Teil II: Mathematische Statistik

1 Einführung

Die Statistik stellt Methoden zur zahlenmäßigen Erfassung und Untersuchung von Datenmaterial bereit. Das Wort Statistik leitet sich von dem lateinischen Wort status (Zustand, Staat) ab, mit dem ursprünglich die Beschreibung des Staates, seiner Bevölkerung, Gewerbe usw. gemeint war.

Die **Statistik** wird unterteilt in die **deskriptive Statistik**, die sich mit der Erhebung, Strukturierung und Darstellung des Datenmaterials befasst, und die **induktive Statistik**, die das Datenmaterial mit wahrscheinlichkeitstheoretischen Methoden untersucht.

Die Entwicklung der Statistik als Wissenschaft begann am Anfang des 20. Jahrhunderts in England. Wichtige Beiträge zur Entwicklung der Statistik wurden z. B. von den folgenden Personen geleistet:

K. Pearson (1857 - 1936) schuf wesentliche Grundlagen für die Statistik im 20. Jahrhundert. Er trug zur Weiterentwicklung der Biometrie bei und beschäftigte sich unter anderem mit der χ^2-Verteilung, dem χ^2-Anpassungstest und der Klassifikation von Verteilungen.

W. S. Gosset (1876 - 1935), der unter dem Pseudonym Student veröffentlichte, führte die t-Verteilung ein.

R. A. Fisher (1890 - 1962) gilt als bedeutendster Statistiker des 20. Jahrhunderts. Er leistete maßgebliche Beiträge auf nahezu allen Gebieten der Statistik. Sein Name ist insbesondere verknüpft mit der Varianzanalyse und der Maximum-Likelihood-Methode.

E. S. Pearson (1895 - 1980) begründete zusammen mit **J. S. Neyman** (1894 - 1981) die moderne Testtheorie.

A. Wald ist Schöpfer der statistischen Entscheidungstheorie und leistete wichtige Beiträge zur Sequentialanalyse.

2 Deskriptive Statistik

2.1 Betrachtung eines Merkmals von Untersuchungsobjekten

Man unterscheidet die folgenden Merkmalstypen:

- qualitative Merkmale (z.B. Familienstand, Studienrichtung),
- Rangmerkmale (z.B. Interesse am Vorlesungsgegenstand) und
- quantitative Merkmale, die entweder diskret (z.B. Anzahl der Geschwister) oder stetig (z.B. Körpergröße) sein können.

Falls qualitative Merkmale bzw. Rangmerkmale durch Zahlen repräsentiert werden, spricht man von **nominalskalierten** Variablen bzw. **ordinalskalierten** Variablen. Quantitative Merkmale werden mit Hilfe einer **metrischen Skala** beschrieben.

Im Folgenden werden im Wesentlichen quantitative Merkmale betrachtet.

Ausgangspunkt der Betrachtungen sind n Messwerte $x_1, \ldots, x_n$, die sogenannte **Urliste**.
Bei der Darstellung und Untersuchung der Daten werden häufig die folgenden Vorgehensweisen angewandt:

- **Ordnen nach der Größe** liefert die im Folgenden mit $x_1^*, \ldots, x_n^*$ bezeichnete Darstellung, die auch **Variationsreihe** genannt wird ($x_i^* \leq x_{i+1}^*,\ i = 1, \ldots, n-1$).
- Methoden zum **Verdichten** der Daten sind zum Beispiel das Aufstellen einer **Häufigkeitstabelle** oder eines **Stem-Leaf-Display**s.

  Der Häufigkeitstabelle liegt in der Regel eine Klasseneinteilung zugrunde. Bei der Klasseneinteilung kann man sich nach den folgenden Empfehlungen richten:
  Für die Anzahl k der Klassen sollte $6 \leq k \leq 20$ oder $k \leq 5 \cdot \lg n$ gelten;
  die Klassenbreite sollte möglichst konstant sein;
  die Klassen „im Kern“ der Tabelle sollten besetzt sein.

  Hinweis: Werden statistische Maßzahlen nach einer Klasseneinteilung ermittelt, rechnet man mit den „Klassenmittelpunkten“.

- Als **graphische Darstellungen** werden häufig das **Stabdiagramm** bei einem diskreten Merkmal oder das **Histogramm** bei einem stetigen Merkmal (in der Regel nach Klasseneinteilung mit konstanter Klassenbreite) verwendet.

  Falls für die „Höhe“ der einzelnen Säulen im Histogramm die Werte

  $$h = \frac{\text{relative Häufigkeit der Klasse}}{\text{Klassenbreite}}$$

  gewählt wird, kann das Histogramm als „Schätzung“ für die Dichte benutzt werden.

- Die **empirische Verteilungsfunktion** ist ein wichtiges Hilfsmittel zur Beurteilung des Verteilungstyps. Sie wird folgendermaßen gebildet:

  $$\tilde{F}_n(x) := \begin{cases} 0 & \text{für} \quad x < x_1, \\ \frac{m}{n} & \text{für} \quad x_m^* \leq x < x_{m+1}^*, \quad m = 1, 2, \ldots, n-1, \\ 1 & \text{für} \quad x \geq x_n^*. \end{cases}$$

  Durch die Einzelwahrscheinlichkeiten $P(X = x_i) = \frac{1}{n},\ i = 1, \ldots, n,$ wird die **empirische Verteilung** bestimmt.

- Die folgenden **empirische**n **Kennwerte** liefern Informationen über Lage und Gestalt der Verteilung, die den Daten zugrunde liegt:

  α) **Lageparameter** sind das

  - **arithmetische Mittel** $\bar{x}_n = \frac{1}{n} \sum\limits_{i=1}^{n} x_i$ und der
  - **empirische Median** $x_{0,5} = \begin{cases} x_{k+1}^* & \text{für} \quad n = 2k+1, \\ \frac{x_k^* + x_{k+1}^*}{2} & \text{für} \quad n = 2k. \end{cases}$

  β) Ein wichtiges **Streuungsmaß** ist die

  - **empirische Varianz** $s_n^2 = \frac{1}{n-1} \sum\limits_{i=1}^{n} (x_i - \bar{x}_n)^2 = \frac{1}{n-1} \left[\sum\limits_{i=1}^{n} x_i^2 - \frac{1}{n} (\sum\limits_{j=1}^{n} x_j)^2 \right].$

  Die Richtigkeit der Gleichung zeigt die folgende Rechnung:

  $$\begin{aligned} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 &= \sum_{i=1}^{n} (x_i^2 - 2x_i \tfrac{1}{n} \sum_{j=1}^{n} x_j + \tfrac{1}{n^2} \left(\sum_{j=1}^{n} x_j \right)^2) \\ &= \sum_{i=1}^{n} x_i^2 - \tfrac{2}{n} \sum_{i=1}^{n} x_i \sum_{j=1}^{n} x_j + \tfrac{n}{n^2} \left(\sum_{j=1}^{n} x_j \right)^2. \end{aligned}$$

$s_n > 0$ heißt **empirische Standardabweichung**.

Weitere Streuungsmaße sind die

- **Spannweite** $x_n^* - x_1^*$ und der
- **empirische Variationskoeffizient** $\frac{s_n}{\bar{x}_n}$.

γ) Die **empirischen Momente** sind in folgender Weise erklärt:

$\hat{m}_k := \frac{1}{n} \sum\limits_{i=1}^{n} x_i^k$ ist das empirische (gewöhnliche) Moment k-ter Ordnung,
$\hat{\mu}_k := \frac{1}{n} \sum\limits_{i=1}^{n} (x_i - \bar{x}_n)^k$ ist das empirische zentrale Moment k-ter Ordnung.

- Box-Whisker-Plots, kurz **Box-Plots**, dienen der übersichtlichen Darstellung eines Datensatzes mittels weniger Kenngrößen. In der Regel werden die folgenden 5 Größen zur Darstellung verwendet:
  - kleinster Wert der Beobachtungsreihe x_1^*,
  - empirisches unteres Quartil (= Quantil der Ordnung 0,25) $x_{0,25}$,
  - empirischer Median $x_{0,5}$,
  - empirisches oberes Quartil (= Quantil der Ordnung 0,75) $x_{0,75}$,
  - größter Wert der Beobachtungsreihe x_n^*.

  Bemerkungen:

  1.) Die empirischen Quantile können z.B. nach der folgenden allgemeinen Vorschrift ermittelt werden. (In Spezialfällen hat man bessere Möglichkeiten, eine wurde oben für den Median angegeben.)

  $$\mathrm{x}_p = \begin{cases} x_{np}^*, & \text{falls } np \text{ ganzzahlig,} \\ x_{[np+1]}^* & \text{sonst.} \end{cases}$$

  2.) $x_{0,75} - x_{0,25}$ heißt **Quartilsabstand** des Datensatzes.

2.2 Betrachtung von zwei Merkmalen X, Y eines Untersuchungsobjektes

Ausgangspunkt der Betrachtungen sind Messwertpaare (x_i, y_i) $i = 1, 2, \ldots, n$. Die folgenden Darstellungen und Kenngrößen können (abhängig von der jeweiligen Aufgabenstellung) wichtige Aufschlüsse für das weitere Vorgehen liefern:

- **Graphische Darstellung im $(x, y)-$Koordinatensystem**
- (bivariate) **Kontingenztafel** (auch Kreuztabelle)

  Es handelt sich dabei um eine bivariate Häufigkeitstabelle, die häufig nach einer Klasseneinteilung aufgestellt wird.

	$b_1 \dots b_2$	$b_2 \dots b_3$	$\dots$	$b_{k_y} \dots b_{k_y+1}$	$\sum$
$a_1 \dots a_2$	h_{11}	h_{12}	$\dots$	h_{1k_y}	$h_{1\cdot}$
$a_2 \dots a_3$	h_{21}	h_{22}	$\dots$	h_{2k_y}	$h_{2\cdot}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$a_{k_x} \dots a_{k_x+1}$	h_{k_x1}	h_{k_x2}	$\dots$	$h_{k_xk_y}$	$h_{k_x\cdot}$
$\sum$	$h_{\cdot 1}$	$h_{\cdot 2}$	$\dots$	$h_{\cdot k_y}$	n

  Dabei bezeichnet h_{ij} die Anzahl der Messwertpaare (x, y) mit $a_i < x \leq a_{i+1}$ und $b_j < y \leq b_{j+1}$.

- **Statistische Maßzahlen**

  α) Die Größen $\bar{x}, \bar{y}, s_x^2, s_y^2$ werden wie im vorhergehenden Abschnitt aus den Werten $x_1, \dots, x_n$ bzw. $y_1, \dots, y_n$ ermittelt.

  β) Zur Beurteilung des Abhängigkeitsverhaltens spielen die **empirische Kovarianz**

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} [(x_i - \bar{x}_n)(y_i - \bar{y}_n)] = \frac{1}{n-1} \left[\sum_{i=1}^{n} x_i y_i - \frac{1}{n} \sum_{j=1}^{n} x_j \sum_{k=1}^{n} y_k \right]$$

  und abgeleitete Größen wie der empirische Korrelationskoeffizient eine wichtige Rolle.

  γ) Der **empirischer Korrelationskoeffizient** (nach Bravais-Pearson) wird nach der Vorschrift

$$r_{xy} := \frac{s_{xy}}{s_x s_y}$$

  ermittelt.

  Der empirische Korrelationskoeffizient kann durch „Ausreißer“ stark verfälscht werden. Eine von Ausreißern weniger stark beeinflusste Variante ist der Rangkorrelationskoeffizient von Spearman, der insbesondere beim Vorliegen von Rangmerkmalen benutzt wird.

δ) Der **Rangkorrelationskoeffizient** von Spearman ist der für die getrennt bestimmten Rangwerte $r(x_i), r(y_i)$ ermittelte empirische Korrelationskoeffizient:

$$r_s = \frac{\sum_{i=1}^{n}(r(x_i)-\bar{r})(r(y_i)-\bar{r})}{\sqrt{(\sum_{i=1}^{n}(r(x_i)-\bar{r})^2)\cdot(\sum_{j=1}^{n}(r(y_j)-\bar{r})^2)}}, \qquad \text{wobei}$$

$\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r(x_i) = \frac{1}{n}\sum_{i=1}^{n} r(y_i) = \frac{n+1}{2}$ gilt.

Bemerkungen:

1.) Kommen bei der Ermittlung der Rangzahlen gleiche Werte (Bindungen) vor, werden mittlere Ränge vergeben.

2.) Sind alle x-Werte und alle y-Werte unterschiedlich, vereinfacht sich r_s zu

$$r_s = 1 - \frac{6\sum_{i=1}^{n}(r(x_i)-r(y_i))^2}{n(n^2-1)}.$$

Sind ***p*-Tupel von Messwerten** gegeben, wird häufig eine paarweise Untersuchung der einzelnen Merkmale durchgeführt, oder es kommen clusteranalytische oder faktoranalytische Methoden zur Verdichtung der Daten und Reduktion der Dimension zum Einsatz.

3 Grundbegriffe der Induktiven Statistik

3.1 Grundgesamtheit und Stichprobe

Zielstellung der folgenden Überlegungen ist die Schaffung eines Modells zur Einbettung der mathematischen Statistik in die Wahrscheinlichkeitstheorie.

Betrachtet wird ein Merkmal X. Als zugehörige **Grundgesamtheit** wollen wir die Menge aller für die statistische Untersuchung relevanten Merkmalsträger bezeichnen. Durch eine Vollerhebung der Daten könnte man den einzelnen Merkmalswerten ihre relativen Häufigkeiten zuordnen und so eine diskrete Wahrscheinlichkeitsverteilung erhalten und ggf. die erhaltene Verteilung durch eine stetige Wahrscheinlichkeitsverteilung approximieren.

Da eine Vollerhebung häufig nicht möglich oder sinnvoll ist, wird aus der Grundgesamtheit eine Stichprobe gezogen mit dem Ziel, Rückschlüsse auf die Verteilung der Merkmalswerte der Grundgesamtheit zu ziehen. Die zufällige Auswahl eines Merkmalsträgers kann als Zufallsexperiment aufgefasst werden. Man kann das Merkmal X daher identifizieren mit einer Zufallsgröße, deren Verteilung gerade durch das beschriebene Wahrscheinlichkeitsmaß gegeben ist.

Bemerkung: Gelegentlich wird als Grundgesamtheit auch die Menge aller Merkmalswerte für die relevanten Merkmalsträger bezeichnet.

Die n-malige unabhängige und zufällige Auswahl eines Merkmalsträgers wird dann (unter Vernachlässigung des Sachverhaltes, dass der gleiche Merkmalsträger nicht mehrfach ausgewählt werden kann) aufgefasst als Realisierung eines Zufallsvektors, dessen Komponenten unabhängig und identisch wie X verteilt sind.

Definition:

> Unter einer **mathematischen Stichprobe** vom Umfang n aus der zu X gehörenden Grundgesamtheit versteht man n unabhängige, identisch wie X verteilte Zufallsgrößen, die zu einem Zufallsvektor zusammengefasst werden: $(X_1, \ldots, X_n)$. Jede Realisierung $(x_1, \ldots, x_n)$ einer mathematischen Stichprobe ist eine **konkrete Stichprobe**.

Analog werden eine zu (X, Y) gehörende Grundgesamtheit und „zwei verbundene Stichproben" $((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))$ definiert. (Die beim Zufallsvektor verwendete Schreibweise als Spaltenvektor wird hier vernachlässigt.)

Definition:

> Es sei $(X_1, \ldots, X_n)$ eine Stichprobe. Eine Funktion $f(X_1, \ldots, X_n)$ der Stichprobenvariablen $X_1, \ldots, X_n$, die wieder eine Zufallsgröße ist, wird **Stichprobenfunktion** oder **Statistik** genannt.

Bemerkung: Die Stichprobenfunktion oder Statistik ist hier eine Zufallsgröße $f(X_1, \ldots, X_n)| \; \Omega \to R^1$. Gelegentlich wird nur die Abbildung $f| \; R^n \to R^1$ (die geeignete Messbarkeitsvoraussetzungen erfüllen muss) als Stichprobenfunktion bezeichnet.

Im Folgenden sollen **Beispiele für Stichprobenfunktionen** näher betrachtet werden. Dabei sei stets $(X_1, \ldots, X_n)$ eine Stichprobe aus einer zu X

gehörenden Grundgesamtheit mit $\mathrm{E}(X) = \mu$ und $\mathrm{var}(X) = \sigma^2$. Aus diesen Annahmen folgt, dass auch $\mathrm{E}(X_i) = \mu$ und $\mathrm{var}(X_i) = \sigma^2$) gilt.

a) Für das **Stichprobenmittel** $\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ ergibt sich

$$\mathrm{E}(\bar{X}_n) = \mathrm{E}(\tfrac{1}{n}\sum_{i=1}^{n} X_i) = \tfrac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i) = \mu,$$

$$\mathrm{var}(\bar{X}_n) = \mathrm{var}(\tfrac{1}{n}\sum_{i=1}^{n} X_i) = \tfrac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}(X_i) = \frac{\sigma^2}{n}.$$

Nach dem Starken Gesetz der Großen Zahlen von Kolmogorov gilt weiterhin
$P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1.$

b) Für die **Stichprobenvarianz** $S_n^2 := \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ (die man verwendet, falls $\mathrm{E}(X_i)$ unbekannt ist) gilt

$$\begin{aligned}
\mathrm{E}(S_n^2) &= \mathrm{E}(\tfrac{1}{n-1}\sum_{i=1}^{n}((X_i-\mu) - \tfrac{1}{n}\sum_{j=1}^{n}(X_j-\mu))^2] \\
&= \mathrm{E}(\tfrac{1}{n-1}\{\sum_{i=1}^{n}(X_i-\mu)^2 - \tfrac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-\mu)(X_j-\mu)+ \\
&\quad +\tfrac{n}{n^2}\sum_{j=1}^{n}\sum_{k=1}^{n}(X_j-\mu)(X_k-\mu)\}] \\
&= \mathrm{E}(\tfrac{1}{n-1}\{\sum_{i=1}^{n}(X_i-\mu)^2(1-\tfrac{1}{n}) - \tfrac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}(X_i-\mu)(X_j-\mu)\}] \\
&= \tfrac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i-\mu)^2 = \sigma^2.
\end{aligned}$$

Mit Hilfe des Starken Gesetzes der Großen Zahlen von Kolmogorov kann auch
$P(\lim_{n\to\infty} S_n^2 = \sigma^2) = 1$ gezeigt werden.

Bemerkung: Ist μ bekannt, verwendet man als „Schätzung" für σ^2 die Stichprobenfunktion $S_0^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$.

3.2 Punktschätzungen

Im Folgenden wird angenommen, dass die Verteilung P_X eines Merkmales X bis auf einen Parameter ϑ bekannt ist, d.h., es gelte $P_X \in \{P_\vartheta : \vartheta \in \Theta \subset R^k\}$. Der Parameter ϑ oder der Wert der Funktion $\tau \mid \Theta \to R^1$ an der

Stelle ϑ soll geschätzt werden.

Definition:

> Eine **Punktschätzung** (auch Schätzfunktion, Schätzer) für $\tau(\vartheta)$ ist eine Stichprobenfunktion $T_n(X_1, \dots, X_n)$.
> Der Wert $T_n(x_1, \dots, x_n)$ heißt **konkrete Punktschätzung** oder **Schätzwert** für $\tau(\vartheta)$.

Beispiele:

a) $\hat{\mu} := \bar{X}_n$ ist eine Punktschätzung für $\mathrm{E}(X)$.

b) S_n^2 ist eine Punktschätzung für $\mathrm{var}(X)$.

c) X sei gleichmäßig stetig verteilt im Intervall $[0, a]$, $a \in (0, \infty)$. Dann ist $\hat{a} := \max(X_1, \dots, X_n)$ eine Punktschätzung für a.

d) $\hat{\varrho}_{XY} := \dfrac{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sqrt{S_{n,X}^2 \cdot S_{n,Y}^2}}$ ist eine Punktschätzung für den Korrelationskoeffizienten ϱ_{XY}. $S_{n,X}^2$ und $S_{n,Y}^2$ werden dabei wie S_n^2 aus den Stichproben
$(X_1, \dots, X_n)$ bzw. $(Y_1, \dots, Y_n)$ gebildet.

3.2.1 Eigenschaften von Punktschätzungen

Im Folgenden werden Gütekriterien für Punktschätzungen betrachtet.

a) **Erwartungstreue**

Definition:

Die Punktschätzung $T_n(X_1, \ldots, X_n)$ heißt **erwartungstreu** (unverzerrt) für $\tau(\vartheta)$, falls

$$\mathrm{E}_\vartheta[T_n(X_1, \ldots, X_n)] = \tau(\vartheta) \quad \forall \vartheta \in \Theta$$

gilt.
Eine Folge von Punktschätzungen $(T_n(X_1, \ldots, X_n))_{n \in N}$ heißt **asymptotisch erwartungstreu** für $\tau(\vartheta)$, wenn

$$\lim_{n \to \infty} \mathrm{E}_\vartheta[T_n(X_1, \ldots, X_n)] = \tau(\vartheta) \quad \forall \vartheta \in \Theta$$

gilt.

Beispiele:

- $\bar{X}_n$ ist erwartungstreu für $\mu = \mathrm{E}(X)$.
- S_n^2 ist erwartungstreu für $\sigma^2 = \mathrm{var}(X)$.
- $\hat{\varrho}_{XY}$ ist asymptotisch erwartungstreu für ϱ_{XY}.

b) **Konsistenz**

Definition:

Die Folge $(T_n(X_1, \ldots, X_n))_{n \in N}$ von Punktschätzungen für $\tau(\vartheta)$ heißt

α) **stark konsistent**, falls

$$P_\vartheta(\lim_{n \to \infty} T_n(X_1, \ldots, X_n) = \tau(\vartheta)) = 1 \quad \forall \vartheta \in \Theta$$

gilt,

β) **schwach konsistent**, falls

$$\forall \varepsilon > 0 \quad \lim_{n \to \infty} P_\vartheta(\mid T_n(X_1, \ldots, X_n) - \tau(\vartheta) \mid \geq \varepsilon) = 0 \quad \forall \vartheta \in \Theta$$

gilt.

Beispiele:

- $\bar{X}_n$ ist stark konsistent für $\mathrm{E}(X)$.
- S_n^2 ist stark konsistent für $\mathrm{var}(X)$.
- $\hat{\varrho}_{XY}$ ist stark konsistent für ϱ_{XY}.

Hinreichend für die schwache Konsistenz einer erwartungstreuen Punktschätzung T_n für $\tau(\vartheta)$ ist die Beziehung $\lim\limits_{n\to\infty} \mathrm{var}_\vartheta(T_n(X_1,\ldots,X_n)) = 0 \quad \forall\vartheta\in\Theta$, denn mit der Tschebyschevschen Ungleichung gilt

$$\begin{aligned} 0 &\le P_\vartheta(\mid T_n(X_1,\ldots,X_n) - \tau(\vartheta) \mid\ge \varepsilon) \\ &= P_\vartheta(\mid T_n(X_1,\ldots,X_n) - \mathrm{E}_\vartheta(T_n(X_1,\ldots,X_n)) \mid\ge \varepsilon) \\ &\le \frac{\mathrm{var}_\vartheta(T_n(X_1,\ldots,X_n))}{\varepsilon^2} \qquad \forall\vartheta\in\Theta. \end{aligned}$$

Sind für ein $\tau(\vartheta)$ mehrere erwartungstreue Punktschätzungen bekannt, benutzt man diejenige mit der kleinsten Varianz (wirksamste Schätzung).

3.2.2 Konstruktionsmethoden für Punktschätzungen

a) **Maximum-Likelihood-Methode** (R. A. Fischer)

Für die Anwendung der Maximum-Likelihood-Methode muss der Verteilungstyp der Grundgesamtheit bis auf einen Parameter $\vartheta\in\Theta\subset R^k$ bekannt sein.

Ausgehend von einer konkreten Stichprobe $(x_1,\ldots,x_n)$ wird die sogenannte **Likelihood-Funktion** L definiert:

$$L(x_1,\ldots,x_n;\vartheta) := \begin{cases} \prod\limits_{i=1}^{n} P_\vartheta(X=x_i), & \text{falls X diskret verteilt,} \\ \prod\limits_{i=1}^{n} f_{X,\vartheta}(x_i), & \text{falls X stetig verteilt.} \end{cases}$$

Als Schätzwert $\hat\vartheta$ für ϑ verwendet man einen Wert, für den die Likelihood-Funktion maximal wird. Gesucht ist also die Lösung der Optimierungsaufgabe

$$(*) \quad \max_{\vartheta\in\Theta} L(x_1,\ldots,x_n;\vartheta)$$

Besitzt $(*)$ eine Lösung $\hat\vartheta = T(x_1,\ldots,x_n)$, so heißt $T(x_1,\ldots,x_n)$ **Maximum-Likelihood-Schätzwert** für ϑ. Ersetzt man formal x_i durch X_i, so erhält man die **Maximum-Likelihood-Schätzung** $T(X_1,\ldots,X_n)$.
Bei der Lösung von $(*)$ macht man sich zunutze, dass L für genau die Werte ϑ_o maximal wird, für die auch $\ln(L)$ maximal wird.
Damit ergibt sich im Fall der Differenzierbarkeit als **notwendige Bedingung** für die Maximumstelle

$$\frac{\partial \ln L(x_1,\ldots,x_n;\vartheta)}{\partial\vartheta_i} = 0, \quad i=1,\ldots,k, \quad \left(\vartheta = \begin{pmatrix}\vartheta_1\\ \vdots \\ \vartheta_k\end{pmatrix}\right).$$

Beispiel:

Betrachtet wird ein Zufallsexperiment, bei dem das Ereignis A eintreten kann. $p = P(A)$ ist zu schätzen. Dazu werden Zufallsgrößen X_i eingeführt:

$$X_i = \begin{cases} 1, & \text{falls Ereignis A im i-ten Versuch eingetreten ist,} \\ 0 & \text{sonst.} \end{cases}$$

n Wiederholungen liefern dann eine Stichprobe $(X_1, \ldots, X_n)$ bzw. eine konkrete Stichprobe $(x_1, \ldots, x_n)$ mit $x_i \in \{0, 1\}$.

α) Es seien nicht alle x_i gleich Null oder gleich Eins. Dann muss $p \in (0, 1)$ gelten, und man kann - unter Verwendung der Abkürzung $s := \sum\limits_{i=1}^{n} x_i$ - wie folgt vorgehen:

$$L(x_1, \ldots, x_n; p) = \prod_{i=1}^{n} P(X = x_i) = p^s (1-p)^{n-s}$$

$$\ln L(x_1, \ldots, x_n; p) = s \cdot \ln p + (n-s) \cdot \ln(1-p)$$

$$\frac{\partial}{\partial p} \ln L(x_1, \ldots, x_n; p) = \frac{s}{p} - \frac{n-s}{1-p} \stackrel{!}{=} 0.$$

$p = \dfrac{s}{n} = \dfrac{\sum\limits_{i=1}^{n} x_i}{n}$ erfüllt daher die notwendige Optimalitätsbedingung. Man kann dann (z.B. durch Betrachtung des Verhaltens von L für $p \to 0$ und $p \to 1$) zeigen, dass an der Stelle $p = \frac{s}{n}$ tatsächlich ein Maximum vorliegt.

Die relative Häufigkeit ist also die Maximum-Likelihood-Schätzung für p: $\hat{p} = \frac{\sum\limits_{i=1}^{n} X_i}{n} = H_n(A)$.

β) Es gelte $x_i = 1,\ i = 1, \ldots, n$. Die Likelihoodfunktion hat dann die Gestalt $L(x_1, \ldots, x_n; p) = p^n$. Diese Funktion nimmt ihr Maximum für $p = 1$ an.

γ) Es gelte $x_i = 0,\ i = 1, \ldots, n$. Die Likelihoodfunktion hat dann die Gestalt $L(x_1, \ldots, x_n; p) = (1-p)^n$. Diese Funktion nimmt ihr Maximum für $p = 0$ an.

Auch in den Fällen β) und γ) kann die Lösung in der Form $p = \frac{\sum\limits_{i=1}^{n} x_i}{n}$ geschrieben werden, folglich ist die relative Häufigkeit die Maximum-Likelihood-Schätzung für p.

b) **Methode der kleinsten Quadrate** (Gauß)

Es wird keine Kenntnis des Verteilungstyps vorausgesetzt. Die Methode der kleinsten Quadrate spielt insbesondere eine Rolle bei Regressionsanalysen und Varianzanalysen. Hier soll das Vorgehen lediglich an einem Beispiel demonstriert werden.

Beispiel:

Der Schätzwert nach der Methode der kleinsten Quadrate für den Erwartungswert μ eines Merkmals ist definiert als Lösung der Optimierungsaufgabe
$\min\limits_{\mu \in R} \sum\limits_{i=1}^{n} (x_i - \mu)^2$.

Man erhält als notwendige Optimalitätsbedingung

$$\frac{\partial}{\partial \mu} \sum_{i=1}^{n} (x_i - \mu)^2 = -2 \sum_{i=1}^{n} (x_i - \mu) = -2 (\sum_{i=1}^{n} x_i - n\mu) \stackrel{!}{=} 0$$

und überzeugt sich leicht, dass bei $\mu = \bar{x}_n$ tatsächlich ein Minimum vorliegt. Die Schätzung nach der Methode der kleinsten Quadrate erhält man dann, indem man die Messwerte x_i durch die Stichprobenvariablen X_i ersetzt. Also ist $\hat{\mu} = \bar{X}_n$ die Schätzung nach der Methode der kleinsten Quadrate für den Erwartungswert.

c) **Momentenmethode**

Zu schätzende Parameter lassen sich oft als Funktion von Momenten darstellen. Man erhält Punktschätzungen, indem man die Momente $\mathrm{E}(X^k)$ durch die Stichprobenmomente $\frac{1}{n} \sum\limits_{i=1}^{n} (X_i)^k$ ersetzt.

3.3 Die Verteilung wichtiger Stichprobenfunktionen

Für die Angabe von Konfidenzschätzungen und die Durchführung von statistischen Tests wird die Kenntnis der Verteilung von gewissen Stichprobenfunktionen benötigt. Dabei wird in der Regel davon ausgegangen, dass die

Stichprobenvariablen X_i normalverteilt sind. Die folgenden Verteilungstypen ergeben sich als Verteilung wichtiger Stichprobenfunktionen.

3.3.1 Die χ^2-Verteilung (Helmert, K. Pearson)

Es seien $Z_1, \ldots, Z_n$ unabhängige, $N(0,1)$-verteilte Zufallsgrößen. Die Verteilung der Zufallsgröße

$$\sum\nolimits_n := Z_1^2 + \ldots + Z_n^2$$

heißt **χ^2 -Verteilung mit n Freiheitsgraden**.

Die χ^2-Verteilung besitzt die Dichtefunktion

$$f_{\sum_n}(x) = \begin{cases} \dfrac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}, & \text{falls} \quad x > 0, \\ 0, & \text{falls} \quad x \leq 0, \end{cases}$$

wobei $\Gamma(\alpha) := \int\limits_0^\infty t^{\alpha-1}e^{-t}dt$ gilt. Die **Gammafunktion** Γ besitzt die Eigenschaften

$$\Gamma(\alpha+1) = \alpha \cdot \Gamma(\alpha), \ \Gamma(1) = 1, \ \Gamma(\frac{1}{2}) = \sqrt{\pi}.$$

Es gilt:

a) Die Summe zweier unabhängiger, χ^2-verteilter Zufallsgrößen mit n bzw. m Freiheitsgraden ist χ^2-verteilt mit $n+m$ Freiheitsgraden.

b) Für große n ist $\dfrac{\sum_n - n}{\sqrt{2n}}$ näherungsweise $N(0,1)$-verteilt.

<u>Satz:</u>

Es sei $(X_1, \ldots, X_n)$ eine Stichprobe aus einer zu $X \sim N(\mu, \sigma^2)$ gehörenden Grundgesamtheit. Dann ist die Stichprobenfunktion
$$\frac{n-1}{\sigma^2}S_n^2 = \sum_{i=1}^n (\frac{X_i - \bar{X}_n}{\sigma})^2$$
χ^2-verteilt mit $n-1$ Freiheitsgraden.

3.3.2 Die t-Verteilung (Student)

X sei eine $N(0,1)$-verteilte Zufallsgröße, und Q sei χ^2-verteilt mit n Freiheitsgraden. X und Q seien unabhängig. Die Verteilung von

$$T_n := \frac{X}{\sqrt{\frac{Q}{n}}}$$

heißt **t-Verteilung mit n Freiheitsgraden**.

Die Dichtefunktion hat die Gestalt

$$f_{T_n}(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}(1 + \frac{x^2}{n})^{-\frac{n+1}{2}} \qquad -\infty < x < \infty.$$

Es gilt:

a) Die t-Verteilung mit n Freiheitsgraden besitzt Momente bis zur Ordnung $k \leq n-1$.

b) Für große n ist T_n näherungsweise $N(0,1)$-verteilt.

Satz:

> Es sei $(X_1, \ldots, X_n)$ eine Stichprobe aus einer zu $X \sim N(\mu, \sigma^2)$ gehörenden Grundgesamtheit. Dann sind die Zufallsgrößen $\bar{X}_n$ und S_n^2 unabhängig, und die Stichprobenfunktion $T_n := \dfrac{\bar{X}_n - \mu}{S_n}\sqrt{n}$ ist t-verteilt mit $n-1$ Freiheitsgraden.

3.3.3 Die F-Verteilung (Fisher)

Die Zufallsgrößen Y_m und $\tilde{Y}_n$ seien unabhängig und χ^2-verteilt mit m bzw. n Freiheitsgraden. Die Verteilung von

$$F := \frac{\frac{Y_m}{m}}{\frac{\tilde{Y}_n}{n}}$$

heißt **F-Verteilung mit (m, n) Freiheitsgraden**.

Sie besitzt die Dichtefunktion

$$f_F(x) = \begin{cases} \frac{(\frac{m}{n})^{\frac{m}{2}} \cdot x^{\frac{m}{2}-1}}{B(\frac{m}{2}, \frac{n}{2})}(1 + \frac{m}{n}x)^{-\frac{m+n}{2}}, & \text{falls } x > 0, \\ 0, & \text{falls } x \leq 0, \end{cases}$$

wobei $B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}dt$ und $B(p,q) = \dfrac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ gelten. B heißt **Betafunktion**.

Für das Quantil $F_{m,n;\alpha}$ der Ordnung α der F-Verteilung mit (m,n) Freiheitsgraden gilt $F_{m,n;\alpha} = \dfrac{1}{F_{n,m;1-\alpha}}$.

<u>Satz:</u>

> $(X_1, \cdots, X_m)$ sei eine Stichprobe aus einer zu $X \sim N(\mu_1, \sigma^2)$ gehörenden Grundgesamtheit, und $(Y_1, \ldots, Y_n)$ sei eine Stichprobe aus zu $Y \sim N(\mu_2, \sigma^2)$ gehörenden Grundgesamtheit. Beide Stichproben seien unabhängig. Dann besitzt die Stichprobenfunktion $\dfrac{S^2_{m,X}}{S^2_{n,Y}} = \dfrac{\frac{1}{m-1}\sum\limits_{i=1}^{m}(X_i - \bar{X}_m)^2}{\frac{1}{n-1}\sum\limits_{i=1}^{n}(Y_i - \bar{Y}_n)^2}$ eine F-Verteilung mit $(m-1, n-1)$ Freiheitsgraden.

3.4 Konfidenzschätzungen

In diesem Abschnitt sollen zufällige Bereiche, die den wahren Wert des Parameters ϑ (mindestens) mit einer vorgegebenen Wahrscheinlichkeit ε überdecken, hergeleitet werden. Falls $\vartheta \in R^1$ gilt, wird in der Regel ein Intervall angegeben.

Konfidenzschätzungen kann man aus Stichprobenfunktionen gewinnen, in denen der betrachtete Parameter ϑ vorkommt, deren Verteilung bekannt und unabhängig von ϑ ist und die sich nach ϑ auflösen lassen. Die folgenden Konfidenzintervalle können auf diese Weise gewonnen werden. Ausgangspunkt der Betrachtungen sind ein normalverteiltes Merkmal $X \sim N(\mu, \sigma^2)$ und eine Stichprobe $(X_1, \ldots, X_n)$.

a) **Konfidenzintervall für $\boldsymbol{\mu}$ bei bekanntem $\boldsymbol{\sigma^2}$** zum Konfidenzniveau $\varepsilon = 1 - \alpha$:

   Die Zufallsgröße $\tilde{Y}_n := \dfrac{\bar{X}_n - \mu}{\sigma}\sqrt{n}$ ist unter den vorausgesetzten Verteilungsannahmen $N(0,1)$-verteilt, und es gilt

   $$P(\lambda_{\frac{\alpha}{2}} \leq \tilde{Y}_n \leq \lambda_{1-\frac{\alpha}{2}}) = 1 - \alpha,$$

wobei λ_α das Quantil der Ordnung α der $N(0,1)$-Verteilung bezeichnet.

Weiterhin folgt

$$\begin{aligned}
1-\alpha &= P(\lambda_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \bar{X}_n - \mu \le \lambda_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}) \\
&= P(\bar{X}_n - \lambda_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X}_n - \lambda_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}) \\
&= P(\bar{X}_n - \lambda_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X}_n + \lambda_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}).
\end{aligned}$$

Damit ergibt sich das Konfidenzintervall als

$$I = [\bar{X}_n - \lambda_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \bar{X}_n + \lambda_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}].$$

b) **Konfidenzintervall für $\boldsymbol{\mu}$ bei unbekanntem $\boldsymbol{\sigma^2}$** zum Konfidenzniveau $\varepsilon := 1-\alpha$:

Aufgrund des Sachverhaltes, dass $T_n := \frac{\bar{X}_n - \mu}{S_n}\sqrt{n}$ eine t-Verteilung mit $n-1$ Freiheitsgraden besitzt, erhält man durch ähnliche Überlegungen wie im Fall a)

$$I = [\bar{X}_n - t_{n-1;1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}; \bar{X}_n + t_{n-1;1-\frac{\alpha}{2}}\frac{S_n}{\sqrt{n}}],$$

wobei $t_{n-1;1-\frac{\alpha}{2}}$ das Quantil der Ordnung $1-\frac{\alpha}{2}$ der t-Verteilung mit $n-1$ Freiheitsgraden bezeichnet.

c) **Konfidenzintervall für $\boldsymbol{\sigma^2}$** zum Konfidenzniveau $\varepsilon = 1-\alpha$:

Aufgrund des Sachverhaltes, dass $\frac{(n-1)S_n^2}{\sigma^2}$ eine χ^2-Verteilung mit $n-1$ Freiheitsgraden besitzt, erhält man

$$I = [\frac{(n-1)S_n^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}}; \frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{\alpha}{2}}}],$$

wobei $\chi^2_{n-1;\frac{\alpha}{2}}$ das Quantil der Ordnung $\frac{\alpha}{2}$ der χ^2-Verteilung mit $n-1$ Freiheitsgraden bezeichnet.

3.5 Tests

3.5.1 Grundbegriffe

Mit einem statistischen Test wird eine Annahme, die sogenannte Hypothese, über die vollständig oder teilweise unbekannte Verteilung eines oder mehrerer Merkmale anhand einer Stichprobe überprüft. Eine Nullhypothese H_0 wird gegen eine Alternativhypothese H_1 getestet.

Zwei Fehlentscheidungen sind möglich:

- Wenn H_0 wird abgelehnt wird, obwohl H_0 richtig ist, spricht man von einem Fehler 1.Art.
- Wenn H_0 angenommen wird, obwohl H_0 falsch ist, spricht man von einem Fehler 2.Art.

In der Praxis werden im Allgemeinen sogenannte **Signifikanztests** durchgeführt. Bei diesen Tests wird für die Wahrscheinlichkeit des Fehlers 1. Art eine obere Grenze, das sogenannte **Signifikanzniveau** α, vorgegeben. Über die Wahrscheinlichkeit des Fehlers 2. Art werden keine Aussagen gemacht.

Beispiel 1: („Zweiseitige Fragestellung")

Betrachtet wird ein normalverteiltes Merkmal $X \sim N(\mu, \sigma^2)$ mit bekanntem σ^2. Zu überprüfen ist die Hypothese

$H_0 : \mu = \mu_0$.

Wenn man davon ausgehen kann, dass Abweichungen nach beiden Seiten unerwünscht sind, wählt man in der Regel die Alternativhypothese

$H_1 : \mu \neq \mu_0$.

Falls H_0 richtig ist, besitzt die Stichprobenfunktion $T = \frac{\bar{X}_n - \mu_0}{\sigma}\sqrt{n}$ eine $N(0,1)$-Verteilung, und es gilt $P_{\mu_0}(-\lambda_{1-\frac{\alpha}{2}} \leq T \leq \lambda_{1-\frac{\alpha}{2}}) = 1 - \alpha$.

Man geht nun folgendermaßen vor: Aus der konkreten Stichprobe wird der Wert $t = \frac{\bar{x}_n - \mu_0}{\sigma}\sqrt{n}$ berechnet. H_0 wird zugunsten der Alternativhypothese abgelehnt, wenn t in der Menge

$K^* := \{\tilde{t} : \ |\tilde{t}| > \lambda_{1-\frac{\alpha}{2}}\}$, dem sogenannten **kritischen Bereich oder Ablehnungsbereich**, liegt. Anderenfalls ist „aufgrund dieses Tests nichts gegen H_0 einzuwenden".

(Gelegentlich wird auch das Urbild von K^*, nämlich
$K = \{(x_1, \dots, x_n) : | \frac{\bar{x}_n - \mu_0}{\sigma}\sqrt{n} | > \lambda_{1-\frac{\alpha}{2}}\}$, als kritischer Bereich oder Ablehnungsbereich bezeichnet.

Bemerkungen:

- Die Wahrscheinlichkeit dafür, dass H_0 abgelehnt wird, obwohl H_0 richtig ist, beträgt hier α.
- Angenommen, H_0 sei nicht richtig, d.h., $\mu \neq \mu_0$. Dann besitzt die Stichprobenfunktion $T = \frac{\bar{X}_n - \mu_0}{\sigma}\sqrt{n}$ eine $N(\frac{\mu-\mu_0}{\sigma}\sqrt{n}, 1)$-Verteilung. Die Wahrscheinlichkeit dafür, dass Ablehnung erfolgt, wenn H_0 falsch ist, ist dann stets größer als α.

  Die Wahrscheinlichkeit dafür, dass keine Ablehnung der falschen Hypothese erfolgt, ist stets kleiner als $1 - \alpha$; sie wird mit betragsmäßig wachsendem μ immer kleiner.
- Es besteht ein enger Zusammenhang zu den Konfidenzintervallen, denn es gilt
  $\frac{\bar{X}_n - \mu_0}{\sigma}\sqrt{n} \in [-\lambda_{1-\frac{\alpha}{2}}; \lambda_{1-\frac{\alpha}{2}}] \quad \Leftrightarrow \quad \mu_0 \in [\bar{X}_n - \lambda_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; \bar{X}_n + \lambda_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}]$.

Beispiel 2: („Einseitige Fragestellung")

Betrachtet werden Waschmittelpakete mit dem Nenngewicht (Sollgewicht) von 3000 g. Das Gewicht kann als normalverteilte Zufallsgröße $G \sim N(\mu, \sigma^2)$ aufgefasst werden. σ^2 sei bekannt.
Eine Verbraucherorganisation erhebt den Vorwurf, die Waschmittelpakete seien zu leicht, d.h., $\mu < 3000$ (in g).
Um die Berechtigung des Vorwurfs zu überprüfen, wird das Gewicht von (z.B.) $n = 30$ Waschmittelpaketen ermittelt. Dabei erhält man die Werte $x_1, \dots, x_{30}$.

Es wird die Nullhypothese $H_0 : \mu \geq 3000$ $(=: \mu_0)$
(auch $H_0 = 3000$)

gegen die Alternativhypothese

$H_1 : \mu < 3000$

getestet.

Bemerkung: Ein Vorwurf sollte stets durch die Alternativhypothese beschrieben werden, da nur für die Wahrscheinlichkeit, dass man sich beim Ablehnen der Nullhypothese (und damit der Bestätigung des Vorwurfs) irrt, eine obere Grenze angegeben werden kann. Das gilt in analoger Weise bei anderen Aufgabenstellungen für Aussagen, bei denen man sich möglichst nicht irren möchte.

H_0 wird zugunsten der Alternativhypothese abgelehnt, wenn $\frac{\bar{x}_n - \mu_0}{\sigma}\sqrt{n}$ zu klein ist, d.h., wenn

$$t = \frac{\bar{x}_n - \mu_0}{\sigma}\sqrt{n} \in K^* = (-\infty, \lambda_\alpha) = (-\infty, -\lambda_{1-\alpha})$$

erfüllt ist.

Anhand einer Skizze überzeugt man sich davon, dass

$$P_\mu(T \in K^*) \leq \alpha \quad \forall \mu \geq \mu_0 = 3000$$

gilt.

Bemerkung: Statistiksoftware gibt in der Regel das sogenannte **empirische Signifikanzniveau** (p-Wert) an. Für eine zweiseitige Aufgabenstellung ist das empirische Signifikanzniveau α^* durch die Gleichung $P_{H_0}(|T| \geq t_D) = \alpha^*$ definiert. Dabei bezeichnen T die Testgröße und t_D den Wert der Testgröße für die konkrete Stichprobe $(x_1, \ldots, x_n)$. P_{H_0} weist darauf hin, dass die Wahrscheinlichkeit bei Gültigkeit der Nullhypothese berechnet wird. Bei vorgegebenem Signifikanzniveau α wird die Nullhypothese daher abgelehnt, wenn $\alpha^* \geq \alpha$ gilt.

3.5.2 Parametertests bei normalverteiltem Merkmal

Im Folgenden wird stets $X \sim N(\mu, \sigma^2)$ vorausgesetzt.

a) Für bekanntes σ^2 soll die Hypothese

$H_0 : \mu = \mu_0$

überprüft werden. Der entsprechende Test wird in der Literatur in der Regel **Gauß-Test** genannt. Seine Durchführung lässt sich aus den Beispielen in 3.5.1 ersehen.

b) **Der einfache t-Test**

Dieser Test wird angewandt, wenn eine **Hypothese über den Erwartungswert** eines normalverteilten Merkmals bei **unbekannter Varianz** zu überprüfen ist.

Die Testgröße $T = \frac{\bar{X}_n - \mu_0}{S_n}\sqrt{n}$ besitzt bei Gültigkeit der Hypothese $H_0 : \mu = \mu_0$ eine t-Verteilung mit $n-1$ Freiheitsgraden.

- Bei einer zweiseitigen Fragestellung, die in der Regel durch die Hypothesen

  $H_0 : \mu = \mu_0, \qquad H_1 : \mu \neq \mu_0$
  gekennzeichnet ist, verwendet man den kritischen Bereich
  $K^* = \{t : \mid t \mid > t_{n-1;1-\frac{\alpha}{2}}\}$.

- Einseitige Fragestellungen treten in der Form

  $H_0 : \mu \overset{(>)}{=} \mu_0; \quad H_1 : \mu < \mu_0; \quad K^* = (-\infty, -t_{n-1;1-\alpha}) = (-\infty, t_{n-1;\alpha})$
  oder
  $H_0 : \mu \overset{(<)}{=} \mu_0; \quad H_1 : \mu > \mu_0; \quad K^* = (t_{n-1;1-\alpha}, \infty)$ auf.

Bemerkung: Der t-Test ist relativ robust gegen Verletzung der vorausgesetzten Normalverteiltheit.

c) **Der χ^2-Varianztest**

Es ist die Hypothese

$H_0 : \sigma^2 = \sigma_0^2 \quad (\sigma_0^2 > 0)$

bei unbekanntem Parameter μ zu überprüfen. Hier und im Folgenden wird nur die zweiseitige Fragestellung näher betrachtet.
Die Testgröße $T = \frac{1}{\sigma_0^2} \sum\limits_{i=1}^{n} (X_i - \bar{X}_n)^2 = \frac{(n-1)S_n^2}{\sigma_0^2}$ besitzt bei Gültigkeit von H_0 eine χ^2-Verteilung mit $n-1$ Freiheitsgraden. Damit ergibt sich der Ablehnungsbereich

$K^* = (0, \chi^2_{n-1;\frac{\alpha}{2}}) \cup (\chi^2_{n-1;1-\frac{\alpha}{2}}, \infty)$.

3.5.3 Prüfung einer unbekannten Wahrscheinlichkeit

Es soll ein Test zum Überprüfen von $H_0 : p = p_0$ angegeben werden, wobei $p = P(A)$ die Wahrscheinlichkeit für das Eintreten des Ereignisses A bezeichnet. Ausgangspunkt ist eine

Stichprobe $(X_1, \ldots, X_n)$ mit $X_i = \begin{cases} 1, & \text{falls } A \text{ im } i\text{-ten Versuch eingetreten ist,} \\ 0 & \text{sonst.} \end{cases}$

Bei Gültigkeit von H_0 ist $\sum\limits_{i=1}^{n} X_i$ binomialverteilt mit $p = p_0$ und n. Daraus kann ein Test für „kleine" Stichprobenumfänge n abgeleitet werden.

Für hinreichend großes n ist $T = \dfrac{\sum\limits_{i=1}^{n} X_i - np_0}{\sqrt{np_0(1-p_0)}}$ nach dem Zentralen Grenzwertsatz von de Moivre-Laplace näherungsweise $N(0,1)$-verteilt. In diesem Fall ergibt sich als kritischer Bereich für die zweiseitige Aufgabenstellung $K^* = \{t : \; |t| > \lambda_{1-\frac{\alpha}{2}}\}$.

3.5.4 Anpassungstests und visuelle Methoden zur Beurteilung der Güte der Anpassung

Anpassungstests sind Tests zum Überprüfen der Hypothese

$H_0 : F_X(x) = F_0(x) \quad \forall x,$

d.h., es wird überprüft, ob die vorliegenden Daten mit der Hypothese, dass das zugrunde liegende Merkmal entsprechend F_0 verteilt ist, verträglich sind. Ausgangspunkt der Untersuchungen ist wieder eine Stichprobe $(X_1, \ldots, X_n)$ aus einer zu X gehörenden Grundgesamtheit.

a) **Der χ^2-Anpassungstest**

Die Testgröße $\chi^2 := \sum\limits_{i=1}^{p} \dfrac{(H_n(i) - np_i)^2}{np_i} = \sum\limits_{i=1}^{p} \dfrac{(H_n(i))^2}{np_i} - n$ ist asymptotisch χ^2-verteilt mit $p-1$ Freiheitsgraden. Dabei bezeichnen

- p die Anzahl der auf der reellen Achse gewählten Klassen $K_1, \ldots, K_p$,
- $H_n(i)$ die absolute Häufigkeit der i-ten Klasse und
- $p_i = P(X \in K_i)$, wobei P das durch F_0 gegebene Wahrscheinlichkeitsmaß bedeutet.

Der Ablehnungsbereich hat die Gestalt $K^* = (\chi^2_{p-1;1-\alpha}, \infty)$.

Bemerkungen:

1.) Der χ^2-Test ist meistens wenig wirksam, (d.h., falsche Hypothesen werden „selten" abgelehnt) aber universell einsetzbar, da keine weiteren Voraussetzungen an F_0 gestellt werden.

2.) Für die Klasseneinteilung gibt es verschiedene Empfehlungen, z.B wird häufig gefordert, dass für jede Klasse $np_i \geq 5$ gilt.

3.) Wurden zur Festlegung von F_0 r Parameter (durch Maximum-Likelihood-Schätzungen) geschätzt, so ist das Quantil der χ^2-Verteilung mit $p - r - 1$ Freiheitsgraden zu benutzen.

Beispiel:

In einer Maschinenhalle arbeiten 6 Maschinen. In $n = 200$ Schichten wurde die Anzahl der pro Schicht ausgefallenen Maschinen registriert. (Da die Maschinen sofort nach Ausfall repariert werden, kann eine Maschine auch mehrmals pro Schicht ausfallen.) Die Daten wurden in einer Häufigkeitstabelle zusammengefasst: $h_{200}(i)$ gibt an, wie oft genau i Ausfälle registriert wurden.
Es ist zu überprüfen, ob die Anzahl der Ausfälle (Merkmal X) einer Poissonverteilung genügt.

Der Parameter λ der Poissonverteilung wird geschätzt durch die Maximum-Likelihood-Schätzung $\hat{\lambda} = \bar{x}_{200} = 1,8$.

Die Größen p_i berechnen sich wie folgt:
$p_i = P(X = i) = \frac{\lambda^i}{i!}e^{-\lambda}, \quad i = 0, \ldots, 6, \quad p_7 = P(X \geq 7) = 1 - \sum_{i=0}^{6} p_i.$

Damit ergibt sich folgende Tabelle:

Anzahl der Ausfälle i	Häufigkeit $h_{200}(i)$		$200 \cdot p_i$		$\frac{(h_{200}(i) - 200p_i)^2}{200p_i}$
0	41		33,06		1,91
1	62		59,50		0,11
2	45		53,56		1,37
3	22		32,14		3,20
4	16		14,46		0,16
5	8		$5,20$		
6	4	14	$1,56$	$7,28$	6,20
7	2		$0,52$		
	200		200		$\chi^2 = 12,95$

Die ursprünglich vorgesehenen Klassen (Zahl der Ausfälle) 5, 6, 7 wurden dabei zu einer neuen Klasse zusammengefasst, um $np_i \geq 5$ zu gewährleisten.

Der Wert der Testgröße ergibt sich zu $12,95 > \chi^2_{6-1-1;0,95} = 9,5$. Die Hypothese „Poissonverteilung" wird daher abgelehnt.

b) **Der Kolmogorov-Smirnov-Test**

Bei diesem Test wird ebenfalls die Hypothese $H_0 : F_X = F_0$ überprüft, es wird aber zusätzlich angenommen, dass F_0 die Verteilungsfunktion eines **stetig verteilten Merkmals** bezeichnet.

Die Testgröße hat die Gestalt $D := \sup\limits_{x \in R} \mid \hat{F}_n(x) - F_0(x) \mid$, wobei $\hat{F}_n$ die aus der Stichprobe ermittelte empirische Verteilungsfunktion bezeichnet. (Das Supremum liegt an einer der Sprungstellen von $\hat{F}_n$.)

Falls H_0 wahr ist, hängt die Verteilung von D nur von n und *nicht* von F_0 ab. Das gilt nicht mehr, falls Parameter von F_0 geschätzt werden müssen (Es gibt Modifikationen dieses Tests bei geschätzten Parametern für die Normal- und die Exponentialverteilung). Bei wahrer Hypothese ist $\sqrt{n}D$ asymptotisch Kolmogorov-verteilt.

Für kleine n arbeitet man entweder mit Modifikationen der Testgröße (z.B. $T = (\sqrt{n}+0,12+\frac{0,11}{n})D$) und den Quantilen der Kolmogorov-Verteilung, oder man benutzt die Quantile $d_{n;1-\alpha}$ der exakten Verteilung von D, die man in Tafelwerken finden kann. Der Ablehnungsbereich hat dann die Form $K^* = \{d : \ d > d_{n;1-\alpha}\}$.

c) **Tests auf Normalverteilung**
Der „gewöhnliche" χ^2-Anpassungstest und der Kolmogorov-Smirnov-Test haben sich als weniger geeignet erwiesen. Es gibt allerdings eine empfehlenswerte Modifikation des χ^2-Anpassungstests, der eine Einteilung der reellen Zahlen in gleichwahrscheinliche Klassen zugrunde liegt.

Nach gegenwärtigem Kenntnisstand gilt der **Shapiro-Wilk-Test** als der beste Test auf Normalverteilung.
Leicht durchzuführenden Tests sind der Anderson-Darling-Test und der D-test von D'Agostino.
Für große Stichprobenumfänge ($n \geq 500$) haben sich auch Tests bewährt, die die Stichprobenschiefe

$$\hat{\gamma}_1 = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X}_n)^3}{[\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X}_n)^2]^{\frac{3}{2}}}$$

bzw. den Stichprobenexzess

$$\hat{\gamma}_2 = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X}_n)^4}{[\frac{1}{n}\sum\limits_{i=1}^{n}(X_i - \bar{X}_n)^2]^2}$$

überprüfen. Für die Normalverteilung gilt $\gamma_1 = 0, \gamma_2 = 3$.

d) **Visuelle Methoden zur Beurteilung der Güte der Anpassung**

α) **Q-Q-Plots** vergleichen die **Quantile** der empirischen Verteilung $Q_{p_i}^{emp}$ mit den Quantilen der durch F_0 gegebenen Verteilung $Q_{p_i}^{F_0}$. Bei einer guten Anpassung liegen die Punkte $(Q_{p_i}^{emp}, Q_{p_i}^{F_0})$ für geeignete p_i näherungsweise auf der Winkelhalbierenden. In der Regel werden die Punkte $(x_i^*, F_0^{-1}(\frac{i-0{,}5}{n})),\ i = 1, \ldots, n$ dargestellt (Einbeziehung einer "Stetigkeitskorrektur").

β) **P-P-Plots** vergleichen die **Werte** der empirischen Verteilungsfunktion mit den Werten von F_0. Bei einer guten Anpassung liegen die Punkte $(F_n(x_i^*), F_0(x_i^*))$ $i = 1, \ldots, n$, näherungsweise auf der Winkelhalbierenden. In der Regel werden die Punkte $(\frac{i-0{,}5}{n}, F_0(x_i^*)),\ i = 1, \ldots, n$, dargestellt (Einbeziehung einer "Stetigkeitskorrektur").

γ) Bei **Normal-Wahrscheinlichkeits-Plots** werden die Werte der empirischen Verteilungsfunktion in ein **Koordinatensystem** eingetragen, in dem die Abszissenachse linear, die **Ordinatenachse entsprechend Φ^{-1}** unterteilt ist. Bei einer guten Anpassung liegen die Punkte $(x_i^*, F_n(x_i^*)),\ i = 1, \ldots, n$, näherungsweise auf einer Geraden.
Aus der Lage dieser Geraden können Rückschlüsse auf Erwartungswert und Varianz der Normalverteilung gezogen werden.
Wahrscheinlichkeitsplots werden auch bei Lebensdauerverteilungen angewandt.

3.5.5 Zwei-Stichproben-Vergleiche

Beim Vergleich zweier Stichproben ist zu unterscheiden, ob zwei **unabhängige Stichproben** $(X_1, \ldots X_{n_1})$ und $(Y_1, \ldots, Y_{n_2})$ oder zwei **verbundene Stichproben** $((X_1, Y_1), \ldots, (X_n, Y_n))$ vorliegen.

a) **Der doppelte t-Test (Vergleich der Erwartungswerte)**

Gegeben seien zwei normalverteilte Merkmale $X \sim N(\mu_X, \sigma_X^2)$ und $Y \sim N(\mu_Y, \sigma_Y^2)$, wobei $\sigma_X^2 = \sigma_Y^2$ gelten muss, und zwei unabhängige Stichproben $(X_1, \ldots, X_{n_1}),\ (Y_1, \ldots, Y_{n_2})$.

Falls $\sigma_X^2 = \sigma_Y^2$ nicht erfüllt ist, kann man den Welch-Test oder den Bartlett-Scheffé-Test verwenden. Hinsichtlich dieser Tests wird auf weiterführende Literatur (Behrens-Fisher-Problem) verwiesen.

Überprüft wird die Nullhypothese

$H_0 : \mu_X = \mu_Y$.

Die Testgröße $\tilde{T}_n := \dfrac{(\bar{X}_{n_1} - \bar{Y}_{n_2})\sqrt{n_1 n_2 (n_1 + n_2 - 2)}}{\sqrt{(n_1 + n_2)[(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2]}}$
besitzt bei Gültigkeit der Nullhypothese eine t-Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden.
Der Ablehnungsbereich hat (bei zweiseitiger Fragestellung) die Gestalt
$K^* = \{t : \ |t| > t_{n_1+n_2-2;1-\frac{\alpha}{2}}\}$.

Bemerkungen:

1. Die Voraussetzung $\sigma_X^2 = \sigma_Y^2$ kann mit Hilfe des F-Tests überprüft werden. (In der Regel wird bei „vorgeschalteten” Tests das Signifikanzniveau 2α gewählt).

2. Bei verbundenen Merkmalen X, Y kann man das „Merkmal $Z = X - Y$" und damit die Stichprobe $(X_1 - Y_1, \ldots, X_n - Y_n)$ betrachten und die Hypothese $\mathrm{E}(Z) = 0$ überprüfen.

b) **Der F-Test (Vergleich der Varianzen)**

Gegeben seien zwei normalverteilte Merkmale $X \sim N(\mu_X, \sigma_X^2)$ und $Y \sim N(\mu_Y, \sigma_Y^2)$ sowie zwei unabhängige Stichproben $(X_1, \ldots, X_{n_1})$ und $(Y_1, \ldots, Y_{n_2})$.

Überprüft wird die Nullhypothese

$H_0 : \sigma_X^2 = \sigma_Y^2$.

Die Testgröße $F = \frac{S_X^2}{S_Y^2}$ besitzt bei Gültigkeit der Nullhypothese eine F-Verteilung mit $(n_1 - 1, n_2 - 1)$ Freiheitsgraden.
Der Ablehnungsbereich hat (bei zweiseitiger Fragestellung) die Gestalt
$K^* = (0, F_{n_1-1,n_2-1;\frac{\alpha}{2}}) \cup (F_{n_1-1,n_2-1;1-\frac{\alpha}{2}}, \infty)$.

c) **Der Zeichentest (Vorzeichentest)**

Gegeben seien zwei verbundene, stetig verteilte Merkmale X und Y mit den Verteilungsfunktionen F_X und F_Y und verbundene Stichproben $((X_1, Y_1), \ldots, (X_n, Y_n))$.

Überprüft wird die Nullhypothese

$H_0 : F_X = F_Y$.

Die Testgröße

$Z_n^+ :=$ Anzahl der Paare(X_i, Y_i) für die $X_i > Y_i$ gilt,

ist bei Gültigkeit der Nullhypothese binomialverteilt mit n und $p = \frac{1}{2}$, denn aufgrund der Voraussetzungen hat man $P(X_i > Y_i) = P(X_i < Y_i) = \frac{1}{2}$.
Der Ablehnungsbereich hat die Gestalt $K^* = \{t :\ t < b_{n;\frac{\alpha}{2}} \ \vee\ t > n - b_{n;\frac{\alpha}{2}}\}$, wobei $b_{n;\frac{\alpha}{2}}$ das Quantil der Ordnung $\frac{\alpha}{2}$ der Binomialverteilung mit n und $p = \frac{1}{2}$ bezeichnet.

Bemerkungen:

1) Für große n kann näherungsweise mit der Normalverteilung gearbeitet werden, denn die folgende Testgröße T ist nach dem Zentralen Grenzwertsatz von de Moivre-Laplace asymptotisch $N(0,1)$-verteilt:

$$T = \frac{Z_n^+ - np}{\sqrt{np(1-p)}} = \frac{Z_n^+ - \frac{n}{2}}{\sqrt{\frac{n}{4}}} = \frac{2Z_n^+}{\sqrt{n}} - \sqrt{n}.$$

2) Tritt in Einzelfällen $X_i = Y_i$ auf, so werden die betreffenden Paare weggelassen, und n wird entsprechend reduziert.

3) Der Zeichentest kann auch für 1-Stichproben-Probleme verwendet werden, z.B:

   - als Test auf **Symmetrie der Verteilung** (Z_n^+ ... Anzahl der X_i mit $X_i > 0$),
   - als Test der Hypothese **Median=M_0** (Z_n^+ ... Anzahl der X_i mit $X_i > M_0$).

d) **Der U-Test** (auch Wilcoxon-Test, Mann-Whitney-Test)

Gegeben seien zwei stetig verteilte Merkmale und zwei unabhängige Stichproben
$(X_1, \ldots, X_{n_1}), (Y_1, \ldots, Y_{n_2})$. Überprüft wird die Nullhypothese

$H_0 : F_X = F_Y$.

Zur Berechnung der Testgröße U werden die $n_1 + n_2$ Messwerte gemeinsam in eine Variationsreihe geordnet, und jedem Wert wird seine Rangzahl zugeordnet. Im Folgenden bezeichnen R_X die Summe der Rangzahlen aller X-Werte und R_Y die Summe der Rangzahlen aller Y-Werte.

Die Testgröße ergibt sich dann als $U = R_X - \frac{n_1(n_1+1)}{2}$, wobei R_X die Summe der Rangzahlen aller X-Werte bezeichnet.

Bemerkungen:

1) Die Merkmale X und Y sind gleichberechtigt.

2) Ermittelt man auch die Summe der Rangzahlen aller Y-Werte R_Y, kann man die Rechenkontrolle $R_X + R_Y = \frac{(n_1+n_2)(n_1+n_2+1)}{2}$ verwenden.

Der Ablehnungsbereich hat die Gestalt $K^* = \{u : \; u < u_{n_1,n_2;\frac{\alpha}{2}} \;\vee\; u > n_1 n_2 - u_{n_1,n_2;\frac{\alpha}{2}}\}$,
wobei $u_{n_1,n_2;\alpha}$ das Quantil der Ordnung α der exakten Verteilung der Testgröße U bezeichnet, das man in Tafelwerken finden kann.

Für $n_1 \geq 4,\; n_2 \geq 4\;, n_1 + n_2 \geq 20$ ist U in guter Näherung normalverteilt mit
$\mathrm{E}(U) = \frac{1}{2} n_1 n_2,\; \mathrm{var}(U) = \frac{1}{12} n_1 n_2 (n_1 + n_2 + 1)$.

Man verwendet daher in diesen Fällen die standardisierte Testgröße

$$Z = \frac{R_X - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{1}{12}n_1n_2(n_1+n_2+1)}},$$

die näherungsweise $N(0,1)$-verteilt ist.

Beispiel:

2 Fertigungslinien arbeiten parallel und stellen das gleiche Erzeugnis her. Gemessen wurde die Zeit zur Herstellung eines Erzeugnisses 12-mal bei Linie 1 und 9-mal bei Linie 2. Es ergaben sich folgende Messwerte (Messungen in s):

Linie 1: 334 339 336 321 333 365 347 339 328 329 340 338

Linie 2: 337 335 334 322 327 330 323 326 332

Es ist die Frage zu klären, ob es einen Unterschied zwischen Linie 1 und 2 gibt.

Zur Feststellung der Summe der Rangzahlen können die Messwerte wie folgt angeordnet und ausgewertet werden:

x_i	y_i	Rang(x_i)	Rang(y_i)
321		1	
	322		2
	323		3
	326		4
	327		5
328		6	
329		7	
	330		8
	332		9
333		10	
334		11,5	
	334		11,5
	335		13
336		14	
	337		15
338		16	
339		17	
339		18	
340		19	
347		20	
365		21	
		$\underbrace{160,5}_{R_X}$	$\underbrace{70,5}_{R_Y}$

Die Testgröße Z nimmt den Wert $z = \frac{160{,}5 - \frac{12(12+9+1)}{2}}{\sqrt{\frac{12 \cdot 9}{12} \cdot (12+9+1)}} = 2,025$ an. Da die Voraussetzungen für die Verwendung der Normalverteilung erfüllt sind, kann man die Quantile der standardisierten Normalverteilung verwenden und erhält für $\alpha = 0,05$ den Ablehnungsbereich $K^* = \{z : \ |z| > 1,96\}$.

Da $z \in K^*$ gilt, ist ein signifikanter Unterschied zwischen den Linien 1 und 2 nachgewiesen.

e) **Untersuchung des (einfachen) Korrelationskoeffizienten**

Gegeben seien ein normalverteilter Zufallsvektor $(X, Y)^T$ und zwei verbundene Stichproben $((X_1, Y_1), \ldots, (X_n, Y_n))$.

Der folgende Test kann in diesem Fall zur **Überprüfung der Unabhängigkeit** herangezogen werden:

α) Es wird die Hypothese $H_0 : \varrho_{XY} = 0$ (d.h. X und Y sind unabhängig) überprüft.

Die Testgröße

$$T = \frac{r_{XY}\sqrt{n-2}}{\sqrt{1-r_{XY}^2}}$$

ist bei Gültigkeit von H_0 t-verteilt mit $n-2$ Freiheitsgraden.

Als Ablehnungsbereich wird $K^* = \{t :\ |t| > t_{n-2;1-\frac{\alpha}{2}}\}$ gewählt.

Es ist auch möglich, andere **Werte des Korrelationskoeffizienten** zu überprüfen:

β) Es wird die Hypothese $H_0 : \varrho_{XY} = \varrho_0$ betrachtet, wobei $|\varrho_0| \notin \{0, 1\}$ gelten soll.

Die Testgröße

$$T = (\frac{1}{2}\ln\frac{1+r_{XY}}{1-r_{XY}} - \frac{1}{2}\ln\frac{1+\varrho_0}{1-\varrho_0} - \frac{\varrho_0}{2(n-1)})\sqrt{n-3}$$

ist bei Gültigkeit von H_0 asymptotisch $N(0,1)$-verteilt, und man erhält den Ablehnungsbereich $K^* = \{t :\ |t| > z_{1-\frac{\alpha}{2}}\}$.

Man kann dabei in der Regel davon ausgehen, dass die Approximation durch die Normalverteilung ab $n = 25$ gerechtfertigt ist.

Bemerkungen:

1.) Weitere Korrelationskoeffizienten und ihre Untersuchung findet man in der Literatur unter dem Stichwort „Korrelationsanalyse“.

2.) Aus der Testgröße T im Fall β) kann unter Vernachlässigung des Terms $\frac{\varrho_0}{2(n-1)}$ ein Konfidenzintervall für ϱ_{XY} hergeleitet werden.

f) **Der χ^2-Unabhängigkeitstest**

Der χ^2-Unabhängigkeitstest kann für die Überprüfung der Unabhängigkeit auch im Falle von ordinalskalierten oder nominalskalierten Daten verwendet werden.

Gegeben seien ein Zufallsvektor $(X, Y)^T$ und verbundene Stichproben $((X_1, Y_1), \ldots, (X_n, Y_n))$.

Es wird die Hypothese

$H_0 : X$ und Y sind unabhängig

überprüft.

Zur Berechnung der Testgröße geht man von zwei Zerlegungen $I_1 \cup \ldots \cup I_k = R$ und
$J_1 \cup \ldots \cup J_k = R$ der reellen Zahlen in disjunkte Intervalle und Halbachsen (Klassen) aus. (Bei diskret verteiltem Vektor wählt man häufig $I_i = \{x_i\}, J_j = \{y_j\}$.)

Es seien nun $H_{ij} =$ die Anzahl der Zahlenpaare, die in $I_i \times J_j$ liegen, $H_{i\bullet} = \sum_{j=1}^{l} H_{ij}$ und $H_{\bullet j} = \sum_{i=1}^{k} H_{ij}$. Es sollte gewährleistet sein, dass $\frac{h_{i\bullet} \cdot h_{\bullet j}}{n} \geq 5$ gilt. Anderenfalls werden - wie beim χ^2-Anpassungstest - Klassen zusammengefasst.

Die Testgröße

$$T = n \cdot \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(H_{ij} - \frac{H_{i\bullet} \cdot H_{\bullet j}}{n})^2}{H_{i\bullet} \cdot H_{\bullet j}}$$

ist bei Gültigkeit von H_0 asymptotisch χ^2-verteilt mit $(k-1) \cdot (l-1)$ Freiheitsgraden.

Der Ablehnungsbereich hat die Gestalt $K^* = \{t : \; t > \chi^2_{(k-1)(l-1);1-\alpha}\}$.

Als Spezialfall ergibt sich das **Vier-Felder- χ^2-Prüfverfahren**:

X und Y nehmen jeweils nur 2 Werte (z.B. 0 und 1) an. Dann vereinfacht sich die Testgröße in folgender Weise:

$$T = \frac{n \cdot (H_{11}H_{22} - H_{12}H_{21})^2}{H_{1\bullet} \cdot H_{2\bullet} \cdot H_{\bullet 1} \cdot H_{\bullet 2}}.$$

Für den Ablehnungsbereich gilt $K^* = \{t : \; t > \chi^2_{1;1-\alpha}\}$.

Zur Verbesserung der Näherung für kleine Werte $h_{i\bullet}$ und $h_{\bullet j}$ wird die folgende Korrektur der Testgröße vorgeschlagen (Korrektur von Yates):

$$\tilde{T} := \frac{(\,|H_{11}H_{22} - H_{12}H_{21}| - \frac{n}{2})^2}{H_{1\bullet} \cdot H_{2\bullet} \cdot H_{\bullet 1} \cdot H_{\bullet 2}}.$$