Hallo,

kurze Anmerkung: Diese Scripte stammen von 1999. Ich kann leider dazu

keine Fragen mehr beantworten! :-(

Euch trotzdem viel Erfolg!

Dorthe

dorthe@luebbert.net

Statistik A+B Begriffe+Theorien

1 Allgemeine Begriffe zur Wahrscheinlichkeitsrechnung

1.1 Zufallsvariable

Jede Regel (oder Funktion) X, die jedem Elementarereignis eines Ereignisraumes eine reelle Zahl und gleichzeitig die zu den Elementarereignis gehörende Wahrscheinlichkeit der reellen Zahl zuordnet, heißt Zufallsvariable. Eine **diskrete Zufallsvariable** X liegt dann vor, wenn jedem möglichen Ereignis eines endlichen Ereignisraumes eine Zahl x_i aus der Menge der Zahlen {x1, x2, x3...xk} zugeordnet wird.

Eine **stetige Zufallsvariable** X liegt dann vor, wenn jedem möglichen Ereignis eines endlichen Ereignisraumes eine Zahl x aus einem Intervall I: $a \leq x \leq b$ zugeordnet wird.

1.2 Erwartungswert

Der Erwartungswert drückt aus, welchen Wert die Zufallsvariable im arithmetischen Durchschnitt realisieren wird. Für diskrete Zufallsvariablen ist der Erwartungswert μ die Summe aus dem Produkt aller Werte mit den ihnen zugeordneten Wahrscheinlichkeiten: $E(X) = \sum_{i=1}^{n} x_i \cdot P(x_i) = \mu$

Für eine stetige Variable berechnet sich der Erwartungswert nach der Formel $E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \mu$

1.3 Wahrscheinlichkeitsfunktion

Eine Zufallsvariable X hat endlich oder abzählbar unendlich viele Werte, d.h. der Wertebereich hat die Gestalt $\{x_1, x_2, x_3 ...\}$. Diese Zufallsvariable und auch deren Verteilung heißen diskret. Die Wahrscheinlichkeitsfunktion $P(X=x_i)=P(x_i)$ ordnet jeder reellen Zahl X die Wahrscheinlichkeit zu, mit der sie von X angenommen wird. Die Wahrscheinlichkeitsfunktion sagt also aus, mit welcher Wahrscheinlichkeit eine bestimmte Ausprägung einer Zufallsvariablen bei einem Zufallsexperiment auftritt.

1.4 Verteilungsfunktion

Kumuliert man die Werte der Wahrscheinlichkeitsfunktion für die Werte $_{xi}$, so erhält man die Verteilungsfunktion: $F(t)=P(X \leq x_t)=\Sigma P(x_i)$

1.5 bedingte Wahrscheinlichkeit

Die bedingte Wahrscheinlichkeit $P(B|A)$ bezeichnet die Wahrscheinlichkeit des Ereignisses B unter der Voraussetzung, daß das Ereignis A eingetreten ist. Berechnet werden kann sie z.B. durch $P(B|A) = \frac{P(A \cap B)}{P(A)}$. Die bedingte Wahrscheinlichkeit gilt also für voneinander abhängige Ereignisse.

Die im Bayesschen Theorem erfaßte Beziehung erlaubt die Umkehrung bedingter Wahrscheinlichkeiten. Sie stellt somit eine a-posteriori Wahrscheinlichkeit dar(Transformation von a priori in a posteriori).

1.6 Permutation und Kombination

In der Kombinatorik wird untersucht, auf welche und auf wieviel verschiedene Arten eine gegebene Anzahl von Elementen angeordnete und zu Gruppen zusammengefaßt werden kan. Zählt die Reihenfolge innerhalb der einzelnen Anordnung, handelt es sich um Permutation, zählt sie nicht von Kombination.

1.7 Theorem von Bayes

Mit dem Theorem von Bayes läßt sich die Wahrscheinlichkeit von A_i unter der Bedingung E ausrechnen. Die Formel, die mit Hilfe des allgemeinen Multiplikationssatzes und dem Satz der totalen Wahrscheinlichkeit hergeleitet wird, lautet:

$$P(A_i|E) = \frac{P(E|A_i)P(A_i)}{\sum_{i=1}^{n} P(E|A_i)P(A_i)}$$

$P(A_i|E)$ bezeichnet man als a-posteriori-Warscheinlichkeit. Die a-priori Wahrscheinlichkeit $P(A_i)$ wird in die a-posteriori-Wahrscheinlichkeit transformiert.

2 Allgemeine Begriffe zur Teststatistik

2.1 Statistischer Test

Ein statistischer Test ist ein Verfahren, das auf Grund eines empirischen Befundes, d.h. von Stichprobenergebnissen, darüber entscheidet, ob eine statistische Hypothese (Annahme oder Behauptung über die unbekannte Verteilung einer Zufallsvariablen) akzeptiert oder verworfen wird.

2.1.1 Elemente des Hypothesentests

Ein Hypothesentest besteht aus den folgenden sechs Schritten:

1. Formulierung einer empirisch überprüfbaren Hypothese
2. Konstruieren einer Entscheidungsregel
3. Ziehen einer Stichprobe
4. Berechnen der Prüfvariablen (Stichprobenkennzahl)
5. Anwenden der Entscheidungsregel (Berechnen der Rückweisungspunkte)
   Der Annahme bzw. Ablehnungsbereich einer Hypothese ist durch sogenannte Rückweisungspunkte begrenzt. Diese müssen bei jedem Test berechnet bzw. abgelesen werden.
6. Ableiten einer Entscheidung (H_O wird verworfen oder angenommen)

2.2 Signifikanztests

Wird bei einer statistischen Entscheidung nur eine einzige Hypothese daraufhin überprüft, ob diese Hypothese *nicht falsch* ist, so nennt man die dazu verwendeten Tests *Signifikanztests.*
Signifikanztests prüfen also die Wahrscheinlichkeit dafür, beim Schluß von der Stichprobe auf die Grundgesamtheit einen Fehler erster Art zu begehen.

2.3 Anteilswerttests und Anpassungstests

Anpassungstests prüfen Hypothesen über die Verteilung einer Zufallsvariablen z.B. H_0: F(x)=F0(X), sie vergleichen beobachtete Verteilungen in einer Stichprobe mit einer erwarteten Verteilung. Der Anpassungstest prüft also, mit welcher Wahrscheinlichkeit eine Stichprobe aus einer Grundgesamtheit stammen kann, für die die erwartete Verteilung gilt. Die einfachste Form ist der Binominaltest.
Anteilswerttests prüfen Hypothesen über die Gleichheit von Anteilswerten. Eine typische Hypothese lautet: Die GG und die Stichprobe sind gleich verteilt: F0(x)=F(x) (→ Binomialtest).
Zusammenhang zwischen Anteilswerts- und Anpassungstest
Betrachtet man nun zunächst nur den dichotomen Fall, bei dem die Zufallsvariable nur zwei Werte realisieren kann, so läßt sich diese Verteilungshypothese auch vereinfachend als Hypothese über den Anteilswert in der Grundgesamtheit formulieren: H_0: $\pi=\pi_0$. Die Wahrscheinlichkeit π_0 ist einerseits die behauptete Wahrscheinlichkeit, daß bei einem Zufallszug ein Element einen bestimmten der beiden möglichen Werte realisiert, andererseits handelt es sich um eine Hypothese über die zu erwartende relative Häufigkeit von einer der beiden Klassen, also einem Anteilswert, der gleichzeitig die Verteilung der Zufallsvariablen bestimmt.
Beide, der Anteilswerttest und der Anpassungstest, testen nun inwieweit die Unterschiede zwischen den beobachteten und den theoretischen Werten zufällig oder signifikant sind und beziehen sich dabei auf die relative Häufigkeit der einzelnen Kategorien. Insofern läßt sich z.B. der Binomialtest sowohl als Anteilswerttest als auch als Anpassungstest interpretieren.

2.4 Multipler Test

Verwendet man z.B. anstelle eines H-Tests mit mehr als zwei unabhängigen Stichproben mehrere U-Tests als Einzeltests, so werden diese Einzeltests als multipler Test angesehen. Zu beachten ist hierbei besonders, daß bei den hintereinandergeschalteten Einzeltests sich die Wahrscheinlichkeit des Fehlers 1. Art mit der Anzahl des Tests vergrößert. Bei einem Vergleich muß dies unbedingt berücksichtigt werden.

2.5 Konservativer Test

Bei einem konservativen Test ist die Prüfvariable diskret verteilt (z.B. U-Test). Es gibt für ein vorgegebenes Signifikanzniveau keine Werte zum Beispiel für ur, die die Gleichung P(u≤ur)=5% erfüllen, deshalb ersetzt man die Gleichung durch die Ungleichung P(u≤ur)≤α%. Man wählt also generell als Rückweisungspunkt jenen Wert, der zu einem Signifikanzniveau von höchsten α führt. Das vorgegebene Signifikanzniveau kann also praktisch erheblich unterschritten werden. → Man verhält sich konservativ und begünstigt die Annahme der Nullhypothese.

2.6 Exakter Test

Ein exakter Test ist ein Test, der für die zu testende Prüfvariable die exakt zuständige Stichprobenverteilung verwendet. Ein exakter Test approximiert also nicht. Exakte Test sind z.B. der Fisher-Test, der Binomial-Test, der McNemar-Test. Nicht exakt arbeitet zum Beispiel ein Test, bei dem man nach dem Zentralen Grenzwert-Theorem die Normalverteilung approximativ für eine Binomialverteilung verwendet.

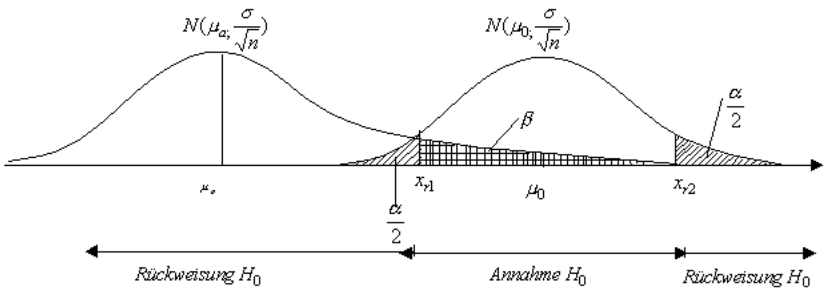
2.7 α- und β-Fehler, Überschreitungswahrscheinlichkeit

Hypothesen können nie letztendlich verifiziert oder falsifiziert werden. Die „Annahme“ einer Hypothese sagt nur: die vorliegende statistische Evidenz reicht nicht aus, um die Hypothese zu verwerfen.
Für die richtige bzw. falsche Schlußfolgerung eines Tests gilt folgende Fehlersystematik.

Schlußfolgerung des Tests ↓	unbekannte Wirklichkeit →	H_0 ist richtig	H_0 ist falsch
H_0-Annahme		richtige Schlußfolgerung	Fehler, 2. Art (β-Fehler)
H_0 -Ablehnung		Fehler 1. Art (α-Fehler)	richtige Entscheidung

Das Signifikanzniveau eines Testverfahrens ist dabei die Wahrscheinlichkeit, mit der die Nullhypothese abgelehnt wird, obwohl diese in Wahrheit zutrifft. Das Signifikanzniveau ist damit gleich dem Fehler 1. Art (α-Fehler). Der α-Fehler heißt auch **Überschreitungswahrscheinlichkeit.** Wenn der α-Fehler ≥ α (bzw α/2), dann wird die H_0 verworfen.

Die Lage der Stichprobenverteilung zur Alternativhypothese bestimmt die Wahrscheinlichkeit des β-Fehlers. α und β stehen in reziprokem Verhältnis zueinander.



Bei gleichen Bedingungen (ceteris paribus) wird der β-Fehler kleiner,

1. wenn die Differenz zwischen μ_0 und μ_a zunimmt (die linke Kurve wandert nach links)
2. wenn der α-Fehler erhöht wird (die Grenze des α-Fehlers wandert nach rechts)
3. Wenn der Umfang n vergrößert wird

Rechnerisch bestimmt man β, indem man die Rückweisungspunkte x_{r1} und x_{r2}, die den Annahmebereich der H_0 begrenzen, mit den Werten der alternativen Stichprobenverteilung standardisiert. Die Fläche des Intervalls (entspricht der Wahrscheinlichkeit im Prüfpunkt) ist der β-Fehler.

Beispiel zur Berechnung des β-Fehlers

In einer Stichprobe von n=100 Zigarettenrauchern, fand man heraus, daß diese im Durchschnitt $\bar{x}$ =7 Zigaretten pro Tag rauchten (σ=4). Dem steht die Annahme entgegen, daß Zigarettenraucher im Durchschnitt μ_0=8 Zigaretten pro Tag rauchen. Der durchgeführte

Signifikanztest für α=5% führte zu den Rückweisungspunkten $\bar{x}_{r1} = 7{,}216$ und $\bar{x}_{r2} = 8{,}784$ ($k_{r2} = 1{,}96$). Die H_0 konnte angenommen werden, da der Stichprobenbefund von 7 im Annahmebereich lag.
Um den β-Fehler zu berechnen, unterstellt man, daß der wahre Grundgesamtheitsparameter μ_a so groß ist wie der Stichprobenbefund $\bar{x} = 7.4$. Die alternative Stichprobenfunktion ist also verteilt nach $N(\mu; \frac{\sigma}{\sqrt{n}})$ =N(7,4;0,4). Gesucht ist $P(\bar{x}_{r1} \le \bar{X}_1 \le \bar{x}_{r2}) = \beta$. Man standardisiert

$\bar{x}_{r1} = 7{,}216$: ${}_1k_{r1} = \frac{X-\mu}{\sigma} = \frac{7{,}216 - 7{,}4}{0{,}4} = -\,0{,}46$

sowie ${}_1k_{r1} = \frac{8{,}784 - 7{,}4}{0{,}4} = 3{,}46$. Die tabellierte Verteilungsfunktion liefert das Ergebnis

$P(\bar{x}_{r1} \le \bar{X}_1 \le \bar{x}_{r2}) = P({}_1k_{r1} \le K_1 \le {}_1\ k_{r2}) = 0{,}677$

Ergebnis: Sollte also der tatsächliche durchschnittliche Zigarettenverbrauch 7,4 betragen, so beträgt die Wahrscheinlichkeit 67,7 Prozent, unter den gegebenen Umständen (μ,σ…) einen Fehler zweiter Art zu begehen.

Fazit
Das Verhältnis von α und β muß optimiert werden, um eine optimale Zuverlässigkeit des Tests zu erreichen.
Will man α und β gleichzeitig verringern, muß man den Stichprobenumfang ehöhen.

2.8 Statistische Hypothesen

2.8.1 Typen von Hypothesen: einfach, zusammengesetzt...

Man unterscheidet **einfache** und **zusammengesetzte** Hypothesen:
eine einfache Hypothese (Punkthypothese) besteht aus der Behauptung eines bestimmten Wertes, z.B. μ=10.000 km.
Eine zusammengesetzte Hypothese umfaßt ein Werteintervall, z.B. $\mu \ge$10.000 oder $\mu \ne$8.000 km.

Man unterscheidet
rechtsseitige Alternativhypothesen:
der Alternativwert μ_a liegt rechts vom μ_o, entweder als Punktwert oder Werteintervall.
Beispiel: ein rechtsseitiger Test liegt vor, wenn bei einem μ_o=8.000 ein Alternativwert von μ_a=10.000 oder sogar ein Werteintervall von μ_o>8000 getestet wird.
linksseitige Alternativhypothesen
beidseitige Alternativhypothesen

Unter einer **spezifizierter Verteilungshypothese** versteht man eine Hypothese, bei der ein oder mehrere explizite Parameter aus dem Stichprobenbefund heraus geschätzt werden müssen.

2.8.2 Homogenitätshypothese

Eine Homogenitätshypothese wird z.B. beim Chi-Quadrat-Homogenitätstest zugrunde gelegt. Sie behauptet, daß ein Merkmal in den zwei oder mehr Stichproben jeweils zugrundeliegenden Grundgesamtheiten jeweils die gleiche Verteilung hat. Genau genommen impliziert die Homogenitätshypothese, daß die empirischen Häufigkeiten mit den theoretischen Häufigkeiten übereinstimmen.

2.9 Kriterien für statistische Prüfverfahren

1. <u>**Meßniveau als wichtigstes Kriterium**</u>

   Das Meßniveau ist das grundsätzliche und wichtigste Kriterium für statistische Prüfverfahren. Man unterscheidet:

   Prüfverfahren bei nominalem Meßniveau
   - Binominaltest (Anteilswerttest)
   - χ^2-Anpassungstest
   - χ^2-Unabhängigkeitstest
   - Mc Nemar Test
   - Fisher Test

   Prüfverfahren bei ordinalem Meßniveau

- Mediantest
- Wilcoxon-Vorzeichen-Rangtest
- Komogoroff/Smirnov-Test
- Mann/Whitney-U-Test
- Kruskal/Wallis-Test
- Test des Rangkorrelationskoeffizienten nach Spearman

Prüfverfahren bei metrischem Meßniveau

- t-Test für das arithmetische Mittel
- f-Test

2. **Anzahl der Untersuchungsvariablen (UV)**

   univariate Prüfverfahren nominalskalierter Untersuchungsgrößen (→ d.h eine Größe wird untersucht, das ist mit allen Prüfverfahren möglich, außer dem Rangkorrelationskoeffizienten)
   bivariate Prüfverfahren (Rangkorrelationskoeffizient)
   multivariate Prüfverfahren

3. **Anzahl der Stichproben**

   a) ein Stichprobenfall:
      → Binominaltest
      → χ^2-Anpassungstest
      → tw. Kolmogoroff/Smirnow

   b) zwei-Stichprobenfall
      → McNemar
      → Fisher

   c) mehr als zwei-Stichprobenfall:
      → H-Test

4. **abhängige/unabhängige Stichproben**

1. **Stichprobenumfang**

   Ein weiteres Klassifizierungskriterium ist die Größe der Stichprobe. Bei genügend großem Stichprobenumfang kann man nach dem Grenzwertsatz von Laplace-Moivre die interessierende Variable mittels einfach zu handhabender Verteilungen ausreichend gut nähern.

2. **parameter/parameterfreie Tests**

   Man unterscheidet parameter und parameterfreie Tests.
   Beim **Parametertests** interessieren konkrete Werte wie π, σ und μ. Ein parametrisches Prüfverfahren macht also Aussagen über Grundgesamtheitsparameter bzw. die in der Verteilungsfunktion einer Untersuchungsvariablen auftretenden Konstanten. Dazu müssen alle Paramter der GG bekannt sein (was oft nicht gegeben ist → Problem). Bei einem Parametertest hat jede der denkbaren Stichproben die gleiche Realisationschance.
   Bei **parameterfreien Tests** (auch nichtparametrische Tests bzw. Verteilungstests genannt) wird der Typ der Zufallsverteilung überprüft: Man entscheidet, ob eine aus *n* Beobachtungen bestehende Häufigkeitsverteilungen bestehende Nullhypothese, die man aus einer Zufallsstichprobe gezogen hat, mit einer Null-Hypothese vereinbar ist, die man über die Verteilung in der Grundgesamtheit aufgestellt hat.

7. **verteilungsfreie/verteilungsgebundene Testverfahren**

   verteilungsfreier Test: über die Verteilung der Grundgesamtheit keinerlei Voraussetzungen gemacht. Man bezeichnet verteilungsunabhängige Tests, da Grundgesamtheitsparameter keine Rolle spielen, auch als parameterfreie Tests. Verteilungsfreie- oder unabhängige Verfahren werden allgemein angewendet bei nicht normalverteilten Grundgesamtheiten, bei ordinal- oder nominalskaldierten Werten, zur Kontrolle eines parametrischen Tests sowie als Schnelltest.
   Bei **verteilungsgebundenen Tests**, wie z.B. dem t-Test, hängt die Verteilung der Prüfvariablen von der Verteilung der Grundgesamtheit (hier Normalverteilung und Streuungsgleichheit) ab. Diese Prüfverfahren betreffen durchweg metrisch skalierte Untersuchungsvariablen. Eine Im allgemeinen haben verteilungsfreie Tests eine geringere Güte als verteilungsabhängige Tests, da sie oft nur einen Teil der im Zahlenmaterial enthaltenden Informationen auswerten.
   verteilungsgebundener Test: die Verteilung der Untersuchungsvariablen hängt von der Verteilung der Grundgesamtheit ab.

2.10 Konstruktion einer geeigneten Prüfvariable

- Die Prüfvariable muß mit dem Problem der Fragestellung korrespondieren, d.h. mit der Hypothese übereinstimmen. Bsp.: Zur Prüfung des Medians ist es z.B. zweckmäßig, die Abweichungen der

Stichprobenwerte von dem entsprechenden Hypothesenwert zu berücksichtigen. Die Prüfvariable des Wilcoxon-Vorzeichen-Rangtests etwa, berücksichtigt neben den Rängen der Abweichungen, über eine 0-1-Kodierung auch deren größer-kleiner Relation hinsichtlich des Medians. Da sich zudem, unter Gültigkeit von H0, ein Erwartungswert formulieren läßt, ist es auch möglich zu prüfen, inwieweit die Abweichungen von ihm signifikant sind.

- Die Verteilung der Prüfvariablen muß bekannt sein, man muß Rückweisungspunkte ermitteln können. Unter Angabe des Fehlers 1. Art zerfällt die Realisation der Prüfvariablen dann in zwei Klassen, dem Annahme- und Rückweisungsbereich.
- Die Verteilung der Prüfvariablen muß bekannt sein für den Fall, daß H_0 in Wahrheit zutrifft.

2.11 Freiheitsgrade

Bezeichnung für die Anzahl von Werten, die innerhalb der Begrenzungen eines Systems von Werten frei variieren oder gewählt werden können. Anders formuliert: Die Anzahl der Freiheitsgrade *v* ist definiert als die Differenz aus dem Stichprobenumfang n und der Anzahl k der aus den n Stichprobenmeßwerten berechneten Parameter *v=n-k.*

Freiheitsgrade werden auch als explizite Parameter verwendet, so ist *v* der einzige explizite Parameter der Chi-Quadrat-Verteilungsfunktion.

3 Stichproben

3.1 Stichprobe

Die Auswahl von Elementen einer Grundgesamtheit wird Stichprobe genannt. Stichproben sind billiger als Totalerhebungen, haben zudem eine kürzere Erhebungs- und Auswertungszeit. Eine Stichprobe stellt somit eine Teilmenge aller Untersuchungseinheiten dar, die die untersuchungsrelevanten Eigenschaften der Grundgesamtheit möglichst genau abbilden soll. Man unterscheidet generell zwischen bewußten Auswahlverfahren und zufälligen Auswahlen.

3.2 Verbundene/unverbundene Stichprobe (= unabhängige/abhängige Stichprobe)

Die Frage nach der statistischen Abhängigkeit ist Untersuchungskriterium bei mehr als einer Stichprobe:

unabhängige Stichprobe:

$$P(A|B) = P(A)$$

Unterschiedliche Werte bei einer Variablen werden gemessen, anders ausgedrückt: gezogene Stichproben werden wieder zurückgelegt, bevor neue Stichproben gezogen werden.

Bei unabhängige Stichproben ist die Wahrscheinlichkeit generell dafür, daß z.B. ein Merkmalsträger der ersten Stichprobe eine bestimmte Merkmalsausprägung trägt, so groß wie die bedingte Wahrscheinlichkeit, daß eben dieser Merkmalsträger die gleiche Merkmalsausprägung trägt unter der Bedingung, daß ein Merkmalsträger aus der zweiten Stichprobe eine bestimmte Merkmalsausprägung trägt. Es besteht also keine Verbindung zwischen beiden Stichproben, die Einzelwahrscheinlichkeiten aus den verschiedenen Stichproben beeinflussen sich nicht gegenseitig.

Beispiel: Der IQ von zwei verschiedenen Personengruppen wird getestet.

Abhängige Stichproben

Mehrere Stichproben werden an ein und derselben Gruppe von Versuchspersonen oder an zwei Gruppen von Versuchspersonen, die einander paarweise zugeordnet sind, erhoben. Dieselbe inhaltliche Variable wird bei derselben Gruppe zwei - der mehrfach gemessen.

Anders ausgedrückt: formal verschiedene Variablen, die zwar inhaltlich das Gleiche messen, aber zu unterschiedlichen Zeitpunkten bzw. in unterschiedlichen Kontexten, d.h. zu verschiedenen Bedingungen.

Als abhängig werden Stichproben dann bezeichnet, wenn die Wahrscheinlichkeit dafür, daß ein Merkmalsträger eine bestimmte Merkmalsausprägung trägt, die Wahrscheinlichkeit für eine bestimmte Merkmalsausprägung in Stichprobe zwei maßgeblich beeinflußt. Dies ist immer dann der Fall, wenn zwischen den Stichproben kausale Zusammenhänge bestehen, z.B. bei einer Untersuchung, ob die Art der Beleuchtung eines Arbeitsplatzes die Fehlerquote beeinflußt. Dabei wird der Grad der Beleuchtung in engem Zusammenhang mit der Fehlerquote stehen, die Wahrscheinlichkeit für das Auftreten einer bestimmten Quote also beeinflussen.

Beispiel: Die mentale Leistungsfähigkeit von einer Personengruppe wird morgens um acht und nach einem Zirkeltraining getestet.

3.3 Stichprobenfehler
Genaugenommen können Aussagen aufgrund von Stichproben nur für die Stichprobe selbst Gültigkeit beanspruchen. Für die GG aus der die Stichprobe gezogen wurde, wird die Gültigkeit nur angenommen. Der Stichprobenfehler ist die Streuung der Stichprobenverteilung bzw. die Differenz zwischen der Maßzahl einer Stichprobe und dem entsprechenden wahren Wert in der GG.
Die durch das Ziehen einer Zufallsstichprobe entstandene Abweichung zwischen dem wahren Wert einer Variablen in der GG u. dem Stichprobenfehler ist um so geringer, je geringer die Varianz einer Verteilung und je größer der Umfang der SP ist. Die Verteilung der SP-Fehler aller möglichen SP gleicher Größe aus einer GG bildet die SP-Verteilung, die zur Prognose von SP-Werten und auch zur Schätzung der GG-Parameter herangezogen wird. Je größer der Stichprobenumfang *n*, desto kleiner ist der Stichprobenfehler. Der Stichprobenfehler wird also nicht vom Umfang der Grundgesamtheit bestimmt, sondern neben der Grundgesamtheitsstreuung vom Stichprobenumfang. Da $\overline{X}$ aufgrund des Gesetztes der großen Zahl für n → ∞ gegen μ strebt, wird der Stichprobenfehler ab einer gewissen Stichprobengröße so klein, daß eine Vergrößerung des Stichprobenumfangs die Mehrausgaben nicht mehr rechtfertigen würde.

4 Die Grenzwertsätze

4.1 Hauptsatz der theoretischen Statistik: Der Zentrale Grenzwertsatz
Der zentrale Grenzwertsatz ist ein Hauptsatz in der theoretischen Statistik und besagt im allgemeinen, daß die Summe von stochastisch unabhängigen Zufallsvariablen annähernd normalverteilt ist. Der zentrale Grenzwertsatz läßt sich nutzen, um die Stichprobenverteilung bestimmter Maßzahlen anzugeben. Er besagt:
Wenn man eine Zufallsstichprobe von genügend großem Umfang n aus einer Grundgesamtheit mit Zurückliegen zieht, in der ein Merkmal X mit dem Erwartungswert μ_x und der Varianz $\sigma 2_x^2$ verteilt ist, dann folgt $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ approximativ der Normalverteilung mit den Parametern μ_x und $\frac{\sigma_x^2}{n}$, <u>ohne daß etwas über die Verteilung von X vorausgesetzt werden muß.</u>

Vereinfacht ausgedrückt heißt das: Wächst n_i, nähert sich die Verteilung von ΣX immer mehr der Normalverteilung an. D.h. Wahrscheinlichkeiten, die mit Hilfe der Summe von unabhängigen, identisch verteilten Zufallsvariablen X_i gebildet werden, lassen sich für großes *n* mittels der Normalverteilung hinreichend genau berechnen.

Voraussetzungen und Bedingungen:
- n sollte nach einer groben Faustformel mindestens 30 sein, damit die Summenformel als so gut wie normalverteilt angesehen werden kann.
- Der Verteilungstyp muß nicht bekannt sein, die Zufallsvariablen müssen nicht symmetrisch verteilt sein, allerdings muß eine Varianz existieren und die Verteilung darf nicht absurd, z.B. einer Arcustangensverteilung folgen.
- Die Stichprobenvariablen müssen nicht voneinander stochastisch unabhängig sein.

Spezialfall: normalverteilte Grundgesamtheiten

Falls die Grundgesamtheit, aus der gezogen wird, normalverteilt ist, ist $\overline{X}$ nicht approximativ, sondern genau normalverteilt. Die Stichprobe kann dann auch von kleinem Umfang sein. Die Grundgesamtheit, aus der gezogen wird, ist $N(\mu_x, \frac{\sigma_x^2}{n})$-verteilt.

4.2 La Place: Lokaler Grenzwertsatz
Der Grenzwertsatz von de Moivre/Laplace gilt als Spezialfall des zentrales Grenzwertsatzes. Der lokale Grenzwertsatz, der sich auf die Wahrscheinlichkeitsfunktion bezieht, besagt, daß eine nach B(n, π) verteilte Zufallsvariable mit wachsendem n an die Dichtefunktion einer Normalverteilung N(nπ; $\sqrt{n\pi(1-\pi)}$) grenzt. Er ist ein lokaler Satz, weil er sich auf die festen Stellen x_i der Binomalverteilung bezieht. Der allgemeine Grenzwertsatz betrifft dann die Verteilungsfunktion Hier kann bei n→∞ die Wahrscheinlichkeit der Realisation einer nach B(n;π) verteilte Zufallsvariablen X in

einen bestimmten Bereich, z.B. (1≤X≤5) approximativ über eine Normalverteilung bestimmt werden. Für die sinnvolle Anwendung existiert eine Faustformel, nach der eine hinreichende Genauigkeit erreicht wird.

4.2.1 Warum sind Stichprobenmittelwerte normalverteilt?

Zu jedem Stichprobenmittelwert gehört eine spezifische Zufallsvariable. Addiert man diese Zufallsvariablen, ergibt sich eine neue Zufallsvariable. Diese ist nach dem *zentralen Zufallstheorem* von LaPlace bei zunehmenden *n* normalverteilt. Die gilt unabhängig davon, wie die ursprüngliche Ausgansvariablen verteilt sind.

4.3 Vergleich lokaler und zentraler Grenzwertsatz

Zentraler und lokaler Grenzwertsatz haben unterschiedliche Einsatzbereiche. Der Zentrale GWS macht eine Aussage über die Summe von unabhängigen Zufallsvariablen (künstliche, normalverteilte Variablen).
Der *lokale Grenzwertsatz* bezieht sich auf die Ausprägung einer Variablen in einem Punkt. Die Aussage wird also über einen Punkt getroffen.

5 Wahrscheinlichkeitsverteilungen

5.1 Normalverteilung (Gaußsche Fehlerkurve)

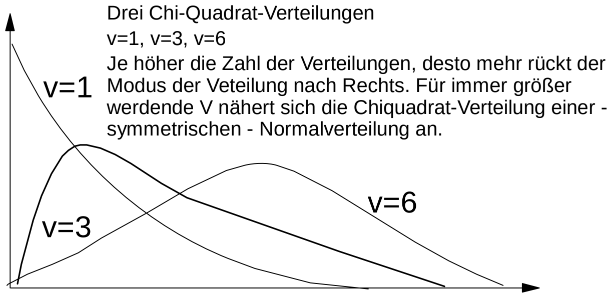
Die Form der Normalverteilung wird durch die Paramter μ und σ bestimmt. Man charakterisiert die Normalverteilung durch N(μ,σ).
Die Normalverteilung hat folgende Eigenschaften:
- arithm. Mittel, Modus und Median fallen zusammen
- die Kurve hat bei x=μ ihr einziges Maximum
- die beiden Äste der Kurve nähern sich asymptotisch der Abzisse
- De Fläche unter der Kurve muß natürlich gleich 1 sein

5.2 Chiquadratverteilung

Die Chiquadratverteilung ist eine **stetige Zufallsverteilung**.
Man geht aus von einer Anzahl v unabhängiger Zufallsvariablen Ki, die standardnormalverteilt sind. Die Summe ihrer Quadrate ergibt die neue Zufallsvariable. Sie wird mit χ^2 bezeichnet.

$$\chi^2 = \sum_{i=1}^{V} K_i^2$$

- χ^2 hat die Dichtefunktion einer Γ-verteilten Variablen mit k=0,5 und λ =0,5v
- Die konkrete Gestalt der χ^2-Funktion hängt davon ab, wie groß v (die Anzahl der Normalverteilungen) ist.

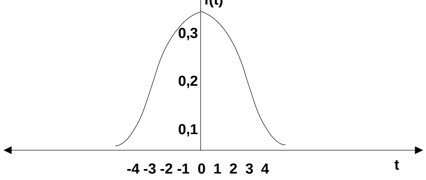


Form der χ^2-Verteilung:

χ^2 kann nur positive Werte realisieren, die χ^2-Verteilung verläuft ab v≥3 unimodal und linkssteil. Mit wachsendem v verlagert sie sich nach rechts, streut stärker und wird zunehmend symmetrisch.

Eine χ^2-Verteilung mit einem größer werdenden Freiheitsgrad nähert sich einer Normalverteilung mit N(*v*, $\sqrt{2v}$) an.

5.3 t-Verteilung



Die t-Verteilung dient u.a. zur Beurteilung der Unterschiede zweier Mittelwerte und zur Berechnung von Vertrauensgrenzen für Mittelwerte und Regressionskoeffizienten.

Der bei der Guinness™-Brauerei angestellte William Gosset fand eine neue Zufallsverteilung, die man nach seinem Pseudonym *student* die t-Verteilung nennt. T-Verteilungen spielen eine große Rolle bei Stichproben mit kleinem Umfang.

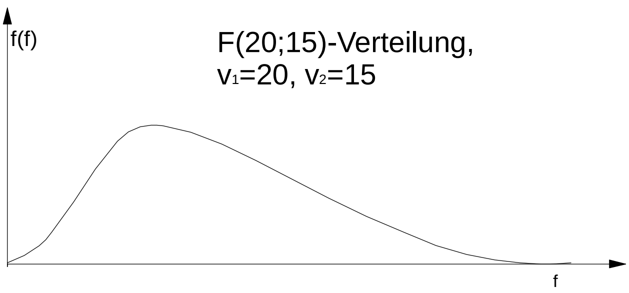
Die T-Variable entsteht als Quotient aus der Standardnormalvariablen K sowie dem Ausdruck $\sqrt{\frac{\chi^2}{v}}$ (*v* sind die Anzahl der Freiheitsgerade. D.h. $t = \frac{K}{\sqrt{\frac{\chi^2}{v}}}$

Die t-Verteilung ist stetig und unimodal. Die t-Variable ist symmetrisch um den Wert 0 verteilt. Modus, Median und Erwartungswert fallen zusammen. Der Wertebereich der T-Verteilung reicht wie der der Normalverteilung von +∞ bis -∞.
Mit größer werdendem Freiheitsgrad (n geht gegen ∞) geht die t-Verteilung über in die Standardnormalverteilung. Normalerweise ist die t-Verteilung gegenüber der Normalverteilung schmaler, dies nimmt allerdings mit zunehmenden Freiheitsgrad ab.

Tabellierung: Die Tabelle auf S. 31 in der Formelsammlung ist für zweiseitiges SN tabelliert. Bei einseitigen Fragestellungen muß mit $\frac{\alpha}{2}$ gearbeitet werden.

5.4 F-Verteilung

R.A.Fisher (1890 bis 1962) entdeckte die F-Verteilung, die z.B. dazu benutzt wird, die Gleichheit zweier Variablen zu testen. Die F-Verteilung ist als Quotient zweier unabhängiger Zufallsvariablen definiert, sie besitzt zwei Parameter: die Zahl der Freiheitsgrade v_1 und v_2 der beiden χ^2-Variablen.

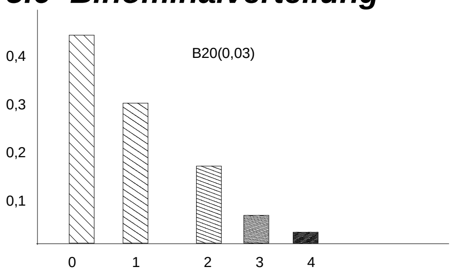


Die Variable kann nur positive Werte realisieren. Sie ist für kleinere Werte von v_1 und v_2 linkssteil, für größere Werte von v_1 und v_2 eine gegen die Normalverteilung konvergierende Zufallsverteilung mit dem Wertebereich [0,∞], d.h. der Wertebereich der F-Verteilung ist **0<f(F)**, der Definitionsbereich *0≤F≤∞*.

5.5 Gamma-Verteilung

- dient zur Berechnung von Flächen unter einer Kurve, deren Summe ungleich 1 ist, hat somit keine Wahrscheinlichkeitsfunktion
- $f(x) = e^x$
- Die Gamma-Verteilung beruht auf der Gamma-Funktion. Sie ermöglicht die Fakultätsberechnung von nicht ganzzahligen Werten
- die Funktion ist abhängig von λ (dem gedachten Erwartungswert) = Häufigkeit des erwarteten Ereignisses
- $\Gamma(\lambda) = Fläche(\lambda - 1)$

5.6 Binominalverteilung



Die Binominalverteilung ist eine der bedeutendsten <u>diskreten</u> Wahrscheinlichkeitsverteilungen.
Das zugrunde liegende Experiment (Bernoulli-Experiment) muß charakterisiert sein durch:

1. Es gibt zwei Ergebnisse: A und $\bar{A}$ [entweder trifft A oder $\bar{A}$ ein]
2. Man kennt die Wahrscheinlichkeit für P(A)=π und P(a)=1-π
3. Es werden *n* Experimente durchgeführt. Die Wahrscheinlichkeit muß bei jedem Einzelversuch gleich groß sein (⇒ konstante Einzelwahrscheinlichkeiten)
4. Die Ergebnisse der Versuche/Einzelexperimente sollen sich nicht gegenseitig beeinflussen, d.h. die Einzelversuchen sollen <u>unabhängig</u> voneinander sein.
   Die Bernoulli-Anordnung ist eine Anordnung mit <u>Zurücklegen</u>

Frage: Wie groß ist die Wahrscheinlichkeit, daß A in *n* Versuchen xi-mal i=1... n+1 realisiert wird? ($0 \leq x_i \leq n$)

n: Anzahl der Versuche/Einzelexperimente A: Ereignisalternative π: Wahrscheinlichkeit für A X_i: Anzahl (wie oft wird A realisiert?)

Durch $B(n;\pi) \hat{=} B(x_i, |n;\pi)$ werden Erfolgswahrscheinlichkeiten ausgerechnet, das heißt, das Eintreten von A wird als Erfolg gewertet.

5.7 Multinominalverteilung

Die Multinominalverteilung ist eine Verallgemeinerung der Binomialverteilung. Das zugrundeliegende Zufallsexperiment besteht aus Ziehen mit Zurücklegen aus einer Trommel, in der sich *mehr als zwei* Kategorien Kugeln befinden. Die Anzahl der Kategorien bestimmt die Dimension der Multinominalvariablen.
Eine Trommel mit *drei* Kugeln liefert eine zweidimensionale Multinominalvariable.
Eine Trommel mit *zwei* Kugeln liefert eine eindimensionale Multinominalvariable, eine Binominalverteilung.

6 Effizienz und Testgüte

6.1 Testgüte (auch Trennschärfe, Mächtigkeit)

Unter der Güte eines Tests versteht man die Wahrscheinlichkeit, keinen Fehler der 2. Art zu begehen, also eine unkorrekte Nullhypothese auch als falsch zu erkennen. Die Höhe der Wahrscheinlichkeit (1-β) bestimmt demnach die Güte (oder Trennschärfe) des Testes und folglich wird ein möglichst kleiner β-Fehler angestrebt. (vergleiche auch α- und β-Fehler, S. 7). Der β-Fehler tritt auf, da statistische Testverfahren auf die Widerlegung der Nullhypothese ausgerichtet sind.
Die Abhängigkeit der Güte (1-β) von dem Wert der Alternativhypothese wird als Gütefunktion bezeichnet, sie bildet bei einem zweiseitigem Test eine napfförmige Kurve. Die Gütefunktion weist jedem denkbaren Wert des Grundgesamtheitsparameters die Wahrscheinlichkeit für den Fehler 2. Art zugeordnet (Trennschärfefunktion).
Je höher die Voraussetzungen für ein Testverfahren sind, desto höher ist im allgemeinen auch die Güte des Tests. Ein Test, der Normalverteilungen und Streuungsleichheit wie z.B. der t-Test voraussetzt, hat im Vergleich zu verteilungsunabhängigen Tests, wie z.B. Schnelltests eine weitaus größere Trennschärfe.

6.2 Effizienz eines Tests

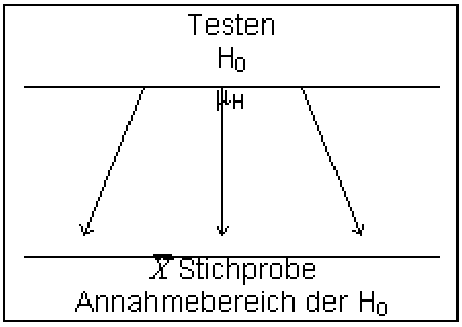
Bei einigen Prüfsituationen können verschiedene Testverfahren konkurrieren (z.B. der Mediantest gegen den Vorzeichentest). Es ist also z.B. möglich, daß ein Verfahren gewählt wird, das nur wenige Voraussetzungen benötigt, das aber bezüglich der Testgüte (Vermeidung des β-Fehlers) eingeschränkt ist.
Die Effizienz macht nun einen qualitativen Vergleich der konkurrierenden Tests möglich. Die finite relative Effizienz wird bemessen durch den Quotienten der Stichprobenumfänge $\frac{n_1}{n_2}$.

n_1 ist der Stichprobenumfang des „besseren“ (z.B. verteilungsgebundenen) Tests, n_2 ist der Umfang des „einfacheren“ (verteilungsfreien) Tests.
Beträgt der Quotient nun beispielsweise n_1=90%, so bedeutet dies, daß die Stichprobe II einem Umfang von n_2=100% hat, der Stichproben I jedoch nur n_1=90% von n_2 genügt, um ein gleichwertiges Ergebnis zu erreichen. So ist das Prüfverfahren I nicht nur ebenbürtig, sondern durch das geringere n sogar billiger und damit Prüfverfahren II vorzuziehen.
Sind die Quotienten jedoch über 90%, sind die „einfachen“ Tests jedoch nur unwesentlich teurer und ungenauer.

interessierender Test	Effizienz	konkurrierender Test
Mediantest	ca. 95% (bei kleiner Stichprobe) ca. 63 % (bei großer Stichprobe)	t-test
U-Test	ca. 95%	t-Test
H-Test	ca. 95%	F-Test

R_{sp}	ca. 91%	R (nach Beavis/Pearson)

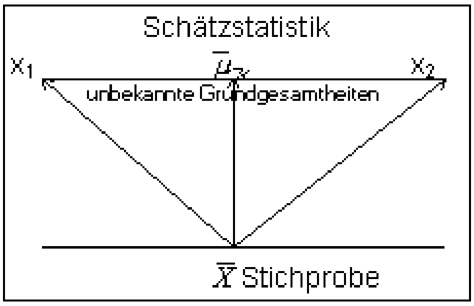
7 Schätzen



Schätzen nennt man das Festlegen von Werten für unbekannte Grundgesamtheiten mittels einer Stichprobe. Unbekannt und gesucht sind Parameter einer Grundgesamtheit, bekannt sind lediglich Stichprobenmaßzahlen, von denen auf entsprechende Grundgesamtheiten geschlossen werden soll.

Die **Punktschätzung** löst Probleme der Art: Der unbekannte Parameter der Grundgesamtheit hat den Wert X. Wie groß ist X? Eine Punktschätzung berechnet also aus einer Zufallsstichprobe ein einziger Schätzwert für den unbekannten Parameter berechnet. Unbekannt bleibt, um welchen Betrag der erhaltene Schätzwert von dem gesuchten Parameter abweicht. Die Bestimmungsvorschrift für den unbekannten Parameter nennt man Schätzer oder Schätzfunktion, sie ist wiederum eine Zufallsvariable. Als Punktschätzverfahren sind vor allem die relativ anspruchslose Momentenmethode und die Maximum-Likelihood-Schätzung zu nennen. Die Momentenmethode macht keinerlei Verteilungsvorgaben über die Grundgesamtheit und schätzt die Momente der Grundgesamtheit mit den Momenten der Stichprobe ab.
Die Maximum-Likelihood-Schätzung setzt die Bekanntheit des Verteilungstyps der Grundgesamtheit voraus und bestimmt diejenigen Werte als Schätzwerte für die unbekannten Parameter, die dem erhaltenen Stichprobenresultat die größte Wahrscheinlichkeit des Auftretens verleihen.

Intervallschätzung: Der unbekannte Parameter der Grundgesamtheit liegt (mit einem Vertrauen von y%) im Bereich x_1 bis x_2, wobei der punktgeschätzte Wert X - der in der Mitte des Intervalls liegt - derjenige ist, der dem wahren Wert vermutlich am nächsten kommt. Die konstruierten Intervalle nennt man Konfidenzintervalle oder Vertrauensbereiche. Die Bereichsgrenzen sind Realisationen von Zufallsvariablen, sie werden aus der Stichprobe berechnet und bilden ein Zufallsintervall.



7.1 *Maximum-Likelihood-Methode*

Die Maximum-Likelihood-Methode ist eine Methode, um „Schätzer" zu finden (wie auch die Methode der kleinsten Quadrate). Der Grundgedanke der Maximum-Likelihood-Methode ist: Man sollte den oder die GG-Parameter so schätzen, daß die entsprechende Stichproben-Realisation die größte Chance hat, realisiert zu werden. Im Ggs zwischen Momentenmethode und zum Intervallverfahren müssen Informationen über die Verteilung vorliegen (→ ein häufiges Problem).
Die Maximum-Likelihood-Methode geht auf die sogenannte Likelihoodfunktion zurück. Sie ist das Produkt aus n identischen Dichte- oder Wahrscheinlichkeitsfunktionen, gibt also die gemeinsame Dichte bzw. Wahrscheinlichkeit von n identischen und unabhängigen Zufallsvariablen an.
Die Zielfunktion lautet: $L(\hat{y}|x_i) \Rightarrow \max$

Der Grundgedanke der Maximum-Likelyhood-Methode läßt sich so beschreiben: Wenn wir einen theoretischen Parameter, z.B. μ, jenen Wert geben, den der aus der ML-Methode hervorgegangene Schätzer für eine konkret vorliegende Stichprobe annimmt, dann haben wir die Vorstellung mit genau jenem Parameterwert herausgefunden, für die das vorgefundene Stichprobenergebnis das „plausibelste" ist. Es ist plausibler, „vernünftiger" oder eben naheliegender, der GG-Verteilung diesen Parameterwert zuzuschreiben als irgeneinen anderen. Die höchste Vermutung („Mutmaßlichkeit") spricht für diesen Parameterwert.

7.2 *Güte von Punktschätzungen*

Für die Güte der Punktschätzungen werden im allgemeinen vier Kriterien angegeben:

1. Die Erwartungstreue
   Der Erwartungswert der Schätzfunktion sollte möglichst gleich dem wahren Parameterwert sein. Dies ist auch in vielen Fällen, wie z.B. bei der Punktschätzung des arithm. Mittels so, andererseits

lassen sich auch Beispiele anführen, wie z.B. die Punktschätzung der Varianz der Grundgesamtheit über die Varianz in der Stichprobe, bei denen diese Erwartungstreue nur asymptotisch, bei großen Stichprobenumfängen gilt.

2. die Effizienz
   eine Schätzfunktion sollte möglichst eine geringe Varianz aufweisen, ein Vergleich zweier erwartungstreuer Schätzfunktionen bezüglich dieses Kriteriums ermöglicht es, die effizientere zu wählen, ihre Einzelschätzungen liegen dann durchschnittlich näher am wahren Parameterwert als bei der Schätzfunktion mit der größeren Varianz. Die Präferenz für die Maximum-Likelihood-Methode gegenüber der Momentenmethode wird des öfteren mit diesem Kriterium begründet.
3. die Konsistenz
   Die Varianz der Stichprobenfunktion sollte möglichst gemäß dem Gesetz der Großen Zahl, mit zunehmenden Stichprobenumfang abnehmen, d.h. der Schätzwert sollte mit steigendem Stichprobenumfang dem Grundgesamtheitsparameterwert beliebig nahe kommen und somit zu ihm stochastisch konvergieren. Anzumerken wäre, daß Punktschätzungen nach der Momentenmethode stets konsistent sind, während das für die ML-Methode nur für die Mehrzahl der Fälle gilt.
4. die Suffizienz und Robustheit
   In der Literatur wird selten auch dieses letzte Kriterium angegeben. Das Kriterium besagt, daß die Schätzfunktion möglichst alle Informationen in der Stichprobe über den Parameter nutzen, und gegenüber Abweichungen und Extremwerten eine gewisse Robustheit vorliegen sollte. So ist z.B. der Median gegenüber Extremwerten robuster als das arithm. Mittel.

7.3 Konfidenzbereiche=Konfidenzintervalle=Vertrauensbereiche

Man versteht unter dem Begriff Vertrauensbereich ein aus Stichprobenwerten berechnetes Intervall, welches den wahren, aber unbekannten Parameter mit einer vorgegebenen Wahrscheinlichkeit überdeckt.

Beispiel: das arithmetische Mittel

Ausgehend von einem Repräsentationsschluß, also von einem Schluß von dem Schätzwert auf den Grundgesamtheitsparameter, ist der aus einer Stichprobe ermittelte Schätzwert $\bar{x}$ nur eine Schätzung des Mittelwertes μ der Grundgesamtheit, d.h. daß bei verschiedenen Stichproben die ermittelten Schätzwerte im allgemeinen variieren. Nun läßt sich allerdings ein Intervall angeben, daß mit einer bestimmten Wahrscheinlichkeit auch den Parameter der Grundgesamtheit enthält. Wählt man also einen Vertrauensbereich von 90%, so enthält der Vertrauensbereich zu P=90% den Parameter der Grundgesamtheit, in 10% der Fälle wird er außerhalb liegen. Letztere Wahrscheinlichkeiten bezeichnet man auch als Irrtumswahrscheinlichkeit α, P und α ergänzen sich zu 1. Wesentlich ist nun, daß ein Gegensatz zwischen Schärfe der Aussage und der Sicherheit der Aussage besteh: ein hoher Wert für P führt zu sicheren, aber unscharfen Aussagen, d.h. der Vertrauensbereich ist sehr breit, ein kleiner Wert für P für zu unsicheren, aber scharfen Aussagen. Grundsätzlich werden Vertrauensbereiche enger, wenn also einerseits die Irrtumswahrscheinlichkeit, aber auch andererseits der Stichprobenumfang erhöht wird.