Konfidenzintervall bei konstantem Konfidenzniveau der **benötigte Stichprobenumfang quadratisch wächst**. Eine Formel zur Bestimmung des Stichprobenumfanges in Abhängigkeit von der Länge L des Konfidenzintervalles und bekanntem bzw. geschätztem Erwartungswert μ gibt BORTZ (1985) an:

$$n = \frac{4 \cdot z_{1-\alpha/2}^2 \cdot \sigma^2}{L^2} \quad (7)$$

Anwendung Schwierigkeiten bei der Ermittlung der notwendigen Stichprobenumfänge bereitet der Standardfehler, der im allgemeinen unbekannt ist, da die Parameter μ und σ^2 nicht vorliegen. Deshalb sollten die unbekannt Parameter durch Voruntersuchungen abgeschätzt werden. Danach wird eine erste Schätzung des Stichprobenumfanges vorgenommen, die eventuell mit neuen Daten verbessert wird, bis sich die Schätzungen der unbekannt Parameter stabilisiert haben.

2.5 Toleranzschätzungen

Statistische Toleranzschätzungen sind Parameterschätzungen unter der Annahme, dass ein bestimmter **Mindestanteil der Grundgesamtheit** mit vorgegebener Wahrscheinlichkeit innerhalb eines Intervalls liegt. Diese Intervalle, die anhand von Stichproben berechnet werden, heißen **statistische Toleranzintervalle** und ihre Grenzen **statistische Toleranzgrenzen**.

Toleranzgrenzen Die untere bzw. obere statistischen Toleranzgrenzen T_u bzw. T_o sind **Zufallszahlen** und werden anhand einer Stichprobe vom Umfang n so bestimmt, dass die Wahrscheinlichkeit dafür, dass zwischen den Grenzen mindestens der Anteil γ der Grundgesamtheit liegt, gleich β sein soll. Die Wahrscheinlichkeit β heißt **Sicherheitsniveau**, statistische Sicherheit oder Vertrauensniveau. γ ist das **Überdeckungs niveau**. Bei gegebenem β und γ werden für eine Stichprobe n die Realisierung τ_u und τ_o als das **zweiseitige statistische Toleranzintervall** (τ_u, τ_o) bestimmt.

Toleranzfaktoren Die Toleranzgrenzen können folgendermaßen berechnet werden:

$$\tau_u = \bar{x} - k_{n;\beta,\gamma} \cdot s \quad (8)$$

und

$$\tau_o = \bar{x} + k_{n;\beta,\gamma} \cdot s \quad (9)$$

wobei \bar{x} und s arithmetisches Mittel und Standardabweichung der Stichprobe sind.

Die **Toleranzfaktoren** $k_{n;\beta,\gamma}$ hängen vom Stichprobenumfang, vom Sicherheitsniveau β und vom Überdeckungs niveau γ ab und liegen in Tabellen vor. Ebenfalls liegen Toleranzfaktoren für die Fälle σ^2 bekannt und μ unbekannt bzw. μ bekannt und σ^2 unbekannt sowie für einseitige Toleranzintervalle vor.

Verteilungsfreie statistische Toleranzgrenzen Werden **keine Annahmen über den Verteilungstyp** in der Grundgesamtheit gemacht, so spricht man von verteilungsfreien statistischen Toleranzgrenzen. Es wird lediglich die Stetigkeit der Verteilungsfunktion vorausgesetzt.

Mindeststichprobenumfang Bei einer verteilungsfreien Toleranzschätzung wird der **minimale Stichprobenumfang** n_0 gesucht, bei dem mit der Wahrscheinlichkeit β mindestens der Anteil γ der Grundgesamtheit zwischen dem kleinsten (x_{min}) und dem größten

(x_{max}) Wert der Stichprobe liegt. x_{min} und x_{max} werden als verteilungsfreie oder **nichtparametrische statistische Toleranzgrenzen** bezeichnet.

Der Mindeststichprobenumfang n_0 ergibt sich bei zweiseitigem verteilungsfreiem statistischem Toleranzintervall aus der Beziehung

$$n \cdot \gamma^{n_0-1} - (n-1) \cdot \gamma^n \leq 1 - \beta \quad (10)$$

und bei einseitigem verteilungsfreiem statistischem Toleranzintervall aus

$$\gamma^{n_0} \leq 1 - \beta \quad (11)$$

Für die einzelnen Werte von n existieren Tafeln, aus denen man den **Mindeststichprobenumfang** ablesen kann.

3 Statistische Tests

3.1 Einleitung

Hypothesen über die Verteilung in der Grundgesamtheit Statistische Tests oder Prüfverfahren dienen dazu, an Hand von Stichproben Hypothesen über das Verteilungsgesetz in der Grundgesamtheit zu überprüfen. Oft existiert über die unbekannte Verteilungsfunktion oder ihre unbekannt Parameter (wie z.B. μ , σ^2) eine bestimmte Vorstellung, die als **statistische Hypothese** H formuliert wird. Wird neben H noch eine weitere Hypothese betrachtet, so nennt man H_0 **Nullhypothese** und H_1 die **Alternativhypothese**.

Überprüfung der Hypothesen Die Überprüfung der Hypothesen erfolgt mittels statistischer Tests. Ein Test entscheidet darüber, ob die Daten einer konkreten Stichprobe zur aufgestellten Hypothese H_0 im Widerspruch stehen oder nicht, d.h. ob H_0 **abzulehnen ist oder nicht**.

Grundsätzlich wird zwischen parametrischen und nicht-parametrischen Tests unterschieden.

Parametrische Tests Parametrische Tests gehen von der Annahme aus, dass die Daten der **Grundgesamtheit ein bestimmtes Skalenniveau und eine bestimmte Verteilung** aufweisen (häufig wird die Normalverteilung angenommen), deren Gestalt bis auf einen oder mehrere unbekannte Parameter bekannt ist. Werden diese Annahmen verletzt, so hat dies Einschränkungen bei der Güte der Testergebnisse zur Folge.

Nicht-parametrische Tests Nicht-parametrische (parameterfreie, verteilungsfreie) Tests hingegen **setzen keine bestimmte Verteilung in der Grundgesamtheit voraus**. Sie haben jedoch auch eine geringere Güte als parametrische Tests. Daher ist ein vorhandener Unterschied bei Anwendung eines nicht-parametrischen Tests **weniger oft signifikant** als bei Anwendung eines parametrischen Tests. Wird ein Unterschied mit einem nicht-parametrischen Test als signifikant bestimmt, so ist dieser auch bei Verwendung eines parametrischen Tests signifikant. Der Umkehrschluss gilt jedoch nicht.

3.2 Tests bei normalverteilter Grundgesamtheit

Existiert eine Hypothese H_0 für den **unbekannten Mittelwert** μ_0 ($H_0 : \mu = \mu_0$) der normalverteilten Grundgesamtheit bei **bekannter Streuung**, so kann dieser anhand einer Stichprobe auf seine Richtigkeit getestet werden. Weicht der, aus der Stichprobe berechnete Mittelwert (z.B. das arithmetische Mittel), vom hypothetischen Wert μ_0 ab, so stellt sich die Frage ob die Hypothese H_0 aufrecht erhalten werden kann oder wie groß die Abweichung höchstens sein darf, damit sie noch als unwesentlich betrachtet werden kann.

Signifikante Abweichung Ist die Wahrscheinlichkeit, dass die Abweichung größer oder gleich einem Wert d ist, kleiner oder gleich einer **Schranke** α , so wird die **Abweichung als signifikant oder statistisch gesichert** bezeichnet. Dies führt zur Ablehnung von H_0 . Oder formal:

$$P(|\bar{X} - \mu_0| \geq d) \leq \alpha \quad (12)$$

. Ist dagegen die Wahrscheinlichkeit für das Ereignis „Abweichung größer oder gleich d “ größer als α , so ist die **Abweichung zufällig oder nicht signifikant**. Demzufolge wird H_0 nicht abgelehnt.

Fehler 1. Art Dies bedeutet jedoch nicht, dass H_0 **angenommen** wird. Es besagt lediglich, dass das Stichprobenergebnis nicht im Widerspruch zur aufgestellten Hypothese steht. Dies bedeutet aber auch, dass eine Aussage darüber, ob eine Hypothese abzulehnen richtig oder falsch ist, nur mit einer bestimmten Wahrscheinlichkeit α getroffen werden kann. Wird die Nullhypothese **fälschlicherweise abgelehnt**, so wird von einem Fehler 1. Art gesprochen.

Signifikanzniveau Dieser Test auf Ablehnung von H_0 wird Signifikanztest genannt und α entsprechend Signifikanzniveau oder auch Irrtumswahrscheinlichkeit. Da α möglichst klein gehalten werden soll, wird in der Praxis mit α -Werten von 0,05, 0,01 oder auch 0,001 gearbeitet.

3.3 Tests bei normalverteilter Grundgesamtheit

Fehler 2. Art Ein Fehler 2. Art wird begangen, wenn die Hypothese H_0 nicht verworfen wird, obwohl sie falsch ist. Dies hängt vom Wert des unbekanntes Parameters in der Grundgesamtheit ab. Der Fehler 2. Art kann bei einem Signifikanztest daher nicht angegeben werden.

Hierzu muss neben der Nullhypothese eine Alternativhypothese H_1 aufgestellt und ein **Alternativtest** durchgeführt werden. Bei einem Alternativtest wird aufgrund einer konkreten Stichprobe entschieden, ob H_0 abgelehnt (und damit H_1 angenommen) wird oder H_0 angenommen (und damit H_1 abgelehnt) wird.

Dabei soll der Test so beschaffen sein, dass der Fehler 2. Art möglichst klein wird. Zur Darstellung dieser Zusammenhänge dient die **Gütefunktion des Tests**, die in Abhängigkeit vom wahren Parameter die Wahrscheinlichkeit für die Ablehnung der Nullhypothese angibt.

Einseitige und zweiseitige Tests Bei einem **einseitigen Test** interessiert nur die Abweichung nach einer Seite, z.B. wenn es nur darauf ankommt, ob ein Stichprobenmittelwert zu groß ist, während ein zu kleiner Mittelwert ohne Bedeutung ist. Im Gegensatz dazu interessieren bei einem **zweiseitigen Test** die Abweichungen nach beiden Seiten.

3.4 Nichtparametrische Tests

Alle bisher dargestellten statistischen Tests setzen voraus, dass die Gestalt der Verteilungsfunktion bis auf einzelne Parameter **bekannt ist**. Die Hypothesen beziehen sich dann auf die unbekannt Parameter.

Nichtparametrische, parameterfreie oder verteilungsfreie Tests prüfen Hypothesen, **ohne die zugrunde liegende Verteilungsfunktion der Grundgesamtheit zu kennen**. Dies ist ein großer Vorteil dieser Methoden, da sie beispielsweise von der Annahme einer normalverteilten Grundgesamtheit unabhängig sind. Allerdings haben die Methoden den Nachteil einer **geringeren Güte** gegenüber den parametrischen Tests, d.h. die statistische Absicherung ist oft geringer. Wird jedoch mit einem nichtparametrischen Test ein Unterschied als signifikant erkannt, so ist dieser auch im parametrischen Test signifikant.

Effizienz eines Tests Die Effizienz oder die Wirksamkeit eines Tests (in Vergleich zu einem anderen) wird ermittelt als **Verhältnis der Stichprobenumfänge** beider Tests die zur Erreichung der gleichen Güte notwendig sind. Dies basiert auf der Tatsache, dass die Steilheit der Gütefunktion mit wachsendem Stichprobenumfang zunimmt. Somit lässt sich für einen nichtparametrischen Test die gleiche Güte erzielen wie für einen parametrischen, wenn nur der Stichprobenumfang genügend groß gewählt wird.

Anwendung Grundsätzlich gilt, dass wenn die Voraussetzungen für einen Parametertest erfüllt sind, ein solcher angewendet werden sollte. Ansonsten ist ein nichtparametrischer Test meist wirksamer.

3.5 Vorzeichentest

Parameterfreier Test für verbundene Stichproben	<p>Der Vorzeichentest zum Vergleich von zwei verbundenen Stichproben ist einer der rechnerisch einfachsten parameterfreien Tests. Er ist auf verbundene oder abhängige Stichproben anwendbar, also Stichproben, bei denen die Werte paarweise einander zugeordnet werden können. Beispielsweise sind dies Messungen mit zwei verschiedenen Messinstrumenten an den gleichen Objekten, so dass für jedes Objekt zwei Messungen vorliegen. Die einzige Voraussetzung für seine Anwendbarkeit ist die Stetigkeit der Verteilungsfunktion in der Grundgesamtheit.</p>
Differenz der Messwertpaare	<p>Für den Vorzeichentest wird die Differenz jedes Messwert-Paares gebildet. Bei einseitiger Fragestellung, d.h. wenn lediglich interessiert, ob die Werte eines Messverfahrens wesentlich größer als die des anderen sind, so wird die Nullhypothese (gleich viele negative wie positive Differenzen) abgelehnt, wenn die Anzahl der positiven Differenzen einen kritischen Wert übersteigt.</p>

4 Anpassungstests

4.1

Nullhypothese und Alternativhypothese Anpassungstests prüfen, ob eine gegebene Stichprobe aus einer Grundgesamtheit mit einer bestimmten Verteilung stammt. Es wird also die Nullhypothese H_0 geprüft:
Das Merkmal x hat die Wahrscheinlichkeitsverteilung $F(x_0)$, d.h. $H_0 : F(x) = F(x_0)$

Als Alternativhypothese H_1 wird in der Regel die zweiseitige Fragestellung $H_1 : F(x) \neq F(x_0)$ betrachtet.

Wichtige Anpassungstests Die wichtigsten Anpassungstests sind:

- χ^2 -Anpassungstest
- Kolmogorow-Smirnow-Test
- Shapiro-Wilk-Test

4.2 Chi-Quadrat-Anpassungstest

Nicht-parametrischer Test Der χ^2 -Anpassungstest ist ein **nicht-parametrischer Test** der überprüft, ob eine Stichprobe aus einer Grundgesamtheit mit einer bestimmten Verteilung stammt. Der Test vergleicht die **beobachtete statistische Verteilung** der Stichprobe mit einer **theoretischen Verteilung** der Grundgesamtheit.

Testgröße Zur Prüfung der **Hypothese** H_0 wird die **Testgröße**

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - n_{j0})^2}{n_{j0}} \quad (13)$$

berechnet, welche die **theoretischen Häufigkeiten** n_{j0} mit den **empirischen Häufigkeiten** n_j der Klassen j vergleicht.

Ablehnung der Nullhypothese Die **Nullhypothese wird abgelehnt**, wenn die Testgröße groß, d.h. die Differenz zwischen beobachteten und theoretischen Häufigkeiten klein ist. Wenn die **Nullhypothese richtig ist**, also nicht abgelehnt wird, dann genügt die zugehörige Zufallsgröße näherungsweise einer χ^2 -Verteilung mit $k-1$ Freiheitsgraden.

Um dies zu entscheiden wird der Wert der Testgröße χ^2 aus den beobachteten Werten mit dem theoretischen Wert $\chi_{k-1;1-\alpha}^2$ verglichen. Dieser kann nach Wahl eines **Signifikanzniveaus** α (z.B. $\alpha=0,05$) aus Tabellen abgelesen werden (von einer Statistiksoftware wird er berechnet).

Gilt

$$\chi^2 \geq \chi_{k-1;1-\alpha}^2 \quad (14)$$

so wird H_0 **abgelehnt**, d.h. die Grundgesamtheit hat mit hoher Wahrscheinlichkeit nicht die Verteilung $F_0(x)$. Für

$$\chi^2 \leq \chi_{k-1;1-\alpha}^2 \quad (15)$$

wird H_0 **nicht abgelehnt**, d.h. die Verteilungsannahme $F_0(x)$ ist gerechtfertigt.

Damit dies gilt, dürfen die theoretischen Häufigkeiten jedoch **nicht zu klein** sein. In der Regel werden Häufigkeiten $n_{j0} \geq 5$ gefordert. Ist dies nicht erfüllt, so müssen benachbarte Klassen K_j, K_{j-1} zusammengefasst werden.

4.2.1 Nebenpfad: Bestimmung der beobachteten und theoretischen Verteilung

Beobachtete Häufigkeiten Die empirische Verteilung wird bestimmt, indem die Merkmalsausprägungen des zu untersuchenden Merkmals x in Klassen K_1, K_2, \dots, K_k eingeteilt werden und die absoluten oder relativen **beobachteten Häufigkeiten** für diese Klassen ermittelt werden.

Theoretische Häufigkeiten Zur Bestimmung der **theoretischen Häufigkeiten** wird berechnet, wie viele Beobachtungen einer Stichprobe der Größe n in einer Klasse K_j liegen müssten, wenn das Merkmal x tatsächlich die Verteilung $F_0(x)$ hätte. Dann ist

$$n_{j0} = p_j \cdot n \tag{16}$$

die unter der hypothetischen Verteilung $F_0(x)$ in der Klasse j zu erwartende Häufigkeit des Merkmals x .

Um p_j berechnen zu können, müssen die Parameter der angenommenen Verteilung **aus der Stichprobe geschätzt** werden.

5 Ausreißer-Tests

5.1

Problemstellung Die Messwerte einer Stichprobe enthalten teilweise einzelne Werte, welche von den übrigen Werten **extrem abweichen**. Dies kann beispielsweise auf Grund **falscher oder verfälschter Messungen** der Fall sein. Das Auftreten solcher Ausreißer kann darauf basierende statistische Berechnungen stark beeinflussen.

Daher ist es wünschenswert, die Stichprobe von solchen Werten bereits **im Vorfeld statistischer Analysen zu bereinigen**. Hierzu muss jedoch entschieden werden, ob ein solcher Ausreißerwert tatsächlich ein verfälschter Wert ist oder ob es sich um einen extremen Wert einer Stichprobe handelt, der in der Stichprobe enthalten bleiben sollte.

Nullhypothese Ein extremer Wert wird somit nur als Ausreißer identifiziert, wenn durch einen Ausreißertest die Nullhypothese abgelehnt wird, die besagt, dass der Ausreißer ein Stichprobenwert aus der Grundgesamtheit ist. Ein so erkannter Ausreißer kann mit einem anschließenden Bereinigungsverfahren aus der Stichprobe eliminiert werden.

Alternativhypothese Bei den meisten Ausreißertests besteht die Alternativhypothese in der Annahme, dass ein Stichprobenwert aus einer **Grundgesamtheit mit einem größeren bzw. kleiner Mittelwert** stammt.

Einseitige und zweiseitige Ausreißer-Tests Sind Abweichungen vom Mittelwert nach beiden Seiten zugelassen, so handelt es sich um einen zweiseitigen Ausreißertest. Interessiert dagegen nur die Abweichung nach einer Seite, so handelt es sich um einen einseitigen Test. Mit solch einem Test kann immer nur ein Ausreißer erkannt werden.

5.2 Faustregeln

Test bei Normalverteilungsannahme Wird die Normalverteilungsannahme für die Grundgesamtheit vorausgesetzt, so kann ein extremer Wert als Ausreisser angesehen werden, wenn er nicht in einem bestimmten **Bereich der Standardabweichung** liegt. Meist wird dieser Bereich durch die 2,5-fache Standardabweichung definiert. Etwa 99% der Beobachtungen einer Normalverteilung liegen in diesem Bereich.

Test ohne Normalverteilungsannahme Kann die Normalverteilungsannahme nicht vorausgesetzt werden, so wird oft ein Bereich von +/- **4 Standardabweichungen** verwendet. Hiermit kann sichergestellt werden, dass mindestens 94% der Beobachtungen in diesem Bereich liegen.

Test bei un-symmetrischen Verteilungen Ist die zugrundeliegende Verteilung unsymmetrisch, so kann der Bereich, außerhalb dessen Werte als Ausreisser definiert werden, durch den **Interquartilsabstand** zwischen dem 1. $x_{0.25}$ und dem 3. Quartil $x_{0.75}$ definiert werden:

$$x_{0.25} - 1.5(x_{0.75} - x_{0.25}) < x_i < x_{0.75} + 1.5(x_{0.75} - x_{0.25}) \quad (17)$$

Die bekannteste Darstellung, wo dieser Test eingesetzt wird, ist der Boxplot.

5.3 Ausreißer-Tests nach Dixon und nach Grubbs

Ausreißertest nach Dixon Mit dem **Ausreißertest bei normalverteilter Grundgesamtheit** nach Dixon kann die **einseitige Fragestellung für den größten bzw. den kleinsten Wert** einer geordneten Stichprobe überprüft werden. Wird der größte Wert $x_{(max)}$ untersucht, so wird die Testgröße

$$t = \frac{x_{(max)} - x_{(max-1)}}{x_{(max)} - x_{(1)}} \quad (18)$$

verwendet.

Wird der kleinste Wert der Stichprobe überprüft, so wird als Testgröße

$$t = \frac{x_{(2)} - x_{(1)}}{x_{(min)} - x_{(1)}} \quad (19)$$

verwendet. In beiden Fällen wird H_0 abgelehnt, wenn $t \geq \tau_{n;\alpha}$ gilt.

Die Werte für $\tau_{n;\alpha}$ sind in Abhängigkeit vom Stichprobenumfang n und der Irrtumswahrscheinlichkeit α [1](#) tabelliert.

Erweiterter Grubbs-Test Eine Erweiterung des Grubbs-Test als **zweiseitiger Test** setzt ebenfalls eine normalverteilte Grundgesamtheit voraus, wobei die Parameter μ und σ^2 nicht als bekannt vorausgesetzt werden.

Zur Prüfung der Nullhypothese wird die Testgröße

$$t = \frac{\max_i |x_i - \bar{x}|}{s_1} \quad (20)$$

verwendet mit \bar{x} als arithmetischem Mittel der Stichprobenwerte und s_1 als Mittel der quadratischen Abweichungen von \bar{x} .

Die Nullhypothese wird **abgelehnt**, wenn $t \geq w_{n;\alpha}$ gilt, wobei die Werte $w_{n;\alpha}$ tabelliert vorliegen. Ablehnung von H_0 bedeutet, dass entweder der größte oder der kleinste Wert der Stichprobe als Ausreißer erkannt wird.

5.3.1 Nebenpfad: Tabellierte Werte für den Test nach Dixon

Die Tabelle zeigt einen Auszug aus den tabellierten Werten $\tau_{n;\alpha}$ in Abhängigkeit vom Stichprobenumfang n und der Irrtumswahrscheinlichkeit α .

$\tau_{n;\alpha}$	0.005	0.01	0.05	0.1	0.2
3	0.994	0.988	0.941	0.886	0.782
5	0.824	0.782	0.643	0.559	0.452
7	0.681	0.636	0.507	0.433	0.344
9	0.596	0.555	0.436	0.370	0.291
10	0.568	0.527	0.412	0.349	0.274
15	0.473	0.438	0.338	0.284	0.220
20	0.426	0.393	0.300	0.251	0.193
25	0.395	0.364	0.277	0.230	0.176
30	0.371	0.342	0.260	0.216	0.165

5.4 Ausreißer-Test nach Walsh

Ausreißertest nach Walsh Der Test von Walsh kann bei **nicht-normalverteilten Daten** angewendet werden. Er benötigt jedoch eine relativ große Anzahl von Beobachtungen (z.B. $n \geq 220$ für ein Signifikanzniveau von $\alpha = 0.05$).

Zur Durchführung des Tests sind folgende Schritte erforderlich:

1. Sortieren des Datensatzes in aufsteigender Reihenfolge x_1, x_2, \dots, x_n .
2. Festlegen einer Anzahl $r \geq 1$ möglicher Ausreisser.
3. Berechnen von $c = \lceil \sqrt{2n} \rceil$ ($\lceil x \rceil$ ist die nächstgrößere ganze Zahl von x , z.B. $\lceil 4.27 \rceil = 5$) und

$$a = \frac{1 + \sqrt{1/\alpha} \cdot \sqrt{\frac{c-1/\alpha}{c-1}}}{c - 1/\alpha - 1} \quad (21)$$

4. Dann gilt:

- Die **r kleinsten Beobachtungen** sind Ausreisser (beim Signifikanzniveau α), wenn gilt:

$$x_r - (1 + a) \cdot x_{r+1} + a \cdot x_k < 0$$

- Die **r größten Beobachtungen** sind Ausreißer, wenn gilt:

$$x_{n+1-r} - (1 + a) \cdot x_{n-r} + a \cdot x_{n+1-k} > 0$$

- Gelten beide Ungleichungen, so können sowohl die r kleinsten als auch die r größten Beobachtungen als Ausreißer betrachtet werden.

5.5 Weitere Ausreißer-Tests

Ausreißer in
hochdimensionalen
Räumen

Üblicherweise können statistische Ausreißertests **nur für ein Attribut** durchgeführt werden. Dies ist jedoch für viele reale Anwendungen zu wenig, meist müssen Ausreißer in hochdimensionalen Räumen gesucht werden.

Außerdem sind für statistische Tests meist **Kenntnisse über bestimmte Parameter** der Grundgesamtheit notwendig, die in der Realität selten vorliegen und in dem Fall, für den kein spezieller Test existiert, kann nicht garantiert werden, dass alle Ausreißer gefunden werden.

Distanzmessungen

Eine weitere Strategie zum Auffinden von Ausreißern basiert auf Distanzmessungen zwischen Objekten.

Ein Objekt wird als Ausreißer $DB(p, d)$ identifiziert, wenn ein Anteil p aller Objekte weiter als d vom betrachteten Objekt entfernt ist.

Die **Komplexität** dieses Problems liegt in der Suche nach Nachbarobjekten innerhalb d für jedes Objekt. Hierfür existieren eine Reihe von Lösungen wie beispielsweise die Verwaltung der Objekte in indizierten Bäumen oder in verschachtelten Schleifen. Problematisch an diesen Algorithmen ist die Bestimmung der Parameter p und d , welche der Benutzer festlegen muss.

6 Variablenauswahl

6.1 Problemstellung

Teilmengen der Merkmale

Obwohl in vielen Fällen eine große Anzahl an Merkmalen zur Untersuchung einer statistischen Fragestellung zur Verfügung steht, ist es häufig wünschenswert nur eine **begrenzte, aussagekräftige Untermenge der Merkmale** zu analysieren. Gründe hierfür sind:

- Ist die Anzahl der verfügbaren Beobachtungen im Verhältnis zur Anzahl der Merkmale relativ klein, so kann dies zu einer **schlechten Anpassung des Modells** an die Daten führen, da die Modellparameter nur ungenau geschätzt werden können.
- Zu viele Merkmale können zu einer **Überanpassung** (Overfitting) des Modells an die Daten führen (z.B. bei der Klassifikation).
- Eine häufige Fragestellung bei der statistischen Analyse ist die nach den „wichtigen“, **einflussreichen Merkmalen**.
- Durch die Identifizierung aussagekräftiger Merkmale kann die **(teure) Erhebung nicht aussagekräftiger Merkmale** vermieden werden.

6.2 Lösungsansätze

Vollständige Enumeration Im Prinzip können **alle möglichen Kombinationen von Merkmalen** ausprobiert und die hiermit erzielten Ergebnissen verglichen werden. Dies ist aber für reale Probleme aufgrund der großen Anzahl möglicher Kombinationen nicht durchführbar.

Es gibt eine Vielzahl von **Ansätzen zur Merkmalsauswahl**, Beispiele sind:

- **I** Mallows' Cp-Statistik
- Schrittweise Prozeduren (Vorwärtsselektion, Rückwärtselimination, Schrittweise Regression)

6.2.1 Nebenpfad: Mallows Cp-Statistik

Mallows Cp-Statistik sucht aus einer Menge von Merkmalen eine **Merkmalskombination mit hoher Aussagekraft** für die Beobachtungen einer Stichprobe. Die Statistik trägt eine Testgröße C_p gegen die reduzierte Anzahl der Merkmale p auf und gibt auf Grundlage dieser Grafik einen Hinweis auf die günstigste Merkmalskombination.

Die Statistik setzt einen **linearen Zusammenhang** zwischen den Merkmalen voraus.

Die Teststatistik wird berechnet zu:

$$C_p = \frac{\sum_{i=1}^n (Y_i - \beta x_i)}{\frac{1}{n-p-1} \sum_{i=1}^n (Y_i - \beta x_i)} - n + 2(p + 1) \quad (22)$$

mit der **Quadratsumme der Residuen für das Regressionsmodell** auf der Basis einer Teilmenge von p Merkmalen

$$\sum_{i=1}^n (Y_i - \beta x_i) \quad (23)$$

und dem Mittelwert der quadrierten Residuen bei der Modellierung mit allen P Merkmalen

$$\sum_{i=1}^n (Y_i - \beta x_i) \quad (24)$$

, n als Anzahl der Beobachtungen und p als Anzahl der Merkmale im reduzierten Modell.

Der Cp-Wert wird **für alle Kombinationen von Merkmalen berechnet** und gegen die jeweilige Anzahl p der Merkmale geplottet. Das Modell bzw. die Merkmalskombination mit dem niedrigsten Cp-Wert, der in etwa gleich dem Wert für p ist, wird als das **aussagekräftigste Modell** betrachtet.

7 Literatur

7.1 Literatur zur Zeitreihenanalyse

Literaturverzeichnis

- Bleymüller, J./Gehlert, G./Gülicher, H.: Statistik für Wirtschaftswissenschaftler. Franz Vahlen, München 1991.
- Bohley, P.: Statistik-Lehrbuch für Wirtschaftswissenschaften. Oldenbourg, München 1989.
- Elpelt, H.: Grundkurs Statistik. Oldenbourg, München 1987.
- Fahrmeir, L./Künstler, R./Pigeot, I./Tutz, G.: Statistik. 5. Aufl., Springer, Berlin Heidelberg New York 2004.
- Hennig, C.: Modellwahl und Variablenselektion in der Statistik. Veröffentlichtes Vorlesungsskript, Universität Hamburg, Fachbereich Mathematik (SPST), SS 2004, auf URL: <http://www.math.uni-hamburg.de/home/hennig/lehre/mskript1.pdf>
- Hochstädter, D.: Statistische Methodenlehre. 8. Aufl., Verlag Harri Deutsch, Frankfurt/Main 1996.
- Litz, H. P.: Statistische Methoden in den Wirtschafts- und Sozialwissenschaften. Oldenbourg Verlag, München Wien 2003.
- Schlittgen, R.: Einführung in die Statistik. Oldenbourg, München 1991.
- Zöfel, P.: Statistik für Wirtschaftswissenschaftler. Pearson Studium 2003.