1 Parameterschätzung

Mittels statistischer Schätzverfahren werden unbekannte Parameter einer Grundgesamtheit oder unbekannte Verteilungen aus den Werten einer Stichprobe geschätzt. Behandelt werden in dieser Kurseinheit ausschließlich Parameterschätzungen, wobei zwischen **Punktschätzung** und **Intervallschätzung** unterschieden wird.

Punktschätzung: Punktschätzung

Eine Punktschätzung entspricht der Angabe eines einzelnen Wertes für den zu schätzenden Parameter.

Intervallschätzung: Intervallschätzung

Abweichungen eines Punktschätzers vom wahren Wert sind in der Regel unvermeidlich. Daher wird eine Punktschätzung oft durch eine Intervallschätzung ergänzt. Dabei wird ein Intervall bestimmt, das den wahren, unbekannten Wert des Parameters mit einer vorgegebenen Wahrscheinlichkeit überdeckt.

1.1 Schätzfunktionen und Punktschätzung

Geschätzt werden soll ein unbekannter Parameter der Grundgesamtheit (z.B. μ, σ^2 oder π), der allgemein mit θ bezeichnet wird. Für die Schätzung wird der Grundgesamtheit eine Zufallsstichprobe der Größe n entnommen. Ihre Elemente werden als Realisationen der Zufallsvariablen $X_1, X_2, ..., X_n$ aufgefasst. Die Zufallsvariablen $X_1, X_2, ..., X_n$ werden in diesem Zusammenhang auch Stichprobenvariablen genannt. Aus den Stichprobenwerten muss nun ein geeigneter Schätzwert $\hat{\theta}$ für den unbekannten Parameter θ berechnet werden. Zur Ermittlung eines Schätzwertes $\hat{\theta}$ wird eine geeignete **Stichprobenfunktion g** mit

Stichprobenfunktion

$$T = g(X_1, ..., X_n),$$

gesucht, die von den Elementen und dem Umfang der Stichprobe abhängt.

> Eine für Schätzungen verwendete Stichprobenfunktion $T = g(X_1, ..., X_n)$ heißt **Schätzfunktion** oder auch **Statistik**.

Schätzfunktion, Statistik

Die Schätzfunktion T wird als Schätzer für den Parameter θ verwendet, d.h. $\hat{\Theta} = T$. Als Schätzer wird in vielen Fällen eine Schätzfunktion gewählt, die einer bekannten Kenngröße der deskriptiven Statistik entspricht.

Beispiel 1.1.1:

a) Die Schätzfunktion

$$\hat{\Theta} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

liefert einen Schätzwert $\hat{\mu}$ für den Parameter μ der Grundgesamtheit bzw. für den unbekannten Erwartungswert $E(X_i) = \mu$ der Zufallsvariablen X_i.

b) Die Schätzfunktion

$$\hat{\Theta} = P = \frac{X}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{mit } X_i = \begin{cases} 0 & \text{für „}\overline{A}\text{ tritt ein“} \\ 1 & \text{für „}A\text{ tritt ein“} \end{cases}$$

liefert einen Schätzwert $\hat{\pi}$ für den Anteilswert π der Grundgesamtheit bzw. für die unbekannte Wahrscheinlichkeit π für das Auftreten des interessierenden Ereignisses A.

Schätzwert, Punktschätzung

Der sich für bestimmte Stichprobenwerte $x_1, x_2, ..., x_n$ ergebende Wert der Schätzfunktion $g(x_1, ..., x_n) = t$ heißt **Schätzwert** oder **Punktschätzung**, d.h. $\hat{\theta} = t$.

Beispiel 1.1.2:

*a) Für eine einfache Zufallsstichprobe von $n = 100$ Studenten der Fernuniversität wurde ein durchschnittlicher Mineralwasserkonsum von $\overline{x} = 1.84$ l/Tag ermittelt. Als Punktschätzung für den durchschnittlichen Konsum μ **aller** Studenten der Fernuniversität ergibt sich damit $\hat{\theta} = \hat{\mu} = \overline{x} = 1.84$ l/Tag.*

*b) Aufgrund einer einfachen Zufallsstichprobe wird ermittelt, dass ein Anteil von 24% der Studenten der Fernuniversität verheiratet ist. Für den unbekannten Anteil π **aller** Studenten der Fernuniversität lautet die Punktschätzung somit $\hat{\theta} = \hat{\pi} = p = 0.24$.*

Die Ausführungen zeigen, dass der Grundgedanke der Parameterschätzung als Punktschätzung verhältnismäßig einfach ist. Probleme ergeben sich u.a. dann, wenn für einen Parameter unterschiedliche Schätzfunktionen verfügbar sind und zu entscheiden ist, welche Schätzfunktion verwendet werden soll. Als Entscheidungskriterien können bestimmte Eigenschaften von Schätzfunktionen betrachtet werden. Auch Gesichtspunkte wie Kosten oder die Einfachheit der Schätzfunktion können eine Rolle spielen.

Hinweis:

Um Missverständnisse bei den Symbolen auszuschließen, wird an dieser Stelle noch einmal auf die in diesem Kurs verwendeten Bezeichnungen hingewiesen:

- Große Buchstaben bezeichnen Zufallsvariablen.
- Kleine Buchstaben bezeichnen Realisationen von Zufallsvariablen oder Konstanten.
- Der Merkmalswert des i-ten Stichprobenelementes ist allgemein eine Zufallsvariable X_i.
- Der Beobachtungswert des i-ten Elementes einer bestimmten Stichprobe ist eine Realisation x_i der Zufallsvariablen X_i.
- Spezielle Punktschätzer und Schätzwerte (z.B. für μ, σ^2, π) werden mit einem Dach gekennzeichnet ($\hat{\mu}, \hat{\sigma}^2, \hat{\pi}$).

1.2 Eigenschaften von Schätzfunktionen

Schätzfunktionen sind durch bestimmte Eigenschaften charakterisiert, die Auskunft darüber geben, wie gut Schätzfunktionen für bestimmte Zwecke geeignet sind. Behandelt werden an dieser Stelle drei wichtige Eigenschaften, nämlich die **Erwartungstreue**, die **Effizienz** und die **Konsistenz**.

Erwartungstreue:

erwartungstreue (unverzerrte) Schätzfunktion

> Eine Schätzfunktion T des Parameters θ heißt **erwartungstreu** oder **unverzerrt** (engl.: „unbiased“), wenn die Beziehung $E(T) = \theta$ gilt, d.h. der Erwartungswert der Zufallsvariablen T entspricht dem wahren Wert des zu schätzenden Parameters.

Beispiel 1.2.1:

Die Stichprobenfunktion $\overline{X}$ besitzt den Erwartungswert $E(\overline{X}) = \mu$.

$$\begin{aligned} E(\overline{X}) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{1}{n} n\ \mu = \mu \end{aligned}$$

erwartungstreue Schätzfunktion für μ

Der Stichprobenmittelwert $\boldsymbol{\overline{X}}$ ist also eine **erwartungstreue Schätzfunktion für den Erwartungswert $\boldsymbol{\mu}$** der Grundgesamtheit.

Um eine erwartungstreue Schätzfunktion für die Varianz der Grundgesamtheit zu erhalten, wird die **Stichprobenvarianz** wie folgt definiert (Beweis s. Aufgabe 1 zu Kapitel 1):

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Eine nicht erwartungstreue Schätzfunktion für die Varianz σ^2 der Grundgesamtheit ist $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ (siehe Definition der empirischen Varianz in Kurseinheit 1, Abschnitt 2.3.2). Für einfache Zufallsstichproben einer unendlichen Grundgesamtheit gilt:

$$E(\tilde{S}^2) = E(\frac{n-1}{n}S^2) = \frac{n-1}{n}\sigma^2.$$

Somit ist also

$$\hat{\sigma}^2 = S^2 = \frac{n}{n-1}\tilde{S}^2$$

erwartungstreue Schätzfunktion für σ^2

eine **erwartungstreue Schätzfunktion für die Varianz $\boldsymbol{\sigma^2}$ einer unendlichen Grundgesamtheit**, d.h. S^2 liefert eine unverzerrte Schätzung.

Für eine **endliche Grundgesamtheit** vom Umfang N und eine **Stichprobe ohne Zurücklegen** vom Umfang n ergibt sich eine erwartungstreue Schätzfunktion für σ^2 durch

$$\hat{\sigma}^2 = \frac{N-1}{N}S^2.$$

Effizienz:

Eine erwartungstreue Schätzfunktion $\boldsymbol{T}$ für den Parameter θ einer Grundgesamtheit heißt **effizient**, wenn T eine endliche Varianz besitzt und wenn es für θ keine andere erwartungstreue Schätzfunktion T^* gibt, welche eine kleinere Varianz als T besitzt. Es wird mit der Effizienz also die Forderung der **Erwartungstreue und minimaler Varianz** gestellt.

effiziente Schätzfunktion

Beispiel 1.2.2:

a) *Für normalverteiltes X ist die Schätzfunktion $\overline{X}$ effizient. Es existiert keine erwartungstreue Schätzfunktion für den Parameter μ einer Grundgesamtheit, die eine kleinere Varianz als der Stichprobenmittelwert $\overline{X}$ hat. Die Varianz von $\overline{X}$ lautet*

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}.$$

b) *Häufig wird als Schätzfunktion für den Parameter μ einer Grundgesamtheit an Stelle von $\overline{X}$ der Stichprobenmedian X_{med} verwendet, der ebenfalls erwartungstreu für μ ist. Meist erfordert die Ermittlung des Medians weniger Aufwand als die Berechnung des arithmetischen Mittels.*

Allerdings hat die Schätzfunktion X_{med} häufig den Nachteil, dass sie nicht effizient ist. So gilt z.B. für eine normalverteilte Zufallsvariable X und für einen großen Stichprobenumfang n approximativ:

$$\sigma^2_{X_{med}} = \sigma^2 \frac{\pi}{2n} = \frac{\pi}{2}\frac{\sigma^2}{n} = \frac{\pi}{2}\sigma^2_{\overline{X}} = 1.571\ \sigma^2_{\overline{X}} > \sigma^2_{\overline{X}}.$$

Die Varianz $\mathrm{Var}(X_{med})$ ist also größer als die Varianz $\mathrm{Var}(\overline{X})$ und folglich nicht minimal. Eine Schätzung des Parameters μ durch X_{med} ist wegen dieser größeren Streuung mit mehr Unsicherheit behaftet.

Eine Schätzfunktion T für einen Parameter θ ist eine Zufallsvariable, die durch eine Dichtefunktion $f_T(t)$ beschrieben werden kann. In Abbildung 1.2.1 sind die Dichtefunktionen von drei Schätzfunktionen für denselben Parameter θ eingezeichnet.

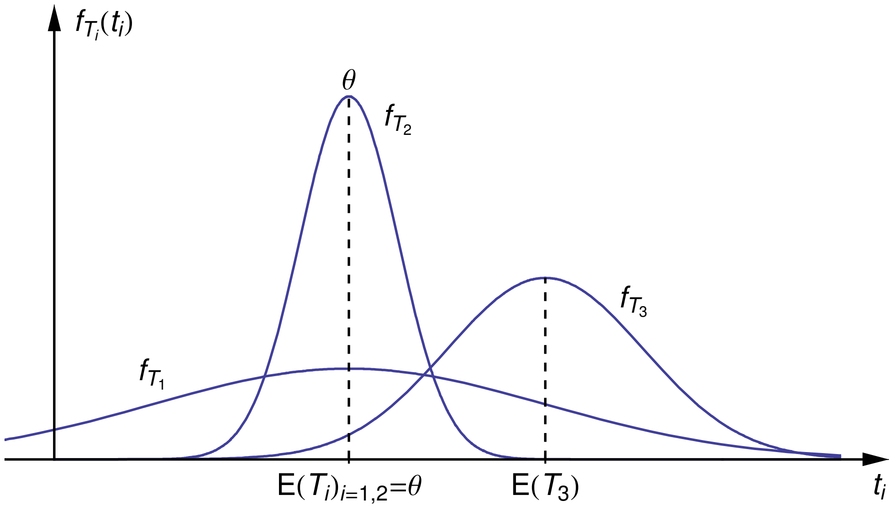

Abbildung 1.2.1: *Dichtefunktion drei verschiedener Schätzfunktionen für denselben Parameter* θ

Die Schätzfunktionen T_1 und T_2 sind erwartungstreu, da $E(T_i) = \theta$ $(i = 1, 2)$ gilt. Die Schätzfunktion T_3 ist dagegen nicht erwartungstreu, denn $E(T_3) \neq \theta$.

Von den beiden erwartungstreuen Schätzfunktionen T_1 und T_2 besitzt T_2 eine kleinere Varianz als T_1, d.h. T_1 ist nicht effizient.

Ein Kriterium, welches die Abweichung einer Schätzfunktion von dem zu schätzenden Parameter beurteilt, ist die mittlere quadratische Abweichung.

MSE

Ist T eine Schätzfunktion für den Parameter θ, dann ist die **mittlere quadratische Abweichung (engl.: mean squared error (MSE))** der Schätzfunktion gegeben durch:

$$\mathrm{MSE}(T) = E[(T-\theta)^2].$$

Für die mittlere quadratische Abweichung gilt:

$$\mathrm{MSE}(T) = \mathrm{Var}(T) + [E(T)-\theta]^2 = \mathrm{Var}(T) + \mathrm{Bias}(T)^2.$$

Die Differenz $T-\theta$ wird als **Schätzfehler** und die Differenz $E(T)-\theta$ als **Verzerrung** oder **Bias** bezeichnet. Für erwartungstreue Schätzfunktionen ist der Bias gleich Null und die Forderung nach minimaler mittleren quadratischen Abweichung ist in dem Fall identisch mit der Forderung nach minimaler Varianz. Anders formuliert: Bei der Bestimmung einer effizienten Schätzfunktion entspricht die Forderung nach minimaler Varianz der Forderung nach minimaler mittleren quadratischen Abweichung, da in dem Fall nur erwartungstreue Schätzfunktionen betrachtet werden.

Schätzfehler (Punktschätzung)

Verzerrung, Bias

Konsistenz:

Eine Schätzfunktion T für den Parameter θ heißt **konsistent im quadratischen Mittel**, wenn gilt:

$$\lim_{n\to\infty} \mathrm{MSE}(T) = 0.$$

Eine Schätzfunktion T für den Parameter θ einer Grundgesamtheit heißt **schwach konsistent**, wenn zu beliebigem $\epsilon > 0$ gilt:

$$\lim_{n\to\infty} \boldsymbol{P}(|T-\theta| \geq \epsilon) = 0.$$

Konsistenz im quadratischen Mittel

schwache Konsistenz

Eine Schätzfunktion heißt schwach konsistent, wenn der Schätzer für große Stichprobenumfänge in der Nähe von θ liegt bzw. die Wahrscheinlichkeit, dass T von θ um mehr als ϵ abweicht, gegen Null konvergiert. Für genügend große Stichprobenumfänge ist diese Wahrscheinlichkeit somit gering. Ist die Schätzfunktion T im quadratischen Mittel konsistent, so ist sie auch schwach konsistent. Eine Schätzfunktion kann somit als konsistent bezeichnet werden, wenn unter zunehmender Stichprobengröße die Schätzfunktion immer genauer wird.

2 Intervallschätzung

Mit einer hohen Wahrscheinlichkeit, stimmt ein Punktschätzer nicht mit dem wahren, unbekannten Parameter θ überein. Von daher ist es sinnvoll, einen Bereich anzugeben, der mit einer hohen Wahrscheinlichkeit den unbekannten Parameter θ überdeckt bzw. mit einer vorgegebenen niedrigen Wahrscheinlichkeit α den Parameter θ nicht überdeckt. Hierbei wird α als **Irrtumswahrscheinlichkeit** bezeichnet und üblicherweise mit 0.1, 0.05 oder 0.01 angegeben.

Irrtums-wahrscheinlichkeit

Konfidenzintervall

Ein Bereich mit solchen Eigenschaften wird **Konfidenzintervall** genannt.

Gegeben sind zwei Stichprobenfunktionen $O = g_1(X_1, ..., X_n)$ und $U = g_2(X_1, ..., X_n)$, für die stets gilt $U < O$. Da U und O Zufallsvariablen sind, bilden sie Grenzen eines Zufallsintervalls $[U; O]$. Mit

$$P(U \leq \theta \leq O) = 1 - \alpha,$$

heißt $[U; O]$ **zweiseitiges Konfidenzintervall** (Vertrauensintervall) für θ zum **Konfidenzniveau $\mathbf{1 - \alpha}$** bei vorgegebener Irrtumswahrscheinlichkeit α. Mit U und O werden die **Konfidenzgrenzen** (Vertrauensgrenzen) bezeichnet. Gilt für das Zufallsintervall $[U; \infty)$ bzw. $(-\infty; O]$

$$P(\theta \geq U) = 1 - \alpha \text{ bzw. } P(\theta \leq O) = 1 - \alpha,$$

dann handelt es sich um ein **einseitiges Konfidenzintervall** für θ.

zweiseitiges Konfidenzintervall, Konfidenzniveau, Konfidenzgrenze

einseitiges Konfidenzintervall

Die Bestimmung des Intervalls bei vorgegebener Irrtumswahrscheinlichkeit α heißt auch Intervallschätzung für den Parameter θ. Für die Interpretation der Grenzen des Konfidenzintervalls ist Folgendes zu beachten:

Die Grenzen des Intervalls werden aus den Werten einer konkreten Stichprobe berechnet. Diese Werte sind Realisationen von Zufallsvariablen und für die Grenzen des Konfidenzintervalls ergeben sich feste Werte. Das ändert nichts an der Tatsache, dass generell ein Konfidenzintervall ein Zufallsintervall ist. Folgende Abbildung, welche sechs Simulationen einer standardnormalverteilten Zufallsvariablen zeigt, veranschaulicht diese Tatsache.

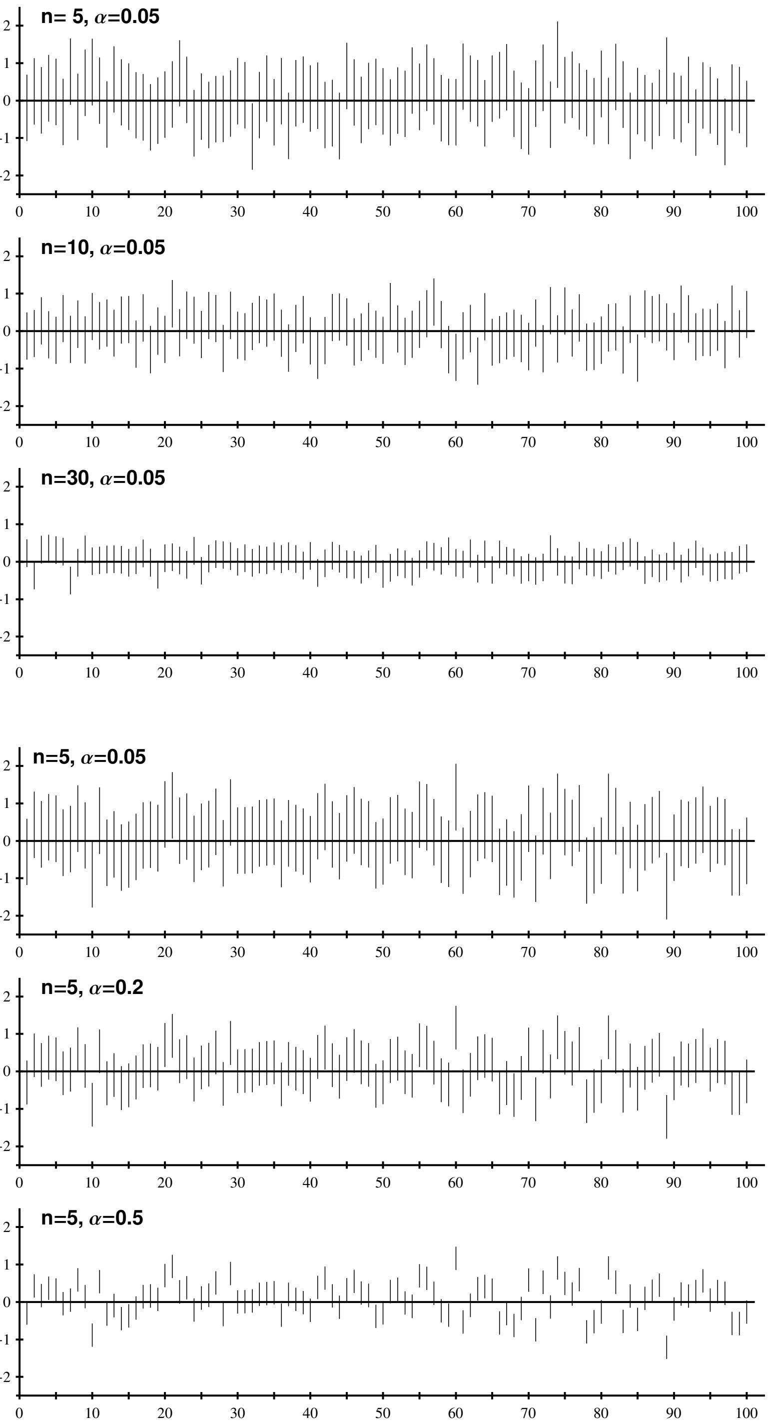

Abbildung 2.0.1: *Konfidenzintervalle als Realisationen eines Zufallsintervalls, 100 Konfidenzintervalle für $X \sim N(0,1)$ und Stichprobenumfang n*

Der in diesen Simulationen bekannte Parameterwert der Grundgesamtheit $\mu = 0$ entspricht der waagerechten Linie. Für die 100 Stichproben sind, als senkrechte Linien, die sich ergebenden Konfidenzintervalle eingezeichnet. Es ist zu erkennen, dass bei steigender Stichprobengröße (α fest) die Länge des Konfidenzintervalls abnimmt. Bei einer Vergrößerung von α (n fest), nimmt die Länge des Konfidenzintervalls ebenfalls ab und es ist zu erkennen, dass bei einem großen α weniger Konfidenzintervalle den wahren Wert überdecken.

> Werden viele Stichproben gezogen, so wird erwartet, dass bei einem $(1-\alpha)\cdot 100\%$-Konfidenzintervall $(1-\alpha)\cdot 100\%$ der Intervalle den unbekannten, aber festen Parameterwert enthalten.

Das hier angewandte Schätzverfahren führt somit mit einer Wahrscheinlichkeit von $1-\alpha$ zu Konfidenzintervallen, die den wahren, unbekannten Wert enthalten. Dies bedeutet jedoch nicht, dass für ein konkret realisiertes Konfidenzintervall der wahre Wert mit einer Wahrscheinlichkeit von $1-\alpha$ enthalten ist. Der Parameter θ kann entweder zwischen den beiden Grenzen des konkret realisierten Konfidenzintervalls liegen oder, sofern er nicht mit einer Grenze zusammenfällt, außerhalb des Intervalls.

> Ein **realisiertes** Konfidenzintervall enthält den unbekannten Parameter θ oder nicht. Eine Wahrscheinlichkeitsaussage der Form „Das Konfidenzintervall enthält mit einer Wahrscheinlichkeit von $1-\alpha$ den Parameter θ“ ist **nicht** mehr möglich.

symmetrisches Konfidenzintervall

Gilt für ein zweiseitiges Konfidenzintervall $\boldsymbol{P}(\theta < U) = \boldsymbol{P}(\theta > O) = \frac{\alpha}{2}$, d.h. ist die Wahrscheinlichkeit dafür, dass θ kleiner als die untere Grenze U ist, genau so groß wie die Wahrscheinlichkeit, dass θ größer als die obere Grenze O ist, so heißt das Konfidenzintervall **symmetrisch bezüglich der Wahrscheinlichkeit**.

Schätzfehler (Intervallschätzung)

Oft liegen die obere und die untere Grenze auch **symmetrisch bezüglich des Punktschätzers** $\hat{\Theta}$. Für ein bezüglich des Punktschätzers $\hat{\Theta}$ symmetrisches Konfidenzintervall wird die halbe Länge des Konfidenzintervalls auch als **Schätzfehler** bezeichnet. Liegen die beiden Grenzen eines Konfidenzintervalls nicht symmetrisch zu $\hat{\Theta}$, dann bezeichnet der größte Abstand der beiden Grenzen zu $T = \hat{\Theta}$ den Schätzfehler. Die Länge eines Konfidenzintervalls berechnet sich als Differenz $O - U$.

2.1 Konfidenzintervall für den Parameter μ eines quantitativen Merkmals

Die in diesem Abschnitt hergeleiteten Konfidenzintervalle sind symmetrisch bezüglich der Wahrscheinlichkeit und bezüglich des Punktschätzers $\hat{\mu} = \overline{X}$.

2.1.1 Konfidenzintervalle bei bekannter Standardabweichung eines normalverteilten Merkmals

Gegeben sei die Grundgesamtheit der Größe N eines $N(\mu, \sigma^2)$-verteilten Merkmals X unter der Voraussetzung, dass $\boldsymbol{\sigma}$ **bekannt** ist. Weiter sei vorausgesetzt, dass es sich im Folgenden um **Stichproben mit Zurücklegen** handelt bzw. die Grundgesamtheit als unendlich betrachtet werden kann, d.h. für den Stichprobenumfang n gilt $n/N < 0.05$. Ausgehend von dem Punktschätzer $\hat{\mu} = \overline{X}$ wird zunächst ein zweiseitiges Konfidenzintervall für den Parameter μ zum Niveau $1-\alpha$ hergeleitet.

Die Stichprobenfunktion $\overline{X}$ ist unter den obigen Voraussetzungen eine normalverteilte Zufallsvariable mit dem Erwartungswert μ und der Varianz $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$, d.h.

$$\overline{X} \sim N(\mu, \tfrac{\sigma^2}{n}) \quad \text{bzw.} \quad Z = \sqrt{n}\,\frac{\overline{X} - \mu}{\sigma} \sim N(0,1).$$

Für die Stichprobenfunktion Z lässt sich mit dem Quantil $z_{1-\frac{\alpha}{2}}$ der Standardnormalverteilung direkt ein zweiseitiges Intervall $[-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}]$ angeben, für das gilt

$$\boldsymbol{P}\left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n}\,\frac{\overline{X} - \mu}{\sigma} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha,$$

wobei $z_{1-\frac{\alpha}{2}}$ dem Wert entspricht, bei dem die Verteilungsfunktion der Standardnormalverteilung (bzw. die Fläche unter der Dichtefunktion bis zu diesem Punkt) den Wert $1 - \frac{\alpha}{2}$ annimmt. Durch Umformung ergibt sich

$$\begin{aligned}
\boldsymbol{P}\left(\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) &= 1 - \alpha \text{ bzw.}\\
\boldsymbol{P}\left(\overline{X} - z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}} \leq \mu \leq \overline{X} + z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}}\right) &= 1 - \alpha.
\end{aligned}$$

Die in der Wahrscheinlichkeitsaussage ermittelten Grenzen $\overline{X} - z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}}$ bzw. $\overline{X} + z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}}$ erfüllen somit die Bedingungen für ein zweiseitiges Konfidenzintervall.

zweiseitige Konfidenzintervall für μ

Gegeben sei eine normalverteilte Zufallsvariable X mit unbekanntem Parameter μ und **bekannter** Varianz σ^2, d.h. $X \sim N(\mu, \sigma^2)$. Das **zweiseitige $(1-\alpha)$-Konfidenzintervall** für μ lautet:

$$\left[\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\ ;\ \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right].$$

Beispiel 2.1.1:

Aus einer Grundgesamtheit mit $N(\mu, 144)$-verteiltem X wurde eine einfache Zufallsstichprobe vom Umfang $n = 36$ gezogen, die $\overline{x} = 26$ liefert. Für $1 - \alpha = 0.95$ ergibt sich für die Standardnormalverteilung das Quantil $z_{0.975} = 1.96$. Gesucht sind die Grenzen des zweiseitigen 95%-Konfidenzintervalls für μ.

$$\mu_u = 26 - 1.96\frac{12}{\sqrt{36}} \qquad \mu_o = 26 + 1.96\frac{12}{\sqrt{36}}$$
$$\mu_u = 22.08 \qquad \mu_o = 29.92$$

Einseitige Konfidenzintervalle ergeben sich aus den obigen Ansätzen bei Verwendung jeweils einer (oberen oder unteren) Grenze und Verwendung des entsprechenden z-Wertes.

einseitige Konfidenzintervalle für μ

Gegeben sei eine normalverteilte Zufallsvariable X mit unbekanntem Parameter μ und **bekannter** Varianz σ^2, d.h. $X \sim N(\mu, \sigma^2)$. Die **einseitigen $(1-\alpha)$-Konfidenzintervalle** für μ lauten:

$$\left[\overline{X} - z_{1-\alpha} \frac{\sigma}{\sqrt{n}}\ ;\ \infty\right) \text{ bzw. } \left(-\infty\ ;\ \overline{X} + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}\right].$$

Beispiel 2.1.2:

Eine Probeserie von 100 Autoreifen, von der angenommen wird, dass es sich um eine einfache Zufallsstichprobe handelt, liefert eine durchschnittliche Lebensdauer von $\overline{x} = 40000$ km bei einer Standardabweichung von $\sigma = 8000$ km.

Unter der Normalverteillungsannahme soll mittels eines einseitigen Konfidenzintervalls geschätzt werden, wie hoch die durchschnittliche Lebensdauer der Autoreifen mit einer Wahrscheinlichkeit von 99% mindestens ist. Es is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{8000}{\sqrt{100}} = 800 \quad \textit{und für } \alpha = 0.01 \textit{ gilt } z_{1-\alpha} = 2.33.$$

Damit ergibt sich als gesuchte Grenze

$$\begin{aligned} \mu_u &= \overline{x} - z_{1-\alpha}\sigma_{\overline{X}} = 40000 - 2.33 \cdot 800 \\ &= 40000 - 1864 = 38136. \end{aligned}$$

Als **Schätzfehler** der zweiseitigen Konfidenzintervalle ergibt sich $z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}} = z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$. Der Schätzfehler hängt vom Stichprobenumfang n und der vorgegebenen Inrrtumswahrscheinlichkeit α ab. Die Länge eines Konfidenzintervalls entspricht dem zweifachen des Schätzfehlers. Für die **Länge der zweiseitigen Konfidenzintervalle** gilt (siehe Abbildung 2.0.1 und [1]):

Schätzfehler

- Mit zunehmendem Stichprobenumfang n, d.h. bei zunehmender Information, nimmt die Länge des Konfidenzintervalls ab.

- Mit wachsendem Konfidenzniveau $1 - \alpha$, d.h. bei zunehmender Sicherheit $1 - \alpha$ bzw. abnehmender Irrtumswahrscheinlichkeit α, nimmt die Länge des Konfidenzintervalls zu.

Liegt eine **Stichprobe ohne Zurücklegen** aus einer endlichen Grundgesamtheit mit $\frac{n}{N} \geq 0.05$ vor, berechnet sich die Varianz des Stichprobenmittelwerts $\overline{X}$ zu:

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$$

Der Faktor $\frac{N-n}{N-1}$ kann als **Endlichkeitskorrektur** betrachtet werden. In diesem Fall wird die Varianz $\frac{\sigma^2}{n}$ durch die Varianz $\frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$ ersetzt. Z.B. lautet das zweiseitige Konfidenzintervall mit Berücksichtigung der Endlichkeitskorrektur:

Endlichkeitskorrektur

$$\boxed{\left[\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\,;\,\overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}\right].}$$

zweiseitige Konfidenzintervall für μ mit Endlichkeitskorrektur

[1] siehe interaktive Mathematica-Applets auf der Homepage des Lehrstuhls http://www.fernuni-hagen.de/ls_statistik/forschung/multimedia/

2.1.2 Konfidenzintervalle bei unbekannter Standardabweichung eines normalverteilten Merkmals

Auch hier sei vorausgesetzt, dass ein normalverteiltes Merkmal X mit den Parametern μ und σ^2 vorliegt und es sich im Folgenden um Stichproben mit Zurücklegen handelt bzw. die Grundgesamtheit als unendlich betrachtet werden kann. Ist $\boldsymbol{\sigma}$ **unbekannt**, wird σ^2 durch die Stichprobenvarianz $S^2 = \frac{1}{n-1}\sum_i^n (X_i - \overline{X})^2$ ersetzt und der Schätzer für $\sigma_{\overline{X}}$ lautet:

Stichprobenvarianz

$$\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}}.$$

Trotz normalverteilter Grundgesamtheit ist $(\overline{X} - \mu)/\hat{\sigma}_{\overline{X}}$ nicht mehr normalverteilt. Die Stichprobenfunktion

$$\sqrt{n}\,\frac{\overline{X} - \mu}{S}$$

ist t-verteilt mit $n - 1$ Freiheitsgraden[2]. Die Konfidenzintervalle ergeben sich analog zu den vorherigen Intervallen, indem die Quantile der Normalverteilung durch die Quantile der t-Verteilung ersetzt werden.

Konfidenzintervalle für $\boldsymbol{\mu}$

Gegeben sei eine normalverteilte Zufallsvariable X mit unbekanntem Parameter μ und **unbekannter** Varianz σ^2, d.h. $X \sim N(\mu, \sigma^2)$. Das **zweiseitige $(1-\boldsymbol{\alpha})$-Konfidenzintervall** für μ lautet:

zweiseitig

$$\left[\overline{X} - t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\ ;\ \overline{X} + t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right]$$

Die **einseitigen $(1-\boldsymbol{\alpha})$-Konfidenzintervalle** ergeben sich zu:

einseitig

$$\left[\overline{X} - t_{1-\alpha}(n-1)\frac{S}{\sqrt{n}}\ ;\ \infty\right) \text{ bzw. } \left(-\infty\ ;\ \overline{X} + t_{1-\alpha}(n-1)\frac{S}{\sqrt{n}}\right]$$

Im Gesamtglossar kann aus der Tabelle direkt das Quantil $t_{1-\alpha}(n-1)$ mit $\boldsymbol{P}(T \le t) = 1 - \alpha$ abgelesen werden. Die Wahrscheinlichkeit $\boldsymbol{P}(T \le$

[2] Die Anzahl der Freiheitsgrade *df* (degrees of freedom) einer Stichprobenfunktion hängt von der Anzahl der unabhängigen Zufallsvariablen X_i ab. Die Schätzfunktion $S = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$ besitzt nur $n-1$ Freiheitsgrade, da S von $\overline{X}$ abhängt bzw. den Wert $\overline{X}$ voraussetzt. Es können nur $n-1$ Zufallsvariablen frei gewählt werden, dass sich der vorgegebene Mittelwert ergibt. Die n-te Zufallsvariable ist durch die Bedingung „Mittelwert ist vorgegeben“ eindeutig festgelegt.

$t) = F(t) = 1-\alpha$ bezeichnet ein einseitiges Intervall, und es gilt aufgrund der Symmetrieeigenschaft der t-Verteilung

$$\boldsymbol{P}(T < t) = \boldsymbol{P}(T > -t) = 1 - \alpha$$

.
Wird ein zweiseitig abgegrenztes Intervall gesucht, für das gilt

$$\boldsymbol{P}(-t \leq T \leq t) = 1 - \alpha,$$

so ist der entsprechende t-Wert in der Tabelle zu $1 - \frac{\alpha}{2}$ nachzuschlagen, denn aus

$$\boldsymbol{P}(T \leq t) = 1 - \frac{\alpha}{2}$$

ergibt sich aufgrund der Symmetrieeigenschaft

$$\boldsymbol{P}(T \leq -t) = \boldsymbol{P}(T \geq t) = 1 - \boldsymbol{P}(T \leq t) = 1 - \left(1 - \frac{\alpha}{2}\right) = \frac{\alpha}{2}$$

und damit

$$\boldsymbol{P}(-t \leq T \leq t) = F(t) - F(-t) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha.$$

Beispiel 2.1.3:

Für eine t-Verteilung mit 10 Freiheitsgraden gilt:

$$\begin{aligned} & \boldsymbol{P}(T \geq -1.812) = \boldsymbol{P}(T \leq 1.812) = 0.95 \\ \textit{aber} \quad & \boldsymbol{P}(-1.812 \leq T \leq 1.812) = 0.9. \end{aligned}$$

Weiter gilt bei der Bestimmung von Intervallgrenzen für Konfidenzintervalle zum Niveau 0.9:

$$\begin{aligned} \text{einseitig:} \quad & \boldsymbol{P}(T \geq -1.372) = \boldsymbol{P}(T \leq 1.372) = 0.9 \\ \text{zweiseitig:} \quad & \boldsymbol{P}(-1.812 \leq T \leq 1.812) = 0.9 \end{aligned}$$

Wird für eine Stichprobe mit $n = 11$ ein einseitiges Intervall zum Niveau 0.9 gesucht, so kann direkt in der Tabelle im Gesamtglossar bei 10 Freiheitsgraden und 0.9 nachgeschlagen werden. Ist weiterhin $\alpha = 0.1$ gegeben, jedoch wird ein zweiseitiges Intervall gesucht, ist bei 10 Freiheitsgraden und der gegebenen Wahrscheinlichkeit $P(T < t) = 1 - \frac{\alpha}{2} = 0.95$ nachzuschlagen. Es ergibt sich dann die obere Grenze des zweiseitigen Intervalls. Aufgrund der Symmetrieeigenschaft der t-Verteilung entspricht die untere Grenze des gesuchten Intervalls der negativen oberen Grenze. Dieser Zusammenhang wird in der folgenden Grafik verdeutlicht.

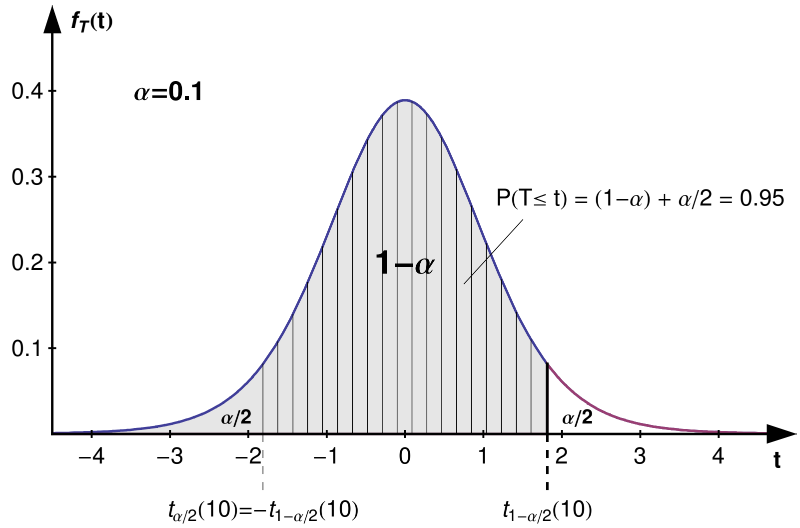

Abbildung 2.1.1: *Zusammenhang einseitiger und zweiseitiger Konfidenzgrenzen bei Vorliegen einer t-Verteilung*

Beispiel 2.1.4:

Eine aus einer Grundgesamtheit mit normalverteiltem X gezogene einfache Zufallsstichprobe vom Umfang $n = 16$ ergibt $\overline{x} = 5$ und $s^2 = 25$. Gesucht ist das 99%-Konfidenzintervall für μ. Aus der Tabelle der Student-Verteilung wird für $n-1 = 15$ Freiheitsgraden und für $1-\alpha = 0.99$ (zweiseitiges Konfidenzintervall!) der Wert $t_{1-\frac{\alpha}{2}}(n-1) = t_{0.995}(15) = 2.947$ abgelesen.

$$\mu_u = 5 - 2.947\tfrac{5}{\sqrt{16}} = 1.32 \qquad \mu_o = 5 + 2.947\tfrac{5}{\sqrt{16}} = 8.68$$

Liegt eine **Stichprobe ohne Zurücklegen** aus einer endlichen Grundgesamtheit mit $\frac{n}{N} \geq 0.05$ vor und σ ist unbekannt, berechnet sich der Schätzer $\hat{\sigma}_{\overline{X}}$ zu:

$$\hat{\sigma}_{\overline{X}} = \frac{S}{n} \cdot \left(\frac{N-n}{N}\right) = \frac{S}{n} \cdot \left(1 - \frac{n}{N}\right)$$

Endlichkeitskorrektur

Auswahlsatz

Der Faktor $(1 - \frac{n}{N})$ wird als **Endlichkeitskorrektur** bezeichnet. Der in der Endlichkeitskorrektur $1 - \frac{n}{N}$ berücksichtigte Bruch $\frac{n}{N}$ wird in der Literatur oft als **Auswahlsatz f** definiert. Im Fall einer Stichprobe ohne Zurücklegen aus einer endlichen Grundgesamtheit mit $\frac{n}{N} \geq 0.05$ ergibt sich somit ein Konfidenzintervall mit Berücksichtigung dieser Endlichkeitskorrektur, z.B.:

$$\left[\overline{X} - t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\sqrt{1-\frac{n}{N}}\ ;\ \overline{X} + t_{1-\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\sqrt{1-\frac{n}{N}}\right]$$

zweiseitiges Konfidenzintervall für μ mit Endlichkeitskorrektur

2.1.3 Konfidenzintervalle für beliebig verteilte Merkmale

Auf die Voraussetzung einer normalverteilten Grundgesamtheit kann verzichtet werden, wenn der Stichprobenumfang über 30 liegt. In allen praktischen Anwendungen kann dann aufgrund des sogenannten zentralen Grenzwertsatzes davon ausgegangen werden, dass auch bei nicht normalverteilter Grundgesamtheit die Stichprobenfunktion $\overline{X}$ approximativ normalverteilt ist. Betrachtet wird hier nur der Fall einer Stichprobe mit Zurücklegen bzw. $n/N < 0.05$. Im Fall ohne Zurücklegen und $n/N \geq 0.05$ muss die Endlichkeitskorrektur $\frac{N-n}{N-1}$ bzw. $\frac{N-n}{N}$ berücksichtigt werden.

Liegt ein Stichprobenumfang mit $n > 30$ vor, ergibt sich für ein beliebig verteiltes Merkmal X ein **approximatives $(1-\alpha)$-Konfidenzintervall** für μ durch:

$$\boldsymbol{\sigma}\ \textbf{bekannt}: \quad \left[\overline{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\ ;\ \overline{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$$

$$\boldsymbol{\sigma}\ \textbf{unbekannt}: \quad \left[\overline{X} - z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\ ;\ \overline{X} + z_{1-\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\right]$$

approximatives zweiseitiges Konfidenzintervall für μ bei beliebig verteiltem Merkmal

Das untere Konfidenzintervall kann in der entsprechenden Form (einseitig/zweiseitig, mit/ohne Endlichkeitskorrektur) für den Fall $n > 30$ in Abschnitt 2.1.2 die dort verwendeten Konfidenzintervalle ersetzen, wobei zu beachten ist, dass die hier angegebenen Intervalle approximativ sind.

Beispiel 2.1.5:

Aus einer normalverteilten Grundgesamtheit vom Umfang $N = 10000$ wird eine Stichprobe vom Umfang $n = 100$ gezogen. Es ist $\overline{x} = 120$ und $s^2 = 80$. Mit $n/N = 0.01 < 0.05$ ist

$$\hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{80}{100}} = \sqrt{0.8} = 0.89.$$

Da $n > 30$ *ist, kann näherungsweise die Normalverteilung anstelle der Student-Verteilung angewendet werden. Für das zweiseitige* 98%-*Konfidenzintervall, d.h.* $\alpha = 0.2$*, ergeben sich mit* $z_{1-\frac{\alpha}{2}} = z_{0.99} = 2.33$ *die Grenzen*

$$\mu_u = 120 - 2.33 \cdot 0.89 = 117.93,$$

$$\mu_o = 120 + 2.33 \cdot 0.89 = 122.07.$$

Anmerkung:

Varianz von $\overline{X}$

Unter der Normalverteilungsannahme wird hier zur Berechnung eines Konfidenzintervalls für den Parameter μ der Punktschätzer $\overline{X}$ herangezogen. Dazu wird die Verteilung des Stichprobenmittelwertes $\overline{X}$ benötigt. Der Erwartungswert von $\overline{X}$ entspricht μ. In der folgenden Tabelle ist zusammenfassend dargestellt, welche Varianz bzw. welcher Varianzschätzer zu verwenden ist:

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Varianz σ^2 der Grundgesamtheit</td></tr>
<tr><td>bekannt</td><td>unbekannt (Schätzung)</td></tr>
<tr><td colspan="2">Stichprobe mit Zurücklegen</td><td rowspan="2">$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$</td><td rowspan="2">$\hat{\sigma}^2_{\overline{X}} = \frac{S^2}{n}$</td></tr>
<tr><td rowspan="2">Stichprobe ohne Zurücklegen</td><td>$\frac{n}{N} < 0.05$</td></tr>
<tr><td>$\frac{n}{N} \geq 0.05$</td><td>$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} \frac{N-n}{N-1}$</td><td>$\hat{\sigma}^2_{\overline{X}} = \frac{S^2}{n} \frac{N-n}{N}$</td></tr>
</table>

2.2 Konfidenzintervall für den Anteilswert $\boldsymbol{\pi}$ einer dichotomen Grundgesamtheit

Gegeben sei eine dichotome Grundgesamtheit $\Omega = \{\omega_1, ..., \omega_N\}$, deren Elemente nur die zwei möglichen Ausprägungen $A, \overline{A}$ aufweisen. Jedes Element kann als dichotome Zufallsvariable X_i mit

$$X_i = X(\omega_i) = \begin{cases} 1 & \text{für } \omega_i \text{ weist die Eigenschaft } A \text{ auf} \\ 0 & \text{für } \omega_i \text{ weist die Eigenschaft } \overline{A} \text{ auf} \end{cases}$$

aufgefasst werden. Für die Wahrscheinlichkeitsverteilung ergibt sich

$$\boldsymbol{P}(X_i = x_i) = \begin{cases} \pi & \text{für } x_i = 1 \\ 1 - \pi & \text{für } x_i = 0 \end{cases}$$

mit $E(X_i) = \pi$ und $\operatorname{Var}(X_i) = \pi(1 - \pi)$.

erwartungstreue Schätzfunktion für $\boldsymbol{\pi}$

Eine **erwartungstreue Schätzfunktion für $\boldsymbol{\pi}$** aus einer Stichprobe $X_1, ..., X_n$ vom Umfang n ist die Stichprobenfunktion

$$P = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Da die Zufallsvariablen X_i nur die Werte 0 und 1 annehmen können, handelt es sich bei P um die relative Häufigkeit für das Auftreten von A in der Stichprobe. Für den Fall einer Stichprobe mit Zurücklegen bzw. einer unendlichen Grundgesamtheit haben alle X_i dieselbe oben angegebene Verteilung (Bernoulli-Verteilung). Die Summe $X = \sum_{i=1}^{n} X_i$ ist $B(n, \pi)$-verteilt mit $E(X) = n\pi$ und $\operatorname{Var}(X) = n\pi(1 - \pi)$ (siehe Binomialverteilung, KE 2, Abschnitt 3.2).

Parameter der Schätzfunktion P

$$\begin{aligned} E(P) &= E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n} \cdot n\pi = \pi \\ \operatorname{Var}(P) &= \frac{1}{n^2}\operatorname{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2} \cdot n\pi(1 - \pi) = \frac{\pi(1 - \pi)}{n} \end{aligned}$$

Für hinreichend großes n (Faustregel: $n\pi \geq 5, n(1-\pi) \geq 5$ bzw. $np \geq 5, n(1-p) \geq 5$) ist P approximativ $N(\pi, \frac{\pi(1-\pi)}{n})$-verteilt. Der Wert p entspricht der in der Stichprobe beobachteten relativen Häufigkeit für das Auftreten der Eigenschaft A. Damit kann ein Intervall für π zum Niveau $1-\alpha$ mittels

$$\boldsymbol{P}\left(P - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}} \leq \pi \leq P + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}}\right) \approx 1-\alpha$$

bestimmt werden. Die daraus resultierenden Intervallgrenzen enthalten noch den unbekannten Anteilswert π. Da die Binomialverteilung mit zunehmendem Stichprobenumfang schnell gegen die Normalverteilung konvergiert, kann ein approximatives Konfidenzintervall angegeben werden, indem π durch den Schätzer P ersetzt wird.

approximatives zweiseitiges Konfidenzintervall für π

Für dichotome Grundgesamtheiten und für große Stichprobenumfänge ($n \geq 30$) lässt sich ein **approximatives zweiseitiges Konfidenzintervall**, welches symmetrisch bezüglich des Punktschätzers P ist, für den Anteilswert π durch

$$\left[P - z_{1-\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}}; P + z_{1-\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}}\right]$$

angeben, wobei der Stichprobenanteil $P = \overline{X}$ dem Schätzer $\hat{\pi}$ entspricht.

Beispiel 2.2.1:

In einer Umfrage stimmten in einer Zufallsstichprobe von 225 *Studenten einer Universität* 45 *für die Einführung mündlicher Nachprüfungen. Da die Grundgesamtheit etwa* 8000 *Studenten umfasst* ($n/N = 0.028$) *liegt der zufälligen Entnahme (ohne Zurücklegen) näherungsweise eine Binomialverteilung mit dem Anteilswert* π *zugrunde. Wegen*

$$p = \frac{45}{225} = 0.2 \quad \text{und} \quad np = 45 \geq 5, n(1-p) = 180 \geq 5$$

kann diese durch die Normalverteilung approximiert werden. Außerdem ist die Forderung $n \geq 30$ *erfüllt. Die Grenzen des approximativen zweiseitigen Konfidenzintervalls für* π *mit* $\alpha = 0.1$ *ergeben sich zu*

$$\begin{aligned}\pi_u &= p - z_{0.95}\sqrt{\frac{p(1-p)}{n}} = 0.2 - 1.65 \cdot \sqrt{\frac{0.16}{225}} = 0.2 - 0.044 = 0.156,\\ \pi_o &= p + z_{0.95}\sqrt{\frac{p(1-p)}{n}} = 0.2 + 0.044 = 0.244.\end{aligned}$$

2.3 Konfidenzintervall für den Parameter σ^2 eines normalverteilten Merkmals

Gegeben sei eine Grundgesamtheit mit $N(\mu, \sigma^2)$-verteiltem Merkmal X, wobei **μ und σ^2 unbekannt** sind. Ausgehend von dem Punktschätzer $S^2 = \frac{1}{n-1}\sum_i^n (X_i - \overline{X})^2$ lassen sich Konfidenzintervalle für σ^2 herleiten, indem eine Statistik konstruiert wird, welche χ^2-verteilt ist mit $n-1$ Freiheitsgraden.

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$$

Mit den Quantilen

$$\chi^2_{\frac{\alpha}{2}}(n-1) \text{ und } \chi^2_{1-\frac{\alpha}{2}}(n-1)$$

der χ^2-Verteilung mit $n-1$ Freiheitsgraden gilt:

$$\begin{aligned} 1-\alpha &= \boldsymbol{P}\left(\chi^2_{\frac{\alpha}{2}}(n-1) \leq \frac{n-1}{\sigma^2} S^2 \leq \chi^2_{1-\frac{\alpha}{2}}(n-1)\right) \\ &= \boldsymbol{P}\left(\frac{\chi^2_{\frac{\alpha}{2}}(n-1)}{(n-1)S^2} \leq \frac{1}{\sigma^2} \leq \frac{\chi^2_{1-\frac{\alpha}{2}}(n-1)}{(n-1)S^2}\right) \\ &= \boldsymbol{P}\left(\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right) \end{aligned}$$

Konfidenzintervall für σ^2 bei unbekanntem μ

Gegeben sei eine normalverteilte Zufallsvariable X mit unbekannter Varianz σ^2 und unbekanntem Erwartungswert μ, d.h. $X \sim N(\mu, \sigma^2)$. Das **zweiseitige $(1-\alpha)$-Konfidenzintervall** für σ^2 lautet:

zweiseitig

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} \,;\, \frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}\right]$$

Die **einseitigen $(1-\alpha)$-Konfidenzintervalle** ergeben sich zu:

einseitig

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)} \,;\, \infty\right) \text{ bzw. } \left(0 \,;\, \frac{(n-1)S^2}{\chi^2_{\alpha}(n-1)}\right]$$

Das hier angegebene zweiseitige Konfidenzintervall ist bezüglich der Wahrscheinlichkeit symmetrisch. Da die Chi-Quadrat-Verteilung keine symmetrische Verteilung ist (s. Kurseinheit 2), ist das obige Konfidenzintervall **nicht** symmetrisch bezüglich des Punktschätzers S^2.

Beispiel 2.3.1:

Untersucht wird das Haushaltsnettoeinkommen von $N = 5000$ *Privathaushalten. Dazu wird eine Stichprobe vom Umfang* $n = 100$ *betrachtet. Als Stichprobenvarianz ergibt sich* $s^2 = 950625$*. Bestimmt werden soll das zweiseitige 95%-Konfidenzintervall für* σ^2*, d.h.* $\alpha = 0.05$*. Mit den Quantilen* $\chi^2_{0.025}(99) = 73.361$ *und* $\chi^2_{0.975}(99) = 128.422$ *ergeben sich die Grenzen*

$$\sigma_o^2 = \frac{99 \cdot 950625}{73.361} = 1282859.762,$$

$$\sigma_u^2 = \frac{99 \cdot 950625}{128.422} = 732832.965.$$

Anhand der Grenzen ist zu erkennen, dass das Konfidenzintervall nicht symmetrisch bezüglich $s^2 = 950625$ *ist.*

Anmerkung:

Ist der Parameter $\boldsymbol{\mu}$ der Grundgesamtheit **bekannt**, so kann zur Bestimmung der Konfidenzgrenzen der Schätzer $S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ verwendet werden. Die Statistik $\frac{nS^{*2}}{\sigma^2}$ ist χ^2-verteilt mit n Freiheitsgraden, so dass sich in dem Fall nachstehendes zweiseitiges Konfidenzintervall ergibt.

zweiseitiges Konfidenzintervall für $\boldsymbol{\sigma^2}$ bei bekanntem $\boldsymbol{\mu}$

$$\left[\frac{nS^{*2}}{\chi^2_{1-\frac{\alpha}{2}}(n)}\,;\,\frac{nS^{*2}}{\chi^2_{\frac{\alpha}{2}}(n)}\right]$$

3 Grundlagen statistischer Testverfahren

Zur Prüfung bestimmter Vermutungen über Parameter oder auch über die Verteilung einer Grundgesamtheit werden **statistische Testverfahren**, oft auch kurz als **Test** bezeichnet, herangezogen. Untersucht werden Unterschiede zwischen einer Stichprobe und deren Grundgesamtheit oder zwischen zwei Stichproben unterschiedlicher Grundgesamtheiten. Die dabei getroffene Vermutung bzw. Annahme wird als **Hypothese** bezeichnet.

Test

Hypothese

Häufig liegt eine **Hypothese über den numerischen Wert eines unbekannten Parameters** vor, zum Beispiel eines Lage- oder Streuungsparameters. Die Prüfung einer derartigen Hypothese wird als **Parametertest** bezeichnet.

Parametertest

Werden **Hypothesen über den Typ einer Verteilung** eines Merkmals überprüft, zum Beispiel, ob eine Normalverteilung vorliegt, so wird von einem **Anpassungstest** gesprochen.

Anpassungstest

Mitunter werden auch **Hypothesen über die Abhängigkeit bzw. Unabhängigkeit** verschiedener Merkmale aufgestellt. Die Überprüfung kann dann mittels eines **Unabhängigkeitstests** erfolgen.

Unabhängigkeitstest

Bei der Durchführung eines Tests wird folgendes Schema verwendet, welches anhand einfacher Parametertests anschließend näher erläutert wird:

(1) Festlegung von Grundgesamtheit und Typ der Verteilung.

(2) Formulierung einer Hypothese und Gegenhypothese.

(3) Vorgabe einer Irrtumswahrscheinlichkeit.

(4) Festlegung einer Prüfgröße und Bestimmung der Verteilung dieser Prüfgröße unter Gültigkeit der Nullhypothese.

(5) Bestimmung kritischer Werte und Angabe eines Ablehnungsbereiches.

(6) Ziehung einer Stichprobe und Berechnung des Wertes der zugrundeliegenden Prüfgröße.

(7) Testentscheidung und Interpretation der Entscheidung.

3.1 Aufbau eines Parametertests

Der Parameter der Grundgesamtheit, über den eine Hypothese aufgestellt wird, wird im Folgenden **allgemein** mit dem Buchstaben θ bezeichnet. **Spezielle** Parameter sind beispielsweise μ, σ und π.

3.1.1 Festlegung von Grundgesamtheit und Typ der Verteilung

Für jeden Test sind vorab folgende Fragen über die Grundgesamtheit zu beantworten:

a) Handelt es sich um ein **quantitatives** oder ein **qualitatives** Merkmal?

b) Ist die Grundgesamtheit **endlich**?

c) Welche **Verteilung** hat die **Zufallsvariable $\boldsymbol{X}$** bei zufälliger Entnahme eines Elements aus der Grundgesamtheit? Diese Angabe ist manchmal nicht oder nur angenähert möglich.

Beispiel 3.1.1:

1) *Es soll überprüft werden, ob bei einem Würfel die „6“ tatsächlich mit der Wahrscheinlichkeit $\frac{1}{6}$ auftritt. Bezeichne X die Binärvariable mit den Realisationen „$1 \mathrel{\hat{=}} 6$ liegt oben“ und „$0 \mathrel{\hat{=}} 6$ liegt nicht oben“.*

   *Hierbei handelt es sich um ein **qualitatives** Merkmal (dichotome Grundgesamtheit). Die Grundgesamtheit besteht aus allen jemals durchgeführten oder noch durchzuführenden (**unendlich** vielen) Würfen mit diesem Würfel. Falls der Würfel „ideal“ ist, gilt:*

$$\boldsymbol{P}(X=1)=\frac{1}{6} \quad \textit{und} \quad \boldsymbol{P}(X=0)=\frac{5}{6}.$$

2) *Es soll überprüft werden, ob die mittlere Füllmenge von Wasserflaschen in einer Lieferung von 10000 Stück gleich 0.5 l ist.*

   *Es handelt sich um ein **quantitatives** (metrisch messbares) Merkmal. Die Grundgesamtheit hat einen Umfang von $N = 10000$. Für die Zufallsvariable X (Füllmenge einer zufällig entnommenen Flasche) kann näherungsweise eine **Normalverteilung** angenommen werden (siehe „Zentraler Grenzwertsatz“).*

3.1.2 Formulierung der Nullhypothese und Alternativhypothese

Ziel eines statistischen Tests ist die Überprüfung einer Hypothese (Behauptung, Vermutung), die sich ergeben kann aus

- einer Theorie,
- Erfahrung,
- einer Güteforderung oder Gütezusage.

In manchen Fällen ist diese Hypothese bereits so formuliert, dass sie direkt als sogenannte Nullhypothese für ein Testverfahren verwendet werden kann.

Die **Nullhypothese H_0** ist die mathematische Formulierung einer Vermutung und kann mit einem statistischen Test überprüft werden.

Nullhypothese

Oft ist es jedoch notwendig, aus der gegebenen Problemstellung erst eine passende Nullhypothese zu formulieren.

Beispiel 3.1.2:

Wenn behauptet wird, durch eine Materialänderung habe sich die durchschnittliche Lebensdauer von Autoreifen verändert, so kann über das eventuelle Ausmaß der vermuteten Änderung nichts gesagt werden. Als Nullhypothese wird deshalb formuliert: „Die durchschnittliche Lebensdauer hat sich ***nicht*** *verändert.“*

Im Wesentlichen werden zwei Typen von Nullhypothesen unterschieden, und zwar zwischen **zweiseitigen Nullhypothesen** (Punkthypothesen) und **einseitigen Nullhypothesen** (Bereichshypothesen).

zweiseitige/ einseitige Nullhypothese

Eine **zweiseitige Nullhypothese** wird immer dann verwendet, wenn behauptet wird, der Parameter θ einer Verteilung entspricht **einem ganz bestimmten Wert θ_0**. Daher die Bezeichnung **Punkthypothese**.

Punkthypothese

$$H_0: \ \theta = \theta_0$$

Eine **einseitige Nullhypothese** wird immer dann verwendet, wenn behauptet wird, dass der Parameter θ einer Verteilung **einen bestimmten Wert θ_0 nicht überschreitet bzw. nicht unterschreitet**, d.h. in einem bestimmten Bereich liegt. Daher der Begriff **Bereichshypothese**.

Bereichshypothese

$$H_0: \ \theta \leq \theta_0 \quad \text{bzw.} \quad H_0: \ \theta \geq \theta_0$$

Alternativhypothese

> Das Gegenteil einer Nullhypothese H_0 wird im Allgemeinen als **Alternativhypothese H_1** bezeichnet.

Formal lauten die Alternativhypothesen H_1 für zweiseitige bzw. einseitige Nullhypothesen H_0 somit:

H_0 zweiseitig:	$H_1: \ \theta \neq \theta_0$
H_0 einseitig:	$H_1: \ \theta > \theta_0 \quad \text{bzw.} \quad H_1: \ \theta < \theta_0$

Beispiel 3.1.3:

Ein Ottomotor soll mit Kolben des Durchmessers 70 *mm bestückt werden. Der Kolbenlieferant garantiert einen mittleren Durchmesser* $\mu = \mu_0 = 70$ *mm bei einer Standardabweichung von* $\sigma = 0.01$ *mm. Da jede große Abweichung nach oben oder unten zur Funktionsunfähigkeit führt, wird die Abnahmekontrolle formulieren:*

Nullhypothese:	$H_0: \ \mu = \mu_0 = 70 \ mm$
Alternativhypothese:	$H_1: \ \mu \neq \mu_0 = 70 \ mm$

Beispiel 3.1.4:

In einer Lieferung von 10000 *Glühlampen soll der Ausschussanteil* π *den Wert* $\pi_0 = 0.02$ *nicht überschreiten. Für den Abnehmer ist es nicht störend, wenn der tatsächliche Ausschussanteil noch niedriger liegt. Ihn interessiert nur, ob die Lieferung wegen zu viel Ausschuss reklamiert werden muss.*

Nullhypothese:	$H_0: \ \pi \leq \pi_0 = 0.02$
Alternativhypothese:	$H_1: \ \pi > \pi_0 = 0.02$

Ein Testproblem wird allgemein mit **H_0 vs. H_1** bezeichnet, wobei „vs." für „versus (gegen)" steht.

3.1.3 Vorgabe der Irrtumswahrscheinlichkeit und mögliche Fehlentscheidungen

Da jede Testentscheidung auf einer Stichprobenziehung basiert, ist diese somit zufällig und mit einem Risiko behaftet. Selbst bei richtiger Nullhypothese wird mit einer Wahrscheinlichkeit α die Nullhypothese abgelehnt. Ein Test wird so konstruiert, dass α beliebig gering vorgegeben werden kann, und zwar üblicherweise mit $0.01, 0.05$ oder 0.1.

> Die **Ablehnung einer zutreffenden Nullhypothese, obwohl sie richtig ist**, wird als **„Fehler 1. Art“** oder auch **„$\boldsymbol{\alpha}$-Fehler“** bezeichnet.
> Die vorgegebene Wahrscheinlichkeit $\boldsymbol{\alpha}$ heißt **Irrtumswahrscheinlichkeit** oder **Signifikanzniveau**.

Fehler 1. Art, α-Fehler

Irrtumswahrscheinlichkeit, Signifikanzniveau

Die Nichtablehnung der Nullhypothese bedeutet keineswegs, dass diese damit bewiesen oder richtig ist. Neben der Ablehnung einer richtigen Nullhypothese besteht noch eine andere Fehlermöglichkeit, nämlich eine falsche bzw. nicht zutreffende Nullhypothese nicht abzulehnen.

> Die Wahrscheinlichkeit $\boldsymbol{\beta}$, mit der **die Nullhypothese nicht abgelehnt wird, obwohl sie falsch ist**, heißt **Fehler 2. Art** oder **$\boldsymbol{\beta}$-Fehler**.

Fehler 2. Art, β-Fehler

Durch die Vorgabe von α kann der Fehler 1. Art kontrolliert werden. Daher kann allgemein formuliert werden:

> Das **Verwerfen der Nullhypothese $\boldsymbol{H_0}$** ist gleichbedeutend mit dem **statistischen Nachweis der Alternativhypothese $\boldsymbol{H_1}$** zum Niveau α.

Kann dagegen die Nullhypothese nicht verworfen werden, so beweist das gar nichts. Es ist nämlich nicht bekannt, wie groß die Wahrscheinlichkeit für den **Fehler 2. Art** ist. Die Wahrscheinlichkeit für den Fehler 2. Art wird im Gegensatz zum Fehler 1. Art nicht kontrolliert. Beide Fehlerwahrscheinlichkeiten können nicht gleichzeitig minimiert werden, da tendenziell folgende Aussage gilt.

> Die Wahrscheinlichkeit für einen Fehler zweiter Art (Nichtablehnung einer falschen Nullhypothese) wächst mit abnehmender Wahrscheinlichkeit für einen Fehler erster Art.[3]

Die Wahrscheinlichkeit β für einen Fehler zweiter Art hängt vom tatsächlichen Wert des unbekannten Parameters ab und kann deshalb nicht explizit berechnet, sondern lediglich geschätzt werden (s. Abschnitt 3.3).

Die folgende Tabelle gibt einen Überblick über die beiden **Fehlermöglichkeiten** bei der Durchführung eines Tests:

Testentscheidung	tatsächlicher Zustand	
	Nullhypothese richtig	Nullhypothese falsch
Nullhypothese nicht ablehnen	richtige Entscheidung	**Fehler 2. Art** (β-Fehler)
Nullhypothese ablehnen	**Fehler 1. Art** (α-Fehler)	richtige Entscheidung

Aus den bisherigen Ausführungen ergibt sich ein wichtiger Grundsatz für das Aufstellen von Null- und Alternativhypothesen:

> Soll durch einen Test der statistische Nachweis einer **Behauptung** erfolgen, so muss die **Nullhypothese** die **Negation dieser Behauptung** sein, denn durch die Vorgabe von α ist eine Verwerfung der Nullhypothese statistisch abgesichert.

Der **Nachweis** der interessierenden Hypothese H_1 ist genau dann erfolgt, **wenn H_0 abgelehnt wird**. Eine derartige Testentscheidung führt bei einer richtigen Nullhypothese mit einer Wahrscheinlichkeit von höchstens α zu einem Irrtum. Diese Wahrscheinlichkeit α wird aber vorgegeben und kann deshalb den jeweiligen Erfordernissen entsprechend klein gehalten werden.

Beispiel 3.1.5:

a) *In einer Großstadt soll der Anteil der Arbeitslosen seit Jahren bei 12% liegen. Dieser Erfahrungswert soll nun durch einen Test überprüft werden. Da sowohl eine Verringerung als auch eine Erhöhung des Anteils von Interesse ist, wird die zweiseitige Fragestellung gewählt.*

$$H_0: \ \pi = \pi_0 = 0.12 \quad \text{vs.} \quad H_1: \ \pi \neq \pi_0 = 0.12.$$

[3]siehe interaktive Mathematica-Applets auf der Homepage des Lehrstuhls http://www.fernuni-hagen.de/ls_statistik/forschung/multimedia/

b) Es soll nachgewiesen werden, dass Weinflaschen eines Anbaugebietes im Mittel weniger als die angegebene Mindestfüllmenge von $\mu_0 = 1\ l$ enthalten. Hierbei interessiert nur die Abweichung nach unten und es wird folgende einseitige Fragestellung gewählt:

$$H_0: \ \mu \geq \mu_0 = 1\ l \quad \text{vs.} \quad H_1: \ \mu < \mu_0 = 1\ l.$$

Gelingt es, H_0 zu verwerfen, so kann die Wahrscheinlichkeit für eine Fehlentscheidung explizit mit α angegeben werden und der geforderte Nachweis von H_1 ist somit statistisch gesichert. Eine konkrete Aussage über die Fehlerwahrscheinlichkeit bei Nichtablehnung der Nullhypothese ist dagegen nicht möglich.

Vor dem Trugschluss, aus der Nichtablehnung einer Nullhypothese auf ihre Richtigkeit schließen zu wollen, muss nachdrücklich gewarnt werden. Statistisch gesichert ist lediglich die Ablehnung der Nullhypothese. Daher wird **nie** von einer *Annahme der Nullhypothese* sondern stets von der **Nichtablehnung der Nullhypothese** gesprochen.

Für stetige Verteilungen gibt die Wahrscheinlichkeit α für einen **zweiseitigen Test** *genau* die Wahrscheinlichkeit des Fehlers 1. Art an.

Für einen **einseitigen Test** gibt α die *obere Grenze* der Wahrscheinlichkeit für einen Fehler erster Art an (s. Abbildung 3.1.3). Der Grund hierfür ist, dass bei einem einseitigen Test die Nullhypothese nicht durch einen einzigen Wert, sondern durch ein Intervall ($H_0 : \ \theta \leq \theta_0$ bzw. $\theta \geq \theta_0$), dargestellt wird.

In fast allen Fällen genügt es, die Irrtumswahrscheinlichkeit mit $\alpha = 0.01, \alpha = 0.05$ oder $\alpha = 0.1$ festzulegen. Welcher Wert im Einzelfall bevorzugt wird, hängt davon ab, welche Folgen ein Irrtum bei der Testentscheidung hat.

Ist α **zu klein**, so wird es kaum möglich sein, eine Nullhypothese zu verwerfen, denn der „α-Fehler" (Verwerfen einer richtigen Nullhypothese) soll fast völlig ausgeschlossen werden. Das erhöht die Gefahr, dass die Nullhypothese auch dann nicht verworfen wird, wenn sie tatsächlich nicht zutrifft. **Eine Verringerung des Fehlers 1. Art entspricht einer Erhöhung des Fehlers 2. Art.**

Ist α **zu groß**, so ist die Gefahr groß, die Nullhypothese nur aufgrund von Zufallseinflüssen abzulehnen, obwohl H_0 doch zutrifft.

Beispiel 3.1.6:

a) *Beim Bau einer Brücke werden Stahlseile verwendet, die nach Herstellerangaben eine mittlere Reißfestigkeit von $\mu_0 = 80$ t haben. In den einzelnen Lieferungen werden Reißfestigkeitsuntersuchungen vorgenommen.*

   Es ist $H_0: \ \mu \geq \mu_0 = 80$ t.

   Hier wird $\alpha = 0.2$ verhältnismäßig groß gewählt. Es erscheint hier vernünftig, das Risiko des Fehlers 1. Art relativ hoch zu veranschlagen, um den Fehler 2. Art möglichst gering zu halten. Der Schaden, der dadurch entsteht, „gute" Seile zurückzuweisen, ist gering gegenüber dem Einsturz der Brücke durch den Einbau minderwertiger Seile.

b) *Bei der Produktion von Leuchtmitteln soll geprüft werden, ob der Ausschussanteil π den Wert 0.02 überschreitet.*

   Es ist $H_0: \ \pi \leq \pi_0 = 0.02$.

   Es wird $\alpha = 0.01$ relativ klein gewählt, da das unnötige Stoppen der Produktion sehr teuer ist.

3.1.4 Festlegung einer Prüfgröße und Bestimmung der Verteilung dieser Prüfgröße

Die Überprüfung einer Nullhypothese über einen unbekannten Parameter θ einer Grundgesamtheit erfolgt mit Hilfe einer Stichprobenfunktion T, die aus den Stichprobenvariablen $X_1, ..., X_n$ bestimmt wird. Beispiele dafür sind der Stichprobenmittelwert, der Stichprobenanteilswert und die Stichprobenvarianz bzw. die Stichprobenstandardabweichung.

Prüfgröße, Teststatistik

Eine Stichprobenfunktion $T = g(X_1, ..., X_n)$, die zur Überprüfung einer Hypothese über einen Parameter θ verwendet wird, heißt **Prüfgröße** oder **Teststatistik**. Die Realisation dieser Prüfgröße wird mit $t = g(x_1, ..., x_n)$ angegeben.

Welche Stichprobenfunktion verwendet wird, hängt von den Gegebenheiten der jeweiligen Fragestellung ab.

> Jede Stichprobenfunktion T, die als Schätzfunktion für einen Parameter θ geeignet ist, kann auch als Teststatistik zur Überprüfung einer Hypothese über diesen Parameter verwendet werden.

Für einen Test über den Mittelwert μ einer Grundgesamtheit findet meist die Prüfgröße $\overline{X}$ Verwendung. Liegt eine symmetrische Verteilung vor, wäre eine andere Möglichkeit die Verwendung des Medians X_{med}. Soll eine Hypothese über die Standardabweichung σ getestet werden, so kann die Stichprobenstandardabweichung S als Prüfgröße verwendet werden oder alternativ die Spannweite W.

Die Unterschiede der Prüfgrößen T drücken sich meist in ihrer Streuung aus. So ist beispielsweise X_{med} größeren Zufallsschwankungen unterworfen als $\overline{X}$ und die Spannweite W hat ebenso eine größere Streuung im Vergleich zu S. Die Stichprobenfunktion X_{med} ist also i.d.R. für den Test einer Hypothese über μ weniger gut geeignet als $\overline{X}$. Eventuelle Abweichungen von der Nullhypothese werden mit einer geringeren Wahrscheinlichkeit erkannt. Die gleichen Überlegungen gelten für die Spannweite W.

Die **Verteilung der Prüfgröße** lässt sich oft aus der Verteilung des Merkmals X in der Grundgesamtheit und den Annahmen des Stichprobenmodells bestimmen. **Vorausgesetzt** wird hierbei immer, **dass die Nullhypothese tatsächlich zutrifft**. Weiterhin muss der Stichprobenumfang n bekannt sein.

Beispiel 3.1.7:

Einem Reifenhersteller ist bekannt, dass seine Reifen eine durchschnittliche Lebensdauer von $\mu = 40000$ km bei einer Standardabweichung von $\sigma = 2000$ km haben. Nach einer Materialänderung soll festgestellt werden, ob dadurch eine Veränderung der Lebensdauer eingetreten ist.

$$H_0 : \mu = \mu_0 = 40000$$

Als Prüfgröße wird hier der Stichprobenmittelwert $\overline{X}$ verwendet. Da der Stichprobenumfang $n = 100$ größer als 30 ist, kann für die Verteilung von

$\overline{X}$ *approximativ von einer Normalverteilung ausgegangen werden mit*

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2000}{10} = 200.$$

Unter der Voraussetzung, dass H_0 zutrifft, muss die Wahrscheinlichkeitsverteilung von $\overline{X}$ den Erwartungswert $E(\overline{X}) = \mu_0$ haben. Es gilt also:

$\overline{X}$ *ist approximativ $N(\mu_0, \sigma^2_{\overline{X}})$-verteilt bzw. $N(40000, 40000)$-verteilt.*

***Beispiel 3.1.8**:*

Vor einer Bundestagswahl soll die Behauptung des Kanzlers überprüft werden, hinter ihm stünden 50% der Bevölkerung. Die Anzahl X der „Hintermänner“ in einer Zufallsstichprobe des Umfangs 400 ist unter Annahme der Behauptung binomialverteilt mit den Parametern $n = 400$ und $\pi_0 = 0.5$, d.h. $X \sim B(400, 0.5)$. Wegen $n\pi_0 = n(1-\pi_0) = 200 \geq 5$ kann $B(n, \pi_0)$ durch die Normalverteilung $N(n\pi_0, n\pi_0(1-\pi_0)) = N(200, 100)$ approximiert werden.

3.1.5 Bestimmung der kritischen Werte

Ablehnungsbereich

Der Bereich, in dem Werte der Prüfgröße liegen, die unter Gültigkeit von H_0 sehr unwahrscheinlich sind, heißt **Ablehnungsbereich**. Fällt die Prüfgröße in diesen Bereich, so wird H_0 abgelehnt.

kritische Werte

Die Grenzen durch die der Ablehnungsbereich charakterisiert wird, heißen **kritische Werte**. Sie werden mit $\boldsymbol{c_u}$ **(untere Grenze)** und $\boldsymbol{c_o}$ **(obere Grenze)** bezeichnet und gehören selbst nicht zum Ablehnungsbereich.

Eine Nullhypothese soll abgelehnt werden, wenn die Prüfgröße zu einer **„signifikanten“** Abweichung von der Nullhypothese führt. Die Wahrscheinlichkeit, dass die Prüfgröße trotz Gültigkeit von H_0 aufgrund einer „zufälligen zu großen“ Abweichung zu einer Ablehnung führt, soll höchstens α betragen.

Signifikanztest

Ein **Signifikanztest** ist dadurch definiert, dass

$$\boldsymbol{P}(H_1 \text{ annehmen} \mid H_0 \text{ richtig}) \leq \alpha$$

gilt, d.h. die Wahrscheinlichkeit für den Fehler 1. Art beträgt höchstens α.

Wenn mehrere Stichproben für dasselbe Testproblem gezogen würden, so würde ein Signifikanztest zum Niveau α im Mittel in $\alpha\%$ der Fälle zur Ablehnung führen.

Anmerkung:

An dieser Stelle wird ausdrücklich daraufhingewiesen, dass der Ausdruck $\boldsymbol{P}(H_1 \text{ annehmen} \mid H_0 \text{ richtig})$ nicht in dem Sinne falsch interpretiert wird, indem den Hypothesen Wahrscheinlichkeiten zugeordnet werden. Der Ausdruck $\boldsymbol{P}(H_1 \text{ annehmen} \mid H_0 \text{ richtig})$ gibt die Wahrscheinlichkeit an, dass unter Gültigkeit der Nullhypothese die verwendete Teststatistik zur Annahme von H_1 führt.

Kritische Werte bei zweiseitiger Nullhypothese

Für ein zweiseitiges Testproblem gilt $H_0: \ \theta = \theta_0 \quad \text{vs.} \quad H_1: \ \theta \neq \theta_0$.

Die Nullhypothese wird nicht abgelehnt, wenn für die Prüfgröße $T = g(X_1, ..., X_n)$ gilt:

$$c_u \leq T \leq c_o.$$

Der Ablehnungsbereich umfasst alle Werte unterhalb von c_u und oberhalb von c_o. Die Nullhypothese wird also abgelehnt, wenn gilt

$$T < c_u \textbf{ oder } T > c_o.$$

Unter Benutzung der Verteilung, die T bei Gültigkeit von H_0 besitzt, werden c_u und c_o bestimmt. Gefordert wird hierbei, dass die Wahrscheinlichkeit, dass T kleiner als c_u oder größer als c_o ausfällt, jeweils $\frac{\alpha}{2}$ beträgt, d.h.

$$\boldsymbol{P}(T < c_u) = \frac{\alpha}{2} \quad \text{und} \quad \boldsymbol{P}(T > c_o) = \frac{\alpha}{2}.$$

Daraus folgt:

$$\boldsymbol{P}(T \in \text{ Ablehnungsbereich}) = \boldsymbol{P}(T < c_u \text{ oder } T > c_o) = \alpha,$$

$$\boldsymbol{P}(T \notin \text{ Ablehnungsbereich}) = \boldsymbol{P}(c_u \leq T \leq c_o) = 1 - \alpha.$$

Beispiel 3.1.9:

In Beispiel 3.1.7 (mittlere Lebensdauer von Reifen) lautete die Nullhypothese

$$H_0: \ \mu = \mu_0 = 40000.$$

Für $n = 100$ *und* $\sigma_{\overline{X}} = 200$ *ist die Prüfgröße* $\overline{X}$ *bei Gültigkeit von* H_0 *näherungsweise* $N(40000, 40000)$*-verteilt. Unter der Annahme, dass* H_0 *zutrifft, und nach Standardisierung* ($Z = \frac{\overline{X}-\mu_0}{\sigma_{\overline{X}}} \overset{a}{\sim} N(0,1)$) *gilt:*

$$\boldsymbol{P}\left(c_u \leq \overline{X} \leq c_o\right) = \boldsymbol{P}\left(\frac{c_u - \mu_0}{\sigma_{\overline{X}}} \leq Z \leq \frac{c_o - \mu_0}{\sigma_{\overline{X}}}\right) \approx 1 - \alpha$$

Werden die sich nach der Standardnormalverteilung ergebenden Quantile

$$-z_{1-\frac{\alpha}{2}} = \frac{c_u - \mu_0}{\sigma_{\overline{X}}} \qquad \text{und} \qquad z_{1-\frac{\alpha}{2}} = \frac{c_o - \mu_0}{\sigma_{\overline{X}}}$$

jeweils nach c_u *und* c_o *umgeformt, so lauten die* ***kritischen Werte***

$$\boldsymbol{c_u = \mu_0 - z_{1-\frac{\alpha}{2}} \sigma_{\overline{X}}} \qquad \textit{und} \qquad \boldsymbol{c_o = \mu_0 + z_{1-\frac{\alpha}{2}} \sigma_{\overline{X}}}.$$

In diesem Beispiel berechnen sich die kritischen Werte mit $\alpha = 0.05$ *konkret zu*

$$c_u = 40000 - 1.96 \cdot 200 = 39608 \quad \textit{und} \quad c_o = 40000 + 1.96 \cdot 200 = 40392.$$

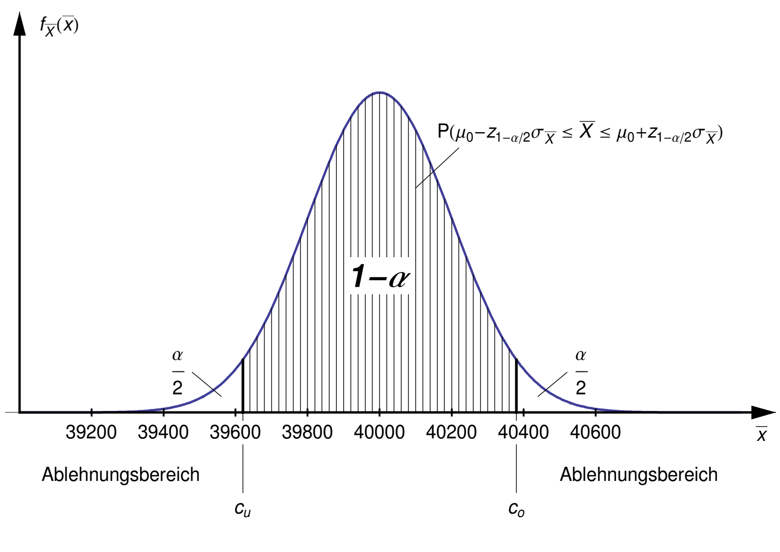

Abbildung 3.1.1: *Ablehnungsbereich bei zweiseitiger Nullhypothese*

Kritische Werte bei einseitiger Nullhypothese

Für ein einseitiges Testproblem gilt

$$\text{Fall 1:} \quad H_0: \ \theta \leq \theta_0 \quad \text{vs.} \quad H_1: \ \theta > \theta_0,$$

$$\text{Fall 2:} \quad H_0: \ \theta \geq \theta_0 \quad \text{vs.} \quad H_1: \ \theta < \theta_0.$$

In Fall 1 wird nur ein kritischer Wert berechnet, der oberhalb des Wertes θ_0 liegt. Die Nullhypothese wird nicht abgelehnt, wenn die Prüfgröße unterhalb des oberen kritischen Wertes c_o liegt: $T \leq c_o$. Gilt $T > c_o$, so wird die Nullhypothese abgelehnt. Dieser Bereich beschreibt den Ablehnungsbereich.

In Fall 2 wird ebenfalls nur ein kritischer Wert berechnet. Die Nullhypothese wird abgelehnt, wenn der Stichprobenwert zu klein ist. Für den Ablehnungsbereich gilt daher $T < c_u$. Die Nullhypothese wird nicht abgelehnt, wenn $T \geq c_u$ gilt.

Einseitige Nullhypothese sind jeweils für einen ganzen Bereich des Parameters θ erfüllt. Um dennoch einen eindeutigen kritischen Wert c_u bzw. c_o bestimmen zu können, wird zur Festlegung der **Verteilung der Prüfgröße $\boldsymbol{T}$** der Parameterwert verwendet, der die Nullhypothese „gerade noch“ erfüllt, nämlich θ_0.

Die Bestimmung von c_o bei gegebener Hypothese $H_0: \ \theta \leq \theta_0$ erfolgt so, dass die Wahrscheinlichkeit, dass T größer als c_o ausfällt, gerade gleich α ist. Abweichungen von der Nullhypothese nach unten sind hierbei nicht von Interesse, da in dem Fall der Behauptung $\theta \leq \theta_0$ nicht widersprochen wird. Zur Bestimmung von c_u bei gegebener Hypothese $H_0: \ \theta \geq \theta_0$ wird analog die Wahrscheinlichkeit, dass T kleiner als c_u ausfällt, mit α angegeben. Die Bestimmungsgleichungen lauten also:

$$\boldsymbol{P}(T > c_o) = \alpha \qquad \text{bzw.} \qquad \boldsymbol{P}(T < c_u) = \alpha$$

$$\text{und} \quad \boldsymbol{P}(T \notin \text{Ablehnungsbereich}) = \boldsymbol{P}(T \leq c_o) = 1 - \alpha$$

$$\text{bzw.} \quad \boldsymbol{P}(T \notin \text{Ablehnungsbereich}) = \boldsymbol{P}(T \geq c_u) = 1 - \alpha.$$

Beispiel 3.1.10:

Wird in dem Beispiel mit den Autoreifen von den Abnehmern vermutet, dass die Materialänderung zu einer ***Verminderung*** *der ursprünglichen Lebensdauer von 40000 geführt hat, so lautet die Nullhypothese H_0 : „Die Lebensdauer hat sich* ***erhöht***“*, denn statistisch gesichert kann nur die Alternativhypothese nachgewiesen werden.*

$$H_0: \ \mu \geq \mu_0 = 40000$$

Unter der Annahme, dass H_0 zutrifft, und nach Standardisierung gilt:

$$\boldsymbol{P}\left(\overline{X} < c_u\right) = \boldsymbol{P}\left(Z < \frac{c_u - \mu_0}{\sigma_{\overline{X}}}\right) \approx \alpha.$$

Wird hier das sich nach der Standardnormalverteilung ergebende Quantil $z_\alpha = \frac{c_u - \mu_0}{\sigma_{\overline{X}}}$ nach c_u umgeformt, so ergibt sich der ***kritische Wert***

$$\boldsymbol{c_u = \mu_0 + z_\alpha \sigma_{\overline{X}} = \mu_0 - z_{1-\alpha}\sigma_{\overline{X}}}.$$

Aus Symmetriegründen gilt $z_\alpha = -z_{1-\alpha}$. Somit berechnet sich der untere kritische Wert mit $\alpha = 0.05$ und $n = 100$ konkret zu

$$c_u = 40000 - 1.65 \cdot 200 = 39670.$$

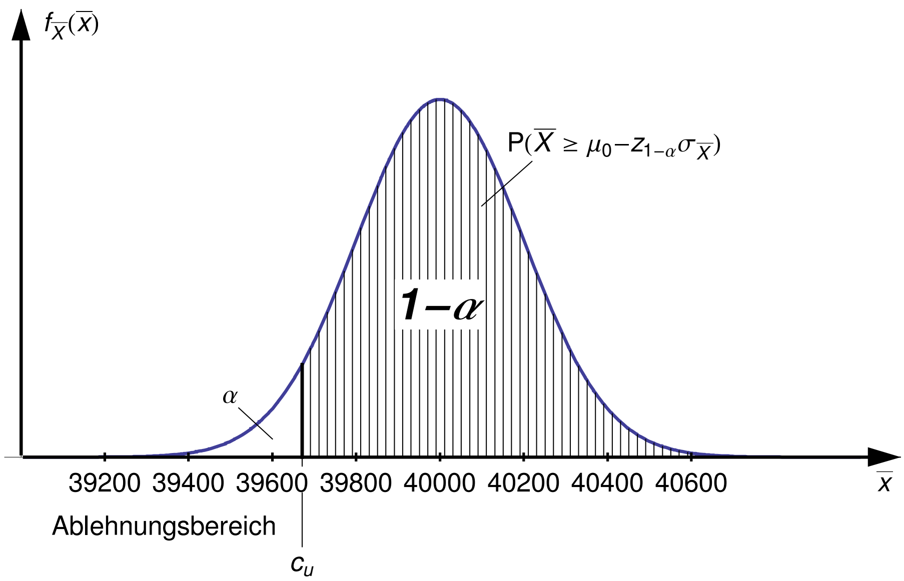

Abbildung 3.1.2: *Ablehnungsbereich bei einseitiger Nullhypothese*

Für den einseitigen Test $H_0: \ \theta \geq \theta_0$ wurde der kritische Wert unter der Bedingung bestimmt, dass die Nullhypothese „gerade noch" erfüllt ist ($\theta = \theta_0$). In diesem Fall ist die Wahrscheinlichkeit für den Fehler 1. Art definitionsgemäß gleich α. Hat nun in Wirklichkeit der Parameter θ den Wert $\theta^* > \theta_0$, so ist H_0 ebenfalls erfüllt. Die Wahrscheinlichkeit, dass T kleiner als das für den Fall $\theta = \theta_0$ berechnete c_u ausfällt, ist aber offenbar kleiner als α. Dies zeigt Abbildung 3.1.3.

Hiermit wird bestätigt, dass für einen einseitigen Test das **Signifikanzniveau $\boldsymbol{\alpha}$** eine **obere Grenze** für die Wahrscheinlichkeit des Fehlers 1. Art angibt.

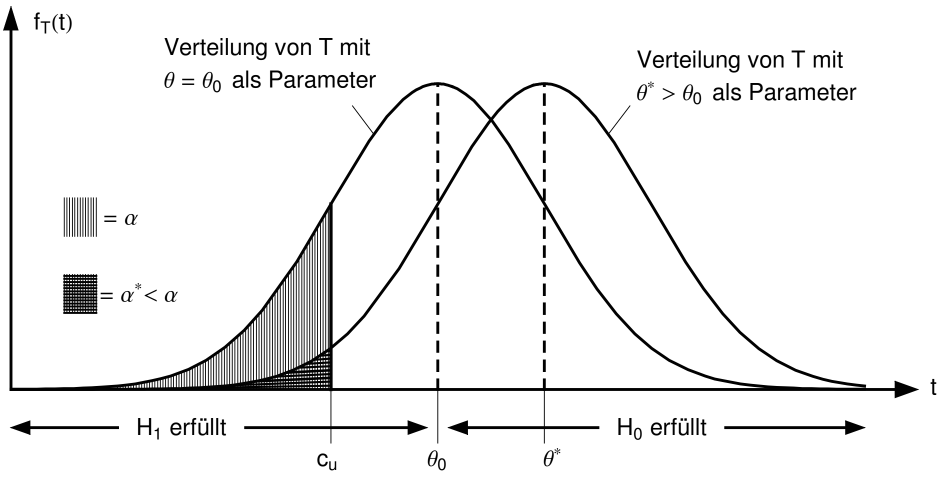

Abbildung 3.1.3: *Tatsächliche Irrtumswahrscheinlichkeit bei dem einseitigen Test $H_0: \ \theta \geq \theta_0$*

Kritische Werte bei standardisierter Prüfgröße:

Die bisher berechneten kritischen Werte hängen eng mit den Grenzen der Konfidenzintervalle zusammen (s. Abschnitt 3.4). Oft wird anstelle der einfachen Stichprobenfunktion T eine standardisierte Prüfgröße, z.B.

$$\frac{T - E(T)}{\sqrt{\mathrm{Var}(T)}},$$

betrachtet. Eine Testentscheidung erfolgt dann im direkten Vergleich dieser Prüfgröße mit den entsprechenden Quantilen der vorliegenden Verteilung.

Beispiel 3.1.11:

Wird in dem Beispiel „Lebensdauer der Autoreifen“ die ***standardisierte Prüfgröße***

$$Z = \frac{\overline{X} - \mu_0}{\sigma_{\overline{X}}}$$

betrachtet, welche für $\mu = \mu_0$ approximativ standardnormalverteilt ist, so gilt für das ***zweiseitige Testproblem***

$$\boldsymbol{P}\left(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}\right) \approx 1 - \alpha.$$

Eine Testentscheidung ist hier direkt möglich. In dem Fall wird die Nullhypothese $H_0 : \ \mu = \mu_0$ abgelehnt, falls für die standardisierte Testgröße $Z < -z_{1-\frac{\alpha}{2}}$ oder $Z > z_{1-\frac{\alpha}{2}}$ gilt.

Für das ***einseitige Testproblem*** *gilt entsprechend aufgrund der Bedingung*

$$\boldsymbol{P}\left(Z \leq z_{1-\alpha}\right) \approx 1 - \alpha \quad \text{bzw.} \quad \boldsymbol{P}\left(Z \geq -z_{1-\alpha}\right) \approx 1 - \alpha,$$

dass die Nullhypothese $H_0 : \ \mu \leq \mu_0$ bzw. $H_0 : \ \mu \geq \mu_0$ abgelehnt wird, falls für die standardisierte Testgröße $Z > z_{1-\alpha}$ bzw. $Z < -z_{1-\alpha}$ gilt.

3.1.6 Ziehung der Stichprobe und Berechnung des Wertes der Prüfgröße

Aus den Stichprobenwerten wird die Realisation $t = g(x_1, ..., x_n)$ der Prüfgröße T berechnet.

Beispiel 3.1.12:

Für die Ermittlung der Lebensdauer von $n = 100$ Reifen in dem Beispiel 3.1.7 kann als Prüfgröße $Z = \frac{\overline{X}-\mu_0}{\sigma_{\overline{X}}}$ bzw. $T = \overline{X}$ verwendet werden. Die Realisation der einfachen Prüfgröße $\overline{X}$ beträgt $\overline{x} = 40350$ km und als Realisation der standardisierten Prüfgröße ergibt sich

$$z = \frac{\overline{x} - \mu_0}{\sigma_{\overline{X}}} = \frac{40350 - 40000}{200} = 1.75.$$

3.1.7 Testentscheidung und Interpretation der Entscheidung

Fällt die Prüfgröße in den Ablehnungsbereich, so wird die Nullhypothese abgelehnt. Ansonsten wird H_0 nicht abgelehnt. Bei der Interpretation des Testergebnisses ist darauf zu achten, dass diese eine **Diskussion der Fehlerrisiken** enthält.

Beispiel 3.1.13:

a) *Mittels des Reifentests (Beispiel 3.1.7) soll durch ein unabhängiges Institut geprüft werden, ob die angegebene Lebensdauer stimmt oder nicht. Für die Nullhypothese*

$$H_0: \ \mu = \mu_0 = 40000$$

fällt $z = \dfrac{\overline{x} - \mu_0}{\sigma_{\overline{X}}} = 1.75$ $(\overline{x} = 40350)$ *nicht in den Ablehnungsbereich, denn es ist*

$$\boldsymbol{c_u} = -z_{0.975} = -1.96 \leq \frac{\overline{x} - \mu_0}{\sigma_{\overline{X}}} \leq 1.96 = z_{0.975} = \boldsymbol{c_o}$$

$$(\boldsymbol{c_u} = \mu_0 - z_{0.975}\sigma_{\overline{X}} = 39608 \leq \overline{x} \leq 40392 = \mu_0 + z_{0.975}\sigma_{\overline{X}} = \boldsymbol{c_o}).$$

Somit wird $\boldsymbol{H_0}$ ***nicht abgelehnt****. Die Nichtablehnung der Nullhypothese* $H_0: \ \mu = \mu_0$ *bedeutet dann, dass kein Grund vorliegt, die alten Lebensdauerdaten zu korrigieren. Das Testergebnis reicht also nicht, um* H_0 *zu verwerfen. Damit wird aber nicht behauptet, dass* H_0 *zutrifft, denn die Wahrscheinlichkeit, eine falsche Nullhypothese nicht zu verwerfen, kann eventuell sehr groß sein.*

b) *Da es in der Untersuchung der Lebensdauer der Autoreifen nach Materialänderung aus Sicht des Herstellers insbesondere auf eine Erhöhung der Lebensdauer der Autoreifen ankommt, kann hier auch folgende einseitige Nullhypothese formuliert werden. Die Realisierung der Prüfgröße* $z = \frac{\overline{x} - \mu_0}{\sigma_{\overline{X}}} = 1.75$ $(\overline{x} = 40350)$ *fällt für die Nullhypothese*

$$H_0: \ \mu \leq \mu_0 = 40000$$

in den Ablehnungsbereich, denn es ist

$$z = \frac{\overline{x} - \mu_0}{\sigma_{\overline{X}}} = 1.75 > 1.65 = z_{0.95} = \boldsymbol{c_o}$$

$$(\overline{x} = 40350 > 40330 = \mu_0 + z_{0.95}\sigma_{\overline{X}} = \boldsymbol{c_o}).$$

Somit wird hier $\boldsymbol{H_0}$ ***abgelehnt****. Die Wahrscheinlichkeit einer Fehlentscheidung, nämlich* H_0 *abzulehnen, obwohl* H_0 *richtig ist, ist hier mit* $\alpha = 0.05$ *gering. Daher wird die Alternativhypothese* H_1 *als „statistisch gesichert“ angenommen und es ist statistisch nachgewiesen, dass sich die Lebensdauer erhöht hat. Die Notwendigkeit einer (eventuell mit hohen Kosten verbundenen) Produktionsumstellung ist dann „genügend“ abgesichert.*

3.1.8 Zusammenfassung

Die folgende Übersicht zeigt noch einmal die angegebenen Schritte bei der Durchführung eines Parametertests:

(1) Festlegung von Grundgesamtheit und Typ der Verteilung.

(2) Formulierung einer Hypothese und Gegenhypothese.

(3) Vorgabe einer Irrtumswahrscheinlichkeit.

(4) Festlegung einer Prüfgröße und Bestimmung der Verteilung dieser Prüfgröße unter Gültigkeit der Nullhypothese.

(5) Bestimmung kritischer Werte und Angabe eines Ablehnungsbereiches.

(6) Ziehung einer Stichprobe und Berechnung des Wertes der zugrundeliegenden Prüfgröße.

(7) Testentscheidung und Interpretation der Entscheidung.

Zum Schluss dieses Abschnitts werden anhand eines Beispiels die einzelnen Schritte des Testschemas angegeben.

Beispiel 3.1.14*:*

Es soll überprüft werden, ob eine gegebene Münze „ideal“ ist, d.h. das Ergebnis „Zahl“ tritt mit der Wahrscheinlichkeit 0.5 *auf. Die Münze wird* $n = 8$ *mal geworfen. Gegeben sind somit Binärvariablen* X_i*,* $i = 1, ..., n$*, mit den Realisationen „*$1 \hat{=}$ *Zahl liegt oben“ und „*$0 \hat{=}$ *Kopf liegt oben“.*

(1) *a) Das Merkmal ist qualitativ.*

*b) Die Grundgesamtheit besteht aus allen jemals durchgeführten oder noch durchzuführenden (**unendlich** vielen) Würfen mit dieser Münze.*

c) Die Binärvariablen X_i, $i = 1, ..., n$, sind unter Gültigkeit der Nullhypothese bernoulliverteilt mit $\pi = 0.5$.

(2) Hier wird ein zweiseitiges Testproblem gewählt.

$$H_0: \ \pi = \pi_0 = 0.5 \qquad \text{vs.} \qquad H_1: \ \pi \neq \pi_0 = 0.5$$

(3) Als Signifikanzniveau wird $\alpha = 0.1$ vorgegeben.

(4) Als Prüfgröße wird die Summe $X = \sum_{i=1}^{n} X_i$, die Anzahl der Ergebnisse „Zahl", verwendet. Unter H_0 ist X binomialverteilt mit den Parametern $n = 8$ und $\pi_0 = 0.5$. Betrachtet wird die Wahrscheinlichkeitsfunktion $f(x) = P(X = x)$.

x	**0**	**1**	2	3	4	5	6	**7**	**8**
$f(x)$	**0.004**	**0.031**	0.109	0.219	0.273	0.219	0.109	**0.031**	**0.004**

(5) Zur Bestimmung des Ablehnungsbereiches wird die obige Wahrscheinlichkeitsfunktion für $n = 8$ und $p = 0.5$ betrachtet. Der Ablehnungsbereich wird so bestimmt, dass für einen Signifikanztest gilt:

$$\boldsymbol{P}(H_1 \text{ annehmen} \mid H_0 \text{ richtig}) \leq \alpha.$$

Der markierte Bereich erfüllt diese Bedingung, denn es gilt:

$$\boldsymbol{P}(X < 2 \text{ oder } X > 6) = 0.07 \text{ und } \boldsymbol{P}(2 \leq X \leq 6) = 0.93$$

.

Bei richtiger Nullhypothese ist die Wahrscheinlichkeit, die Augenzahl 0, 1, 7 oder 8 zu erhalten, so klein (nämlich 0.07), dass in einem solchen Fall mit ausreichender Sicherheit davon ausgegangen werden kann, dass die Nullhypothese nicht zutrifft. Als kritische Grenzen ergeben sich somit $c_u = 2$ und $c_o = 6$.

(6) In einer Stichprobe des Umfangs $n = 8$ wird $X = 3$ ermittelt.

(7) Da die Anzahl der Ergebnisse „Zahl" $X = 3$ zwischen den kritischen Werten liegt, also $c_u = 2 \leq X \leq c_o = 6$ wird die Nullhypothese nicht abgelehnt, was jedoch nicht bedeutet, dass tatsächlich eine ideale Münze vorliegt. Es kann lediglich nicht nachgewiesen werden, dass es sich um eine gefälschte Münze handelt.

3.2 p-Wert

Bisher wurde die beobachtete Prüfgröße mit den kritischen Werten verglichen, um eine Testentscheidung zu treffen. Alternativ kann durch die Angabe der sogenannten **Überschreitungswahrscheinlichkeit**, kurz auch ***p*-Wert** genannt, ein Test entschieden werden. Diese Möglichkeit wird vorallem von statistischen Software-Paketen genutzt, da so auf die konkrete Angabe eines bestimmten Signifikanzniveaus verzichtet werden kann.

Überschreitungswahrscheinlichkeit

***p*-Wert**

Der ***p*-Wert** gibt die Wahrscheinlichkeit an, dass unter der Hypothese H_0 der beobachtete Wert $t = g(x_1, ..., x_n)$ oder ein extremerer Wert der Prüfgröße $T = g(X_1, ..., X_n)$ ermittelt wird, formal:

$$\begin{array}{lll} H_0: & \theta \leq \theta_0 & \boldsymbol{P}(T \geq t) = p, \\ H_0: & \theta \geq \theta_0 & \boldsymbol{P}(T \leq t) = p, \\ H_0: & \theta = \theta_0 & 2 \cdot \min(\boldsymbol{P}(T \geq t), \boldsymbol{P}(T \leq t)) = p. \end{array}$$

Überschreitet das vorgegebene Signifikanzniveau p, d.h. gilt $p < \alpha$, so wird H_0 abgelehnt.

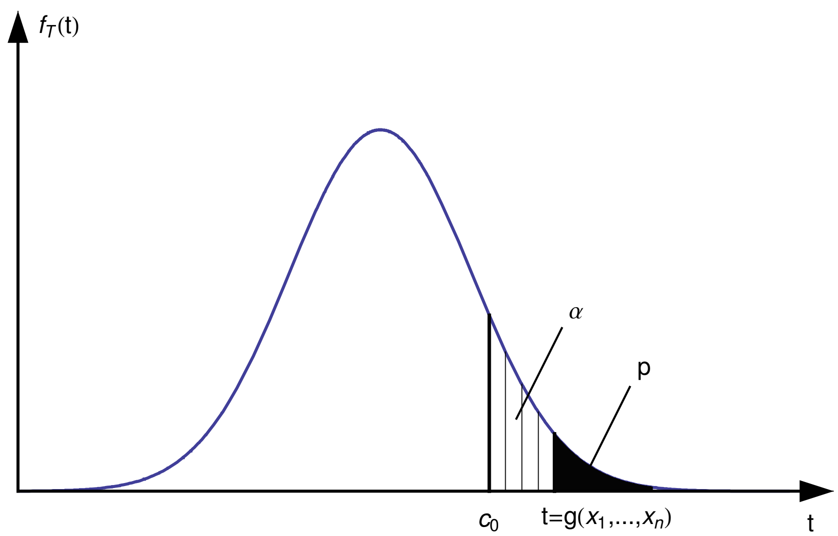

Abbildung 3.2.1: *p-Wert und Signifikanzniveau α bei dem einseitigen Test $H_0 : \theta \leq \theta_0$*

Ist der p-Wert klein, so ist unter H_0 die Wahrscheinlichkeit gering eine extremere oder gleiche Realisation der Prüfgröße zu erhalten. Dies deutet daraufhin, dass H_0 nicht zutrifft. Im Fall $p < \alpha$ gilt, dass t zum Niveau α im Ablehnungsbereich liegen würde.[4]

Alternativ kann der p-Wert interpretiert werden, dass für die vorliegende Realisation der Prüfgröße durch p ein Niveau bestimmt wird, bei dem H_0 gerade eben nicht abgelehnt würde. Der p-Wert gibt für den beobachteten Prüfgrößenwert somit das Nivean an, durch das in diesem Fall genau an der Stelle t der Ablehnungsbereich bestimmt werden würde.

Missbrauch des *p*-Wertes

Diese Interpretation kann zu dem **Missbrauch** führen, dass erst der Test durchgeführt wird und anschließend das Signifikanzniveau so festgelegt wird, dass H_0 noch abgelehnt werden kann. Dies steht jedoch im Widerspruch zu den Prinzipien der Testtheorie.

Beispiel 3.2.1:

Eine Münze wird 10 *mal geworfen. Die Zufallsvariable* X *beschreibt die Anzahl der Ergebnisse „Zahl“. Überprüft werden soll zum Signifikanzniveau* $\alpha = 0.1$ *die Hypothese, es handelt sich um eine faire Münze* $(H_0: \ \pi = 0.5)$. *Der Münzwurf ergibt* $x = 3$. *Unter der Nullhypothese ist* X *binomialverteilt mit* $n = 10$ *und* $\pi = 0.5$. *Als p-Wert ergibt sich*

$$p = 2 \cdot \boldsymbol{P}(X \leq 3) = 2 \cdot 0.1719 = 0.3438.$$

Da $p > \alpha$ *gilt, wird die Nullhypothese nicht abgelehnt.*

Beispiel 3.2.2:

Es soll nachgewiesen werden, dass die Zugfestigkeit von Baustahl bei einer bekannten Standardabweichung von $\sigma = 100\ kg/cm^2$ *mindestens* $5000\ kg/cm^2$ *beträgt* $(\alpha = 0.05)$, *wobei von der Normalverteilung ausgegangen wird. Untersucht werden* 25 *Probestücke mit einer durchschnittlichen Zugfestigkeit von* $5170\ kg/cm^2$. *Betrachtet wird die Nullhypothese* $H_0: \ \mu \leq 5000$. *Als p-Wert ergibt sich*

$$\begin{aligned} p &= \boldsymbol{P}(X \geq 5170) = \boldsymbol{P}(Z \geq \frac{5170 - 5000}{100}) = \boldsymbol{P}(Z \geq 1.7) \\ &= 1 - \boldsymbol{P}(Z \leq 1.7) = 1 - 0.9554 = 0.0446. \end{aligned}$$

Da $p < \alpha$ *gilt, wird die Nullhypothese abgelehnt.*

[4] siehe interaktive Mathematica-Applets auf der Homepage des Lehrstuhls http://www.fernuni-hagen.de/ls_statistik/forschung/multimedia/

3.3 Operationscharakteristik und Gütefunktion von Parametertests

Die Wahrscheinlichkeiten für den Fehler 1. Art und 2. Art hängen eng mit der **Operationscharakteristik** bzw. der **Gütefunktion** zusammen. Betrachtet wird lediglich der Fall eines zweiseitigen Testproblems.

Formal lassen sich die Wahrscheinlichkeiten für den Fehler 1. und 2. Art wie folgt schreiben:

$$\begin{aligned} \boldsymbol{P}(\text{Fehler 1. Art}) &= \boldsymbol{P}(H_0 \text{ ablehnen}|H_0 \text{ richtig}) = \alpha \\ \boldsymbol{P}(\text{Fehler 2. Art}) &= \boldsymbol{P}(H_0 \text{ nicht ablehnen}|H_1 \text{ richtig}) = \beta \end{aligned}$$

Zu jedem **zweiseitigen Test** lässt sich eine Funktion $\beta(\theta)$ bestimmen, welche die Wahrscheinlichkeit angibt, dass T nicht in den Ablehnungsbereich fällt, unter der Bedingung, dass der betreffende Parameter den Wert θ hat.

Operationscharakteristik

Die bedingte Wahrscheinlichkeitsfunktion $\beta(\theta)$

$$\beta(\theta) = \boldsymbol{P}(c_u \leq T \leq c_o|\theta)$$

heißt **Operationscharakteristik (OC)** des zweiseitigen Tests.

Über dem Bereich der Alternativhypothese ist die OC also eine Fehlerwahrscheinlichkeit, nämlich die für den Fehler 2. Art, denn die Nichtablehnung der Nullhypothese ist hier die falsche Entscheidung. Über dem Bereich der Nullhypothese ist die OC keine Fehlerwahrscheinlichkeit, da hier die Nichtablehnung der Nullhypothese die korrekte Entscheidung ist.

OC-Funktion

Die Operationscharakteristik eines Test kann als Funktion tabellarisch oder grafisch als sogenannte **OC-Funktion** dargestellt werden.

Beispiel 3.3.1:

Der normalverteilte Sollwert von Widerständen in einer Lieferung von 500 Stück beträgt nach Herstellerangaben $\mu_0 = 100\ \Omega$ bei einer Standardabweichung von $\sigma = 3\ \Omega$. Zur Prüfung des Sollwertes wird eine Zufallsstichprobe vom Umfang $n = 20$ gezogen ($\alpha = 0.05$). Es soll die OC*-Kurve für $96 \leq \mu \leq 104$ tabellarisch und grafisch dargestellt werden.*

Es ist $n = 20,\ \sigma_{\overline{X}} = 0.67,\ c_u = 98.7,\ c_o = 101.3$. Da für $\beta(\mu)$ gilt

$$\begin{aligned}\beta(\mu) &= \boldsymbol{P}(c_u \leq \overline{X} \leq c_o|\mu) = \boldsymbol{P}\left(\frac{c_u - \mu}{\sigma_{\overline{X}}} \leq Z \leq \frac{c_o - \mu}{\sigma_{\overline{X}}}\right) \\ &= F_Z\left(\frac{c_o - \mu}{\sigma_{\overline{X}}}\right) - F_Z\left(\frac{c_u - \mu}{\sigma_{\overline{X}}}\right)\end{aligned}$$

kann für die Berechnung von $\beta(\mu)$ bei gegebenem μ das folgende Rechenschema verwendet werden:[5]

μ	$\frac{c_o-\mu}{\sigma_{\overline{X}}}$	$\frac{c_u-\mu}{\sigma_{\overline{X}}}$	$F_Z\left(\frac{c_o-\mu}{\sigma_{\overline{X}}}\right)$	$F_Z\left(\frac{c_u-\mu}{\sigma_{\overline{X}}}\right)$	$\beta(\mu)$
96	7.91	4.03	1	1	0
97	6.42	2.54	1	0.9945	0.0055
98	4.93	1.04	1	0.8508	0.1492
99	3.43	-0.45	0.9997	0.3264	0.6733
100	1.94*	-1.94*	0.9738	0.0262	0.9476*
101	0.45	-3.43	0.6736	0.0003	0.6733
102	-1.04	-4.93	0.1492	0	0.1492
103	-2.54	-6.42	0.0055	0	0.0055
104	-4.03	-7.91	0	0	0

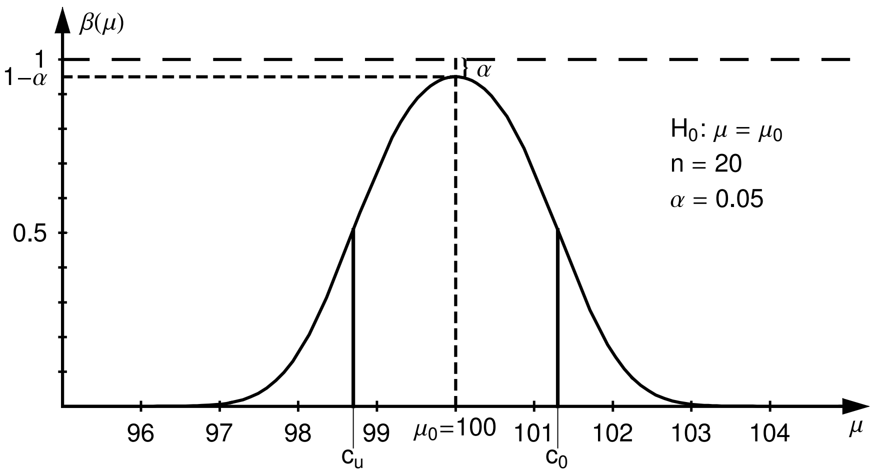

Abbildung 3.3.1: *OC-Kurve zu Beispiel 3.3.1*

[5] Die mit * versehenen Werte weichen wegen der Aufrundung von $\sigma_{\overline{X}}$ auf zwei Stellen leicht von den theoretisch richtigen Werten 1.96 bzw. 0.95 ab.

Aus Abbildung 3.3.1 kann die typische Gestalt einer **OC-Kurve eines zweiseitigen Tests** entnommen werden. An der Stelle $\theta = \theta_0$ (in der Grafik: $\mu = \mu_0$) hat die OC-Kurve ihr Maximum mit

$$\beta(\theta_0) = 1 - \alpha.$$

Das bedeutet, dass die Wahrscheinlichkeit für die Ablehnung der Nullhypothese am kleinsten ist, wenn H_0 erfüllt ist.

Rechts und links vom Maximum nimmt $\beta(\theta)$ ab und geht schließlich gegen Null. Das bedeutet: **Die Wahrscheinlichkeit, H_0 nicht abzulehnen, ist umso kleiner, je weiter der tatsächliche Wert θ von θ_0 entfernt ist**.

Die Operationscharakteristik ist somit vom tatsächlich vorliegenden Parameterwert abhängig. Weiter ist die Operationscharakteristik auch vom Stichprobenumfang n abhängig. Wird der Stichprobenumfang vergrößert, dann wird $\sigma_{\overline{X}}$ kleiner und somit der Ablehnungsbereich größer. Zu einem gegebenen Wert für θ wird insgesamt die Wahrscheinlichkeit β für die Nichtablehnung der Nullhypothese kleiner.

Abbildung 3.3.2 zeigt drei OC-Kurven mit den Daten des vorherigen Beispiels für $\alpha = 0.05$ und die Stichprobenumfänge $n = 10, 20, 50$.

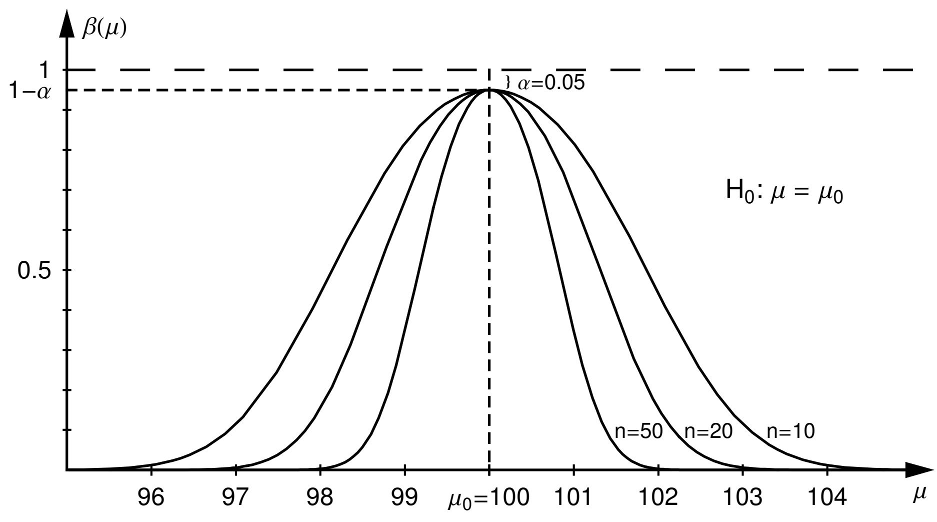

Abbildung 3.3.2: *OC-Kurve für unterschiedliche n*

Alle drei Kurven berühren sich an der Stelle $\mu = \mu_0$ und haben dort den Wert $1 - \alpha = 0.95$.

> Der Wert der Operationscharakteristik ist abhängig vom Stichprobenumfang und vom Wert des Parameters, den es zu überprüfen gilt.

Anhand der Abbildung 3.3.2 ist zu erkennen, dass an jeder beliebigen Stelle μ, für die gilt $\mu \neq \mu_0$ bzw. für die H_0 nicht erfüllt ist, die **Wahrscheinlichkeit β für den Fehler 2. Art mit wachsendem n abnimmt**. Je größer der Stichprobenumfang ist, desto steiler verläuft die Operationscharakteristik, desto schneller fällt also mit zunehmendem Abstand $|\mu - \mu_0|$ das Risiko eines β-Fehlers, und desto „leichter“ lassen sich Abweichungen von der Nullhypothese erkennen.

Gütefunktion (Macht eines Testes)

Die Funktion $\mathbf{1 - \beta(\theta)}$ wird als **Gütefunktion** oder **Macht (engl.: Power)** eines Parametertests bezeichnet.[6] Die **Güte** ist die Wahrscheinlichkeit, mit der H_0 in Abhängigkeit vom tatsächlichen Parameterwert θ abgelehnt wird.

Über dem Bereich der Nullhypothese ist die Güte eine Fehlerwahrscheinlichkeit, nämlich die für den Fehler 1. Art, denn die Ablehnung der Nullhypothese ist dann die falsche Entscheidung. Über dem Bereich der Alternativhypothese ist die Güte keine Fehlerwahrscheinlichkeit, da hier die Ablehnung der Nullhypothese die korrekte Entscheidung ist.

Gütefunktion und OC-Kurve haben einen **komplementären Verlauf**.

Ob mit der Güte- oder mit der OC-Funktion gearbeitet wird, ist Geschmackssache. Jede Funktion charakterisiert einen Test vollständig.

[6] siehe interaktive Mathematica-Applets auf der Homepage des Lehrstuhls http://www.fernuni-hagen.de/ls_statistik/forschung/multimedia/

3.4 Zusammenhang zwischen Konfidenzintervall und Testverfahren

Für eine normalverteilte Zufallsvariable X mit unbekanntem Parameter μ und **bekannter** Varianz σ^2, d.h. $X \sim N(\mu, \sigma^2)$ lautet das zweiseitige $(1-\alpha)$-Konfidenzintervall für μ

$$\text{KI}: \qquad \left[\overline{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \,;\, \overline{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right].$$

Hierbei sind die Grenzen Zufallsvariablen, d.h. die Grenzen sind zufällige Größen. Der Parameter μ ist dagegen eine unbekannte, aber feste Größe.

Wird ein Parametertest für μ durchgeführt, so geben die kritischen Grenzen c_u und c_o einen Bereich für den Stichprobenmittelwert $\overline{X}$ an, bei dem die Nullhypothese $H_0: \ \mu = \mu_0$ nicht abgelehnt werden kann.

$$\text{Test}: \qquad \left[\mu_0 - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \,;\, \mu_0 + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$$

Die Grenzen des Bereiches $[c_u; c_o]$ sind feste Größen, während die Prüfgröße $\overline{X}$ zufällig ist.

Die Grenzen des einen Bereiches können aus den Grenzen des anderen Bereiches hergeleitet werden ($\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$).

$$\begin{aligned}
& \boldsymbol{P}\left(\mu - z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}} \leq \overline{X} \leq \mu + z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}}\right) \\
= \ & \boldsymbol{P}\left(-z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}} \leq \overline{X} - \mu \leq z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}}\right) \\
= \ & \boldsymbol{P}\left(-\overline{X} - z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}} \leq -\mu \leq -\overline{X} + z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}}\right) \\
= \ & \boldsymbol{P}\left(\overline{X} - z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}} \leq \mu \leq \overline{X} + z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}}\right) = 1 - \alpha
\end{aligned}$$

Die Nullhypothese $H_0: \ \mu = \mu_0$ kann auch mittels des Konfidenzintervalls überprüft werden. Überdeckt das Konfidenzintervall nicht den Wert μ_0, so kann die Nullhypothese abgelehnt werden. Die Testentscheidung mittels der Prüfgröße $\overline{X}$ und die Testentscheidung mittels des Konfidenzintervalls liefern dasselbe Ergebnis. Das Konfidenzniveau „$1 - \alpha$" entspricht „1−Signifikanzniveau" beim Testen.

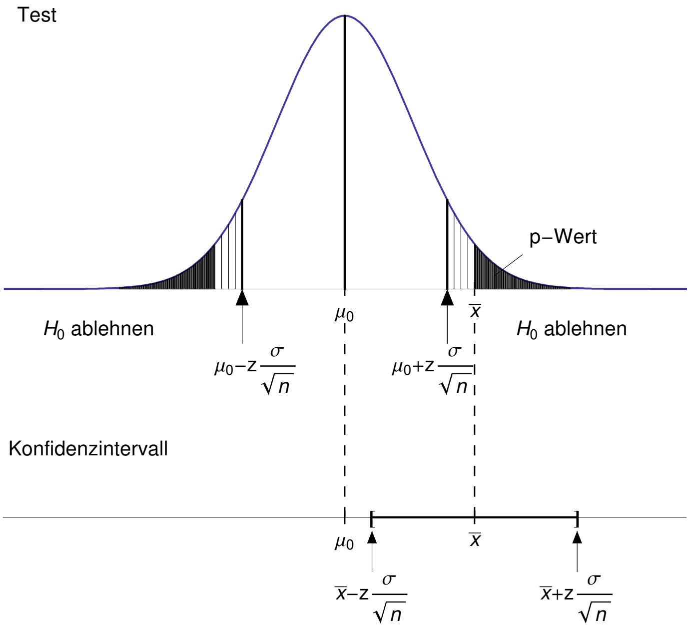

Abbildung 3.4.1: *Zusammenhang zwischen Konfidenzintervall und Testverfahren*

Zusammenfassend kann formuliert werden:

Das Konfidenzintervall mit der Wahrscheinlichkeitsaussage

$$\boldsymbol{P}\left(\overline{X} - z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}} \leq \mu \leq \overline{X} + z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}}\right) = 1 - \alpha$$

ist ein Intervall, dessen Grenzen Zufallsvariablen sind und welches den unbekannten (aber festen) Parameter μ mit der Wahrscheinlichkeit $1 - \alpha$ überdeckt.

Das Intervall mit der Wahrscheinlichkeitsaussage

$$\boldsymbol{P}\left(\mu - z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}} \leq \overline{X} \leq \mu + z_{1-\frac{\alpha}{2}}\sigma_{\overline{X}}\right) = 1 - \alpha$$

hat feste Grenzen und gibt die Wahrscheinlichkeit $1 - \alpha$ an, mit der die Zufallsvariable $\overline{X}$ in den angegebenen Bereich fällt.

4 Spezielle Testverfahren

4.1 Parametertests

In Abschnitt 3.1 wurde der Aufbau eines Parametertests bereits ausführlich erläutert. In diesem Abschnitt wird die Normalverteilungsannahme getroffen, d.h. das Merkmal X ist (approximativ) normalverteilt bzw. bei beliebiger Verteilung der Grundgesamtheit ist die Stichprobe genügend groß ($n > 30$, Anwendung des zentralen Grenzwertsatzes).

4.1.1 Test für den Erwartungswert μ

Prüfgröße

Als Ausgangspunkt bezüglich der Tests für den Erwartungswert[7] wird stets die **standardisierte Prüfgröße** $\frac{\overline{X}-\mu}{\sigma_{\overline{X}}}$ bzw. $\frac{\overline{X}-\mu}{\hat{\sigma}_{\overline{X}}}$ verwendet. Ähnlich wie bei der Bestimmung von Konfidenzintervallen für den Parameter μ einer Grundgesamtheit müssen Fallunterscheidungen vorgenommen werden, die sich auf die Berücksichtigung der Endlichkeitskorrektur beziehen und auf die Frage, ob die Normalverteilung oder Student-Verteilung verwendet werden muss. Es sind also immer die Fragen zu klären:

Stichprobe ohne Zurücklegen

- **Wird eine Stichprobe ohne Zurücklegen aus einer endlichen Grundgesamtheit gezogen?**

  Sofern $\frac{n}{N} \geq 0.5$ ist, muss die sogenannte **Endlichkeitskorrektur**

  $$\frac{N-n}{N-1} \text{ bei } \sigma^2_{\overline{X}} \text{ und } \frac{N-n}{N} = 1 - \frac{n}{N} \text{ bei } \hat{\sigma}^2_{\overline{X}}$$

  berücksichtigt werden (siehe auch Ende des Abschnitts 2.1.2.

- **Ist die Standardabweichung $\boldsymbol{\sigma}$ der Grundgesamtheit bekannt?**

  Gauß-Test

  Bei bekanntem σ wird die (approximativ) $N(0,1)$-verteilte Prüfgröße $Z = \frac{\overline{X}-\mu}{\sigma_{\overline{X}}}$ verwendet. In dem Fall wird der Test als **Gauß-Test** bezeichnet. Muss σ bzw. $\sigma_{\overline{X}}$ mittels der Stichprobenstandardabweichung geschätzt werden, so wird die Prüfgröße $T = \frac{\overline{X}-\mu}{\hat{\sigma}_{\overline{X}}}$ mit $\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}}$ verwendet, die (approximativ) Student-verteilt ist mit $n-1$ Freiheitsgraden. In dem Fall wird der Test als ***t*-Test** bezeichnet.

  ***t*-Test**

[7] siehe interaktive Mathematica-Applets auf der Homepage des Lehrstuhls http://www.fernuni-hagen.de/ls_statistik/forschung/multimedia/

Annahme: $X \sim N(\mu, \sigma^2)$		
	Gauß-Test	t-Test
σ	bekannt	unbekannt
Prüfgröße	$Z = \frac{\overline{X}-\mu_0}{\sigma_{\overline{X}}} \overset{H_0}{\sim} N(0,1)$	$T = \frac{\overline{X}-\mu_0}{\hat{\sigma}_{\overline{X}}} \overset{H_0}{\sim} t(n-1)$
$H_0: \ \mu = \mu_0$	lehne ab, falls $Z < -z_{1-\frac{\alpha}{2}}$ oder $Z > z_{1-\frac{\alpha}{2}}$.	lehne ab, falls $T < -t_{1-\frac{\alpha}{2}}(n-1)$ oder $T > t_{1-\frac{\alpha}{2}}(n-1)$.
$H_0: \ \mu \leq \mu_0$	lehne ab, falls $Z > z_{1-\alpha}$.	lehne ab, falls $T > t_{1-\alpha}(n-1)$.
$H_0: \ \mu \geq \mu_0$	lehne ab, falls $Z < -z_{1-\alpha}$.	lehne ab, falls $T < -t_{1-\alpha}(n-1)$. (für $n > 30$ können in dieser Spalte die t-Quantile durch die z-Quantile ersetzt werden.)

Das Symbol „$\overset{H_0}{\sim}$“ bedeutet, dass unter Gültigkeit der allgemeinen Annahme $\theta = \theta_0$ die angegebene Verteilung vorliegt. Diese Annahme entspricht der zweiseitigen Nullhypothese H_0. Zur Bestimmung der $z-$Werte bzw. t-Werte (Quantile der Standardnormalverteilung bzw. t-Verteilung) kann auch bei den Testverfahren die Tabelle mit häufig vorkommenden z- bzw. t-Werten im Gesamtglossar verwendet werden. Nachfolgende Tests werden nach dem Schema aus Abschnitt 3 durchgeführt.

Beispiel 4.1.1:

Gauß-Test: Stichprobe ohne Zurücklegen mit $\frac{n}{N} < 0.05$

Eine Fabrik stellt Leuchtstoffröhren her, deren Lebensdauer normalverteilt sei mit $\mu = 1500$ h und $\sigma = 100$ h. Eine Änderung des Produktionsverfahrens verspricht bei gleichbleibendem σ eine höhere Lebensdauer.

(1) *Die Lebensdauer X ist normalverteilt. Mit den obigen Angaben gilt $X \sim N(1500, 10000)$.*

(2) *Für den Test auf Erhöhung der Lebensdauer lautet die Nullhypothese, dass sich die Lebensdauer nicht erhöht hat.*

$$H_0: \ \mu \leq \mu_0 = 1500 \quad \text{vs.} \quad H_1: \ \mu > \mu_0$$

(3) *Als maximales Risiko für einen Fehler 1. Art wird $\alpha = 0.05$ vorgegeben.*

(4) *$\overline{X}$ ist $N(1500, \sigma^2_{\overline{X}})$-verteilt. Es liegt eine Stichprobe ohne Zurücklegen vor. Da aber eine laufende Produktion vorliegt, gilt sicherlich $\frac{n}{N} < 0.05$ und somit $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{100}{10} = 10$. Als Prüfgröße wird*

$Z = \frac{\overline{X}-\mu_0}{\sigma_{\overline{X}}} = \sqrt{n}\frac{\overline{X}-\mu_0}{\sigma}$ *verwendet, welche standardnormalverteilt ist.*

(5) *Als kritische obere Grenze ergibt sich* $c_o = z_{0.95} = 1.65$, *so dass* H_0 *abgelehnt wird, falls* $Z > z_{0.95}$.

(6) *In einer Stichprobe ohne Zurücklegen vom Umfang* $n = 100$ *wird ein Mittelwert von* $\overline{x} = 1520$ *Stunden festgestellt, so dass* $z = \frac{1520-1500}{10} = 2$ *ist.*

(7) *Wegen* $z = 2 > 1.65$ *muss die Nullhypothese verworfen werden. Damit ist bei einer Irrtumswahrscheinlichkeit von* $\alpha = 0.05$ *statistisch nachgewiesen, dass sich die Lebensdauer erhöht hat.*

Beispiel 4.1.2:

t-Test: Stichprobe ohne Zurücklegen mit $n < 30$ *und* $\frac{n}{N} < 0.05$

Untersucht wird das Nettoeinkommen von Studenten, welches normalverteilt sei. Es soll die Behauptung des Amtes für Ausbildungsförderung widerlegt werden, dass das durchschnittliche Nettoeinkommen von Studenten in Berlin mindestens 725€ im Monat beträgt.

(1) *Das Nettoeinkommen* X *ist normalverteilt mit unbekanntem* μ *und unbekannter Standardabweichung* σ.

(2) *Es ist*

$$H_0: \ \mu \geq \mu_0 = 725 \ \text{ vs. } \ H_1: \ \mu < \mu_0.$$

(3) *Das Niveau wird mit* $\alpha = 0.05$ *vorgegeben.*

(4) *Da* σ *unbekannt ist, wird* σ *durch die empirische Standardabweichung geschätzt. Es liegt eine Stichprobe ohne Zurücklegen vom Umfang* $n = 20$ *vor, doch sicher gilt* $\frac{n}{N} < 0.05$. *Somit wird der Varianzschätzer* $\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}}$ *verwendet, für den sich der Wert* $\frac{\sqrt{2000}}{\sqrt{20}} = 10$ *ergibt. Mit* $n = 20 < 30$ *wird daher für* $T = \frac{\overline{X}-\mu_0}{\hat{\sigma}_{\overline{X}}} = \sqrt{n}\frac{\overline{X}-\mu_0}{S}$ *die Student-Verteilung mit* $n-1$ *Freiheitsgraden zugrundegelegt.*

(5) *Die Nullhypothese wird abgelehnt, falls* T *kleiner als* $-t_{0.95}(19) = -1.729$ *ist.*

(6) *Aus einer Stichprobe ohne Zurücklegen mit* $n = 20$ *ergibt sich ein Stichprobenmittelwert von* $\overline{x} = 700$, *so dass* $t = \frac{700-725}{10} = -2.5$ *beträgt.*

(7) *Wegen* $t = -2.5 < -1.729$ *muss* H_0 *verworfen werden. Das durchschnittliche Nettoeinkommen liegt also bei einer Irrtumswahrscheinlichkeit von höchstens* 0.05 *unter* 725 €.

Beispiel 4.1.3:

t-Test: Stichprobe ohne Zurücklegen mit $n > 30$ und $\frac{n}{N} \geq 0.05$

Eine Lieferung von 800 *Stahlseilen soll eine garantierte mittlere Reißfestigkeit von mindestens* 5000 *kp haben. Es soll geprüft werden, ob diese Gütezusage in der Lieferung unterschritten worden ist* ($n = 50$).

(1) *Die Verteilung der Reißfestigkeit* X *ist unbekannt. Da* $n = 50 > 30$ *gilt, wird von einer Normalverteilung ausgegangen.*

(2) *Die Nullhypothese und die Alternativhypothese lauten:*

$$H_0: \ \mu \geq \mu_0 = 5000 \ \text{ vs. } \ H_1: \ \mu < \mu_0.$$

(3) *Das Niveau wird mit* $\alpha = 0.05$ *vorgegeben.*

(4) *Da* σ *unbekannt ist, wird* σ *durch* S *geschätzt. Es liegt eine Stichprobe ohne Zurücklegen vor mit* $\frac{n}{N} = \frac{50}{800} = 0.0625 \geq 0.05$. *Somit ergibt sich der Varianzschätzer* $\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N}}$ *zu* $\frac{700}{\sqrt{49}} \cdot \sqrt{\frac{750}{800}} = 95.85$. *Da* $n = 50 > 30$ *gilt, kann für* $T = \frac{\overline{X}-\mu_0}{\hat{\sigma}_{\overline{X}}}$ *anstelle der Student-Verteilung die Normalverteilung zugrundegelegt werden.*

(5) *Die Hypothese* H_0 *wird abgelehnt, falls* T *kleiner als* $-z_{1-\alpha} = -1.65$ *ist.*

(6) *Aus einer Stichprobe ohne Zurücklegen mit* $n = 50$ *ergibt sich* $\overline{x} = 4850$ *und somit* $t = \frac{4850-5000}{95.85} = -1.56$.

(7) *Wegen* $t = -1.56 > -1.65$ *kann damit bei einer Irrtumswahrscheinlichkeit von* $\alpha = 0.05$ *statistisch nicht nachgewiesen werden, dass die Gütezusage in der Lieferung unterschritten worden ist. Die Nullhypothese kann nicht abgelehnt werden.*

4.1.2 Test für die Varianz σ^2

Für Hypothesen über die Varianz einer Grundgesamtheit wird hier vorausgesetzt, dass das interessierende Merkmal normalverteilt ist mit **unbekanntem μ** und **unbekanntem σ**.

Zu testen ist die **Nullhypothese**:

Nullhypothese

a) zweiseitige Fragestellung: $H_0: \ \sigma^2 = \sigma_0^2$

b) einseitige Fragestellung: $H_0: \ \sigma^2 \leq \sigma_0^2$ bzw. $\sigma^2 \geq \sigma_0^2$

Zur Bestimmung der **Prüfgröße** wird die Stichprobenfunktion S^2 zugrundegelegt, für welche die Verteilung nicht direkt angegeben werden kann. Bekannt ist jedoch, dass unter Gültigkeit der Nullhypothese H_0 die transformierte Stichprobenfunktion

Prüfgröße χ^2 (μ unbekannt)

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

Chi-Quadrat-verteilt ist mit $n-1$ Freiheitsgraden. Als oberer bzw. unterer **kritischer Wert** ergeben sich somit die Quantile der Chi-Quadrat-Verteilung.

kritische Werte

In a) wird H_0 abgelehnt, falls $\chi^2 > \chi^2_{1-\frac{\alpha}{2}}(n-1)$ oder $\chi^2 < \chi^2_{\frac{\alpha}{2}}(n-1)$.

In b) wird H_0 abgelehnt, falls $\chi^2 > \chi^2_{1-\alpha}(n-1)$ bzw. $\chi^2 < \chi^2_{\alpha}(n-1)$.

Beispiel 4.1.4:

Eine Konservenfabrik benötigt zur Herstellung eines Produktes Kartoffeln. Aus Verpackungsgründen soll das Gewicht der Kartoffeln nicht zu stark schwanken. Andererseits ist die Fabrik an einem Gewichtsunterschied interessiert, damit die Kartoffeln nicht zu sehr einem Fließbandprodukt ähneln. Sie verlangt deshalb einen mittleren Gewichtsunterschied (Standardabweichung) von 5 *g. Die Nullhypothese und Alternativhypothese lauten:*

$$H_0: \ \sigma^2 = \sigma_0^2 = 25 \quad \text{vs.} \quad H_1: \ \sigma^2 \neq \sigma_0^2.$$

Das Kartoffelgewicht wird als normalverteilt angesehen und für den Test ein Signifikanzniveau von $\alpha = 0.01$ zugrundegelegt.

a) *In einer Stichprobe des Umfangs* $n = 16$ *aus den von Bauer A angebotenen Kartoffeln ergab sich eine Standardabweichung von* $s_A = 3.8$. *Es gilt:*

$$\chi^2_{0.005}(15) = 4.6$$
$$\chi^2_{0.995}(15) = 32.8$$

$$\chi^2 = \frac{(n-1)s_A^2}{\sigma_0^2} = \frac{15 \cdot 3.8^2}{25} = 8.66$$

Wegen $4.6 \leq \chi^2 = 8.66 \leq 32.8$ *kann die Nullhypothese nicht abgelehnt werden. Die Fabrik kann aus der Stichprobe nicht mit ausreichender Sicherheit schließen, dass die Kartoffeln ungeeignet sind bzw. eine von* $\sigma = 5$ *signifikant abweichende Standardabweichung aufweisen.*

b) *In einer Stichprobe des Umfangs* $n = 101$ *aus den von Bauer B angebotenen Kartoffeln ergab sich* $s_B = 6.6$. *Es gilt:*

$$\chi^2_{0.005}(100) = 67.33$$
$$\chi^2_{0.995}(100) = 140.17$$

$$\chi^2 = \frac{(n-1)s_B^2}{\sigma_0^2} = \frac{100 \cdot 6.6^2}{25} = 174.24$$

Wegen $\chi^2 = 174.24 > 140.17$ *ist die Nullhypothese also abzulehnen. Die Kartoffeln sind nicht verwendbar, da statistisch nachgewiesen ist (Irrtumswahrscheinlichkeit* $\alpha = 0.01$*), dass die Standardabweichung von* 5 *g abweicht.*

Anmerkung:

Ist der Parameter $\boldsymbol{\mu}$ der Grundgesamtheit **bekannt**, so wird als Teststatistik

$$\chi^2 = \frac{\sum_{i=1}^{n}(X_i - \mu)^2}{\sigma_0^2}$$

Prüfgröße χ^2 (μ bekannt)

verwendet, welche Chi-Quadrat-verteilt ist mit n Freiheitsgraden. Als oberer bzw. unterer kritischer Wert ergeben sich somit die bisher genannten Quantile der Chi-Quadrat-Verteilung, jedoch mit $\boldsymbol{n}$ **Freiheitsgraden**.

4.1.3 Vergleich zweier Mittelwerte (Differenzentest)

Es liegen zwei verschiedene Grundgesamtheiten mit den Umfängen N und M vor. Betrachtet werden die Merkmale X und Y, welche für die ersten beiden betrachteten Fälle normalverteilt sind. Vorausgesetzt wird hier eine „Stichprobe mit Zurücklegen" bzw. $\frac{n}{N} < 0.05$, $\frac{m}{M} < 0.05$. Im einfachsten Fall soll getestet werden, ob zwischen den unbekannten Parametern der Grundgesamtheiten μ_X und μ_Y ein Unterschied besteht, was gleichbedeutend ist, dass die Differenz $\mu_X - \mu_Y$ von Null abweicht. Daher der Name **Differenzentest**. Die hier betrachtete Nullhypothese lautet im allgemeinen Fall:

Differenzentest

Nullhypothese

$$H_0: \ \mu_X - \mu_Y = \delta_0 \qquad \text{vs.} \qquad H_1: \ \mu_X - \mu_Y \neq \delta_0.$$

Für die Fälle $H_0: \ \mu_X - \mu_Y \geq \delta_0$ und $H_0: \ \mu_X - \mu_Y \leq \delta_0$ lassen sich die Rechnungen leicht übertragen.

Aus jeder Grundgesamtheit wird unabhängig voneinander eine Stichprobe vom Umfang n bzw. m gezogen.

Zur Bestimmung der Prüfgröße wird die Stichprobenfunktion

$$D = \overline{X} - \overline{Y}$$

zugrundegelegt, welche unter H_0 den Erwartungswert

$$\mu_X - \mu_Y = \delta_0$$

besitzt. Wie im Ein-Stichproben-Fall wird hier die standardisierte Prüfgröße

Prüfgröße

$$T = \frac{D - \mu_D}{\sigma_D} = \frac{D - E(D)}{\sqrt{\text{Var}(D)}} = \frac{(\overline{X} - \overline{Y}) - (\mu_X - \mu_Y)}{\sqrt{\text{Var}(\overline{X} - \overline{Y})}}$$

verwendet. Für die Bestimmung der kritischen Werte muss die Verteilung der Prüfgröße unter H_0 bekannt sein.

Fall 1: $\boldsymbol{\sigma_X}$ und $\boldsymbol{\sigma_Y}$ bekannt

Aufgrund der Unabhängigkeit lässt sich mit $\text{Var}(X) = \sigma_X^2$ und $\text{Var}(Y) = \sigma_Y^2$ die Varianz der Differenz D zu $\sigma_D^2 = \sigma_{\overline{X}}^2 + \sigma_{\overline{Y}}^2 = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}$ berechnen, so dass die Teststatistik

$$Z = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}$$

standardnormalverteilt ist. Die Nullhypothese H_0 wird daher abgelehnt, falls $Z < -z_{1-\frac{\alpha}{2}}$ oder $Z > z_{1-\frac{\alpha}{2}}$ gilt.

Fall 2: $\boldsymbol{\sigma_X}$ und $\boldsymbol{\sigma_Y}$ unbekannt und $\boldsymbol{\sigma_X = \sigma_Y}$

Die Standardabweichung der Differenz D wird mit Hilfe der Stichprobenfunktionen S_X^2 und S_Y^2 geschätzt. Die Teststatistik

$$T = \frac{\overline{X} - \overline{Y} - \delta_0}{S_D} = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}}}$$

ist unter der Annahme der **Varianzhomogenität** $(\boldsymbol{\sigma_X = \sigma_Y = \sigma})$ t-verteilt mit $n + m - 2$ Freiheitsgraden. Die Nullhypothese H_0 wird daher abgelehnt, falls $T < -t_{1-\frac{\alpha}{2}}(n+m-2)$ oder $T > t_{1-\frac{\alpha}{2}}(n+m-2)$ gilt.

Varianzhomogenität

Fall 3: $\boldsymbol{n > 30}$ und $\boldsymbol{m > 30}$

Gilt $n > 30$ und $m > 30$, kann die Forderung $\sigma_X^2 = \sigma_Y^2$ und die Normalverteilungsannahme fallengelassen werden. Die Prüfgröße T, auch **Behrens-Fischer-Statistik** genannt,

$$T = \frac{\overline{X} - \overline{Y} - \delta_0}{S_D} = \frac{\overline{X} - \overline{Y} - \delta_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}$$

Behrens-Fischer-Statistik

ist für beliebig verteilte X in dem Fall approximativ standardnormalverteilt, so dass anstelle der t-Quantile die Quantile der Standardnormalverteilung verwendet werden können, d.h. die Nullhypothese H_0 wird abgelehnt, falls $T < -z_{1-\frac{\alpha}{2}}$ oder $T > z_{1-\frac{\alpha}{2}}$ gilt.

Anmerkung:
Sind σ_X und σ_Y unbekannt, $\sigma_X \neq \sigma_Y$ und $n \leq 30$ oder $m \leq 30$, dann ist die Behrens-Fischer-Statistik T approximativ t-verteilt. Die Anzahl der Freiheitsgrade lässt sich nicht direkt angeben, sondern wird anhand des Lösungsansatzes zum „Behrens-Fischer-Problem“ bestimmt [4].

Beispiel 4.1.5:

Autoreifen der Marken A und B sollen bezüglich ihrer Lebensdauer (gemessen in gefahrenen Kilometern) verglichen werden. Zur Überprüfung wird eine Stichprobenerhebung durchgeführt, die folgende Daten liefert:

$$\begin{array}{lll} n = 100 & \overline{x} = 52800\ km & s_X = 4000\ km \\ m = \ \ 45 & \overline{y} = 51500\ km & s_Y = 3000\ km \end{array}$$

Als Signifikanzniveau wird $\alpha = 0.05$ *zugrundegelegt und die Nullhypothese und Alternativhypothese lauten*

$$H_0: \ \mu_X - \mu_Y = 0 \qquad \text{vs.} \qquad H_1: \ \mu_X - \mu_Y \neq 0.$$

Zwar sind σ_X *und* σ_Y *unbekannt, doch es gilt*

$$n = 100 > 30 \quad und \quad m = 45 > 30.$$

Daher ergibt sich

$$t = \frac{52800 - 51500}{\sqrt{\frac{4000^2}{100} + \frac{3000^2}{45}}} = \frac{1300}{600} = 2.167$$

Wegen $2.167 > 1.96 = z_{1-\frac{\alpha}{2}}$ *ist die Nullhypothese abzulehnen. Damit ist bei einer Irrtumswahrscheinlichkeit von* $\alpha = 0.05$ *statistisch nachgewiesen, dass ein Unterschied zwischen den Lebensdauern der Reifen der Marken A und B besteht.*

4.1.4 Vergleich zweier Varianzen (F-Test)

Es wird von zwei verschiedenen, normalverteilten Grundgesamtheiten ausgegangen, deren Mittelwerte μ_X und μ_Y nicht bekannt sind, d.h. X bzw. Y sind $N(\mu_X, \sigma_X^2)$- bzw. $N(\mu_Y, \sigma_Y^2)$-verteilt. Von Interesse ist festzustellen, ob die Varianzen σ_X^2 und σ_Y^2 voneinander abweichen.

Varianzentest (F-Test)

Überprüft werden soll anhand eines **Varianzentest**, auch ***F*-Test** genannt, die Nullhypothese $H_0: \ \sigma_Y^2 = \sigma_X^2$, so dass äquivalent das Testproblem

Nullhypothese

$$H_0: \ \frac{\sigma_Y^2}{\sigma_X^2} = 1 \qquad \text{vs.} \qquad H_1: \ \frac{\sigma_Y^2}{\sigma_X^2} \neq 1$$

betrachtet wird. Der F-Test verwendet die **Prüfgröße**[8]

Prüfgröße F

$$F = \frac{S_Y^2}{S_X^2} \qquad \text{mit} \quad S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

$$\text{und} \quad S_Y^2 = \frac{1}{m-1}\sum_{j=1}^{m}(Y_j - \overline{Y})^2,$$

welche unter Gültigkeit der Nullhypothese F-verteilt ist mit $m-1$ und $n-1$ Freiheitsgraden. Der **obere kritische Wert** entspricht somit dem $(1-\frac{\alpha}{2})$-Quantil der F-Verteilung mit $(m-1, n-1)$ Freiheitsgraden, d.h.

oberer kritischer Wert

$$c_o = F_{1-\frac{\alpha}{2}}(m-1, n-1).$$

Der **untere kritische Wert**, für den $\boldsymbol{P}(F < c_u) = \frac{\alpha}{2}$ gelten soll, ergibt sich aus der Beziehung $F_p(m,n) = \frac{1}{F_{1-p}(n,m)}$, so dass

unterer kritischer Wert

$$c_u = F_{\frac{\alpha}{2}}(m-1, n-1) = \frac{1}{F_{1-\frac{\alpha}{2}}(n-1, m-1)}$$

entspricht. Die Nullhypothese wird abgelehnt, falls $F < c_u$ oder $F > c_o$ gilt. Wird das einseitige Testproblem

$$H_0: \ \frac{\sigma_Y^2}{\sigma_X^2} \le 1 \quad \text{bzw.} \quad H_0: \ \frac{\sigma_Y^2}{\sigma_X^2} \ge 1$$

zugrundegelegt, wird H_0 abgelehnt, falls $F > F_{1-\alpha}(m-1, n-1)$ bzw. $F < F_\alpha(m-1, n-1)$ gilt.

Beispiel 4.1.6:

Ein Großhändler für Obst liefert italienische Orangen (A), die er wegen ihrer geringen Größenunterschiede (Durchmesser approximativ normalverteilt) in Normpackungen kaufen kann. Er erhält ein Angebot von billigeren, spanischen Orangen (B). Die Varianzen der Größen von A und B sollen nach Angabe des Lieferanten gleich sein.

Die zugrunde liegende Nullhypothese lautet:

$$H_0: \ \sigma_A^2 = \sigma_B^2 \text{ bzw. } H_0: \ \frac{\sigma_A^2}{\sigma_B^2} = 1.$$

[8] Mit $\frac{(m-1)}{\sigma_Y^2}S_Y^2 \sim \chi^2(m-1)$ und $\frac{(n-1)}{\sigma_X^2}S_X^2 \sim \chi^2(n-1)$ gilt $\frac{1}{\sigma_Y^2}S_Y^2 / \frac{1}{\sigma_X^2}S_X^2 \sim F(m-1, n-1)$ (s. KE 2 Abschnitt 3.7). Unter Berücksichtigung der Nullhypothese $\frac{\sigma_Y^2}{\sigma_X^2} = 1$ gilt $F = S_Y^2/S_X^2 \overset{H_0}{\sim} F(m-1, n-1)$.

Eine Stichprobe von je 31 *Stück ergab, dass bei gleicher Durchschnittsgröße die Standardabweichung bei italienischen Orangen* $s_A = 0.9$ *und bei spanischen Orangen* $s_B = 1.0$ *beträgt. Mit* $\alpha = 0.1$ *ergeben sich die kritischen Grenzen*

$$\begin{aligned} c_o &= F_{0.95}(30,30) = 1.84, \\ c_u &= F_{0.05}(30,30) = \frac{1}{F_{1-0.95}(30,30)} = \frac{1}{1.84} = 0.543 \end{aligned}$$

Wegen

$$c_u = 0.543 \leq \frac{s_A^2}{s_B^2} = 0.81 \leq 1.84 = c_o$$

kann H_0 *nicht abgelehnt werden. Der Großhändler wird also die spanischen Orangen abnehmen, da ein Unterschied zur Varianz der italienischen Orangen statistisch nicht nachgewiesen werden kann.*

4.1.5 Test für den Anteilswert π (Binomialtest)

Binomialtest

Hypothesen über den Anteilswert π können mittels des **Binomialtests** überprüft werden. Gegeben sei eine dichotome Grundgesamtheit, deren Elemente nur die beiden Ausprägungen $A, \overline{A}$ aufweisen. Jedes Elemente kann als dichotome Zufallsvariable X_i mit

$$X_i = X(\omega_i) = \begin{cases} 1 & \text{für } \omega_i \text{ weist die Eigenschaft } A \text{ auf} \\ 0 & \text{für } \omega_i \text{ weist die Eigenschaft } \overline{A} \text{ auf} \end{cases}$$

Prüfgröße X, P

aufgefasst werden. Für die Prüfung wird eine Stichprobe vom Umfang n gezogen. Als **Prüfgröße** wird die **Anzahl** $X = \sum_{i=1}^n X_i$ der Elemente mit der interessierenden Eigenschaft A in der Stichprobe vom Umfang n oder der **Stichprobenanteilswert** $P = \frac{X}{n}$ verwendet.

exakter Binomialtest

Unter der Voraussetzung einer Stichprobe mit Zurücklegen bzw. einer unendlichen Grundgesamtheit ist X binomialverteilt. Im Folgenden wird zunächst die Vorgehensweise des **exakten Binomialtests** erläutert.

Gegeben sei das **zweiseitige Testproblem**

$$H_0: \ \pi = \pi_0 \ \text{ vs. } \ H_1: \ \pi \neq \pi_0.$$

Mittels der Verteilungsfunktion $F_X(x)$ der Binomialverteilung wird als untere Grenze c_u der Wert x gesucht, bei dem F_X den Wert $\frac{\alpha}{2}$ gerade überschreitet und als c_o wird der Wert x gesucht, bei dem F_X den Wert $1 - \frac{\alpha}{2}$ gerade erreicht oder überschreitet. Die Nullhypothese wird abgelehnt, falls $X < c_u$ oder $X > c_o$ gilt. Für die **kritischen Werte** gilt somit:

kritische Werte (zweiseitig)

$$\boldsymbol{c_u}: \quad F_X(c_u - 1) \leq \frac{\alpha}{2} \quad F_X(c_u) > \frac{\alpha}{2}$$

$$\boldsymbol{c_o}: \quad F_X(c_o) \geq 1 - \frac{\alpha}{2} \quad F_X(c_o - 1) < 1 - \frac{\alpha}{2}$$

Die Werte c_u und c_o sind dabei notwendigerweise ganzzahlig, so dass

$$(\star) \quad \boldsymbol{P}(c_u \leq X \leq c_o) \geq 1 - \alpha$$

gilt. Die Gleichheit in $(\star)$ ist im Allgemeinen nicht erreichbar, da es sich um eine diskrete Verteilung handelt. Ein Test, der das Signifikanzniveau nicht voll ausschöpft, wird als **konservativer Test** bezeichnet.

konservativer Test

Für die **einseitigen Nullhypothesen**

$$H_0: \ \pi \leq \pi_0 \ \text{ bzw. } \ H_0: \ \pi \geq \pi_0$$

wird mittels der Verteilungsfunktion der Binomialverteilung F_X der Wert c_u bzw. c_o gesucht, für den gilt:

kritische Werte (einseitig)

$$\boldsymbol{c_o}: \quad F_X(c_o) \geq 1 - \alpha \quad F_X(c_o - 1) < 1 - \alpha$$

$$\text{bzw.} \quad \boldsymbol{c_u}: \quad F_X(c_u - 1) \leq \alpha \quad F_X(c_u) > \alpha$$

Es gilt dann

$$\boldsymbol{P}(X \leq c_o) \geq 1 - \alpha \quad \text{bzw.} \quad \boldsymbol{P}(X \geq c_u) \geq 1 - \alpha.$$

Auch hier wird die Gleichheit im Allgemeinen nicht erreicht, d.h., das tatsächliche Signifikanzniveau beträgt α^* mit $\alpha^* \leq \alpha$.

Ist der Stichprobenumfang n groß genug ($n\pi \geq 5, n(1-\pi) \geq 5$), so kann der **approximative Binomialtest** durchgeführt werden. Verwendet wird die Teststatistik

approximativer Binomialtest

$$Z = \frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} = \frac{P - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{P - \pi}{\sigma_P},$$

welche unter der Gültigkeit von H_0 approximativ standardnormalverteilt ist. Für das **zweiseitige Testproblem** wird die Nullhypothese abgelehnt, wenn $Z < -z_{1-\frac{\alpha}{2}}$ oder $Z > z_{1-\frac{\alpha}{2}}$ ist. Liegt ein **einseitiges Testproblem** vor, wird als kritischer Wert das Quantil $z_{1-\alpha}$ bzw. $-z_{1-\alpha}$ verwendet.

Beispiel 4.1.7:

Die Wahrscheinlichkeit, beim Werfen einer „fairen“ Münze das Ergebnis „Wappen“ zu erzielen ist $\pi_0 = 0.5$. Für eine unbekannte Münze wird die Hypothese

$$H_0: \ \pi = \pi_0 = 0.5 \quad H_1: \ \pi \neq \pi_0$$

aufgestellt. Die Münze soll durch zehnmaliges Werfen getestet werden. Als Prüfgröße wird die Anzahl X der Würfe „Wappen“ in der Stichprobe verwendet. X ist unter H_0 binomialverteilt.

Für $n = 10$, $\pi_0 = 0.5$ und $\alpha = 0.1$ ergeben sich unter Verwendung der Tabelle der Binomialverteilung die kritischen Werte $c_u = 2, c_o = 8$ ($F_X(1) = 0.0107$, $F_X(2) = 0.0547$, $F_X(8) = 0.9893$, $F_X(7) = 0.9453$).

Beispiel 4.1.8:

Eine studentische Organisation behauptet, bei den kommenden Senatswahlen seien ihr mindestens 30 % der Stimmen sicher. Diese Behauptung soll widerlegt werden ($\alpha = 0.05$).

Es ist $H_0: \ \pi \geq \pi_0 = 0.3$ und $H_1: \ \pi < \pi_0 = 0.3$. Eine Stichprobe vom Umfang $n = 65$ ergibt einen Anteil von 12 Wählern dieser Organisation.

Es ist $n\pi_0 = 19.5 \geq 5, n(1-\pi_0) = 45.5 \geq 5$ und sicher $\frac{n}{N} < 0.05$.

Daher ist

$$Z = \frac{P - \pi_0}{\sqrt{\frac{\pi_o(1-\pi_0)}{n}}}$$

approximativ standardnormalverteilt. Mit $p = \frac{12}{65} = 0.185$ gilt für die Prüfgröße $z = \frac{0.185-0.3}{0.057} = -2.02 < 1.65 = c_u$. Damit ist statistisch nachgewiesen, dass die Behauptung falsch ist bei einer Irrtumswahrscheinlichkeit von $\alpha = 0.05$.

4.2 Nichtparametrische Testverfahren

Die bisher verwendeten Verfahren sind von der Normalverteilungsannahme der Grundgesamtheit ausgegangen. In den dort verwendeten Teststatistiken wurden stets Parameterschätzungen (z.B. $\overline{X}$, S) berücksichtigt. Testverfahren, bei denen keine Annahmen bezüglich des Verteilungstyps der Grundgesamtheit getroffen werden, heißen **nichtparametrische Testverfahren** (verteilungsfreie Testverfahren). Es werden Prüfgrößen verwendet, deren Verteilung von der Grundgesamtheit unabhängig sind. Anstelle der Parameter (z.B. μ, σ^2) werden bestimmte Kenngrößen, wie z.B. der Median X_{med}, betrachtet.

nichtparametrische Testverfahren

Wird eine Hypothese über die Verteilung der Grundgesamtheit aufgestellt, so wird von einem Anpassungstest gesprochen.

> Ein **Anpassungstest** überprüft die Hypothese, dass die Verteilung der Grundgesamtheit durch eine bestimmte Dichtefunktion (Wahrscheinlichkeitsfunktion) bzw. Häufigkeitsverteilung beschrieben werden kann.

Anpassungstest

Ist die Abhängigkeit bzw. Unabhängigkeit zweier Merkmale X und Y von Interesse, so kann dies mit einem Unabhängigkeitstest überprüft werden.

> Ein Test, dem eine Hypothese über die Unabhängigkeit zweier Merkmale X und Y zugrunde liegt, wird als **Unabhängigkeitstest** bezeichnet.

Unabhängigkeitstest

Für beide Problemstellungen wird die Häufigkeitsverteilung einer Zufallsstichprobe ermittelt und die zugrundeliegende Hypothese wird verworfen, wenn die Abweichung zwischen hypothetischer Verteilung und Verteilung der Stichprobe so stark ist, dass die Abweichung nicht mehr als zufällig angesehen werden kann.

Dazu wird ein **Maß für die Abweichungen** zwischen der hypothetischen Verteilung der Grundgesamtheit und der beobachteten Verteilung der Stichprobe benötigt, das als Prüfgröße verwendet werden kann. Sowohl der Anpassungstest als auch der Unabhängigkeitstest verwenden an dieser Stelle eine Prüfgröße, welche Chi-Quadrat verteilt ist.

4.2.1 χ^2-Anpassungstest

Ausgangspunkt ist eine Hypothese über die unbekannte Verteilung einer Zufallsvariablen X, wobei kein bestimmtes Skalenniveau vorausgesetzt werden muss.

Ist X eine **diskrete Zufallsvariable mit wenigen Ausprägungen**, z.B. nominalskaliert oder ordinalskaliert mit m möglichen Ausprägungen $(x_1, ..., x_m)$, dann wird mit

$$h_j = h(x_j) \quad (j = 1, ..., m)$$

die beobachtete absolute Häufigkeit der Ausprägung x_j in der Stichprobe bezeichnet. Mit $\boldsymbol{P}(X = x_j)$ wird die Wahrscheinlichkeit angegeben, dass X der Ausprägung x_j entspricht.

Ist X eine **stetige Zufallsvariable** oder eine **diskrete Zufallsvariable mit vielen Ausprägungen**, erfolgt eine Intervallbildung in m disjunkte, aneinandergrenzende Klassen mit den Klassengrenzen x^*_{j-1} und x^*_j. In dem Fall wird mit

$$h_j = h(x_j) = h(x^*_{j-1} < X \leq x^*_j) \quad (j = 1, ..., m)$$

die beobachtete absolute Häufigkeit der j-ten Klasse in der Stichprobe angegeben. Jede Klasse j wird durch den Wert x_j repräsentiert, so dass die Wahrscheinlichkeit $\boldsymbol{P}(x^*_{j-1} < X \leq x^*_j)$ kurz mit $\boldsymbol{P}(X = x_j)$ angegeben wird.
Durch diese Vereinheitlichung können sowohl im **diskreten** und **stetigen Fall** die Hypothesen

Nullhypothese

$$\begin{aligned} H_0: &\quad \boldsymbol{P}(X = x_j) = \pi_j \text{ für alle } j = 1, ..., m, \\ H_1: &\quad \boldsymbol{P}(X = x_j) \neq \pi_j \text{ für mindestens ein } j = 1, ..., m \end{aligned}$$

aufgestellt werden. Für jede Ausprägung bzw. jedes Intervall kann bei gegebenem Stichprobenumfang n die **erwartete absolute Häufigkeit von $\boldsymbol{X_j}$** unter Gültigkeit der Nullypothese mit $\boldsymbol{n\pi_j}$ angegeben werden.

Der **χ^2-Anpassungstest** vergleicht die beobachteten absoluten Häufigkeiten h_j mit den unter H_0 erwarteten Häufigkeiten $n\pi_j$.

Prüfgröße χ^2

Als **Prüfgröße** wird

$$\chi^2 = \sum_{j=1}^{m} \frac{(h_j - n\pi_j)^2}{n\pi_j}$$

verwendet, die approximativ χ^2-verteilt ist mit $m-1$ Freiheitsgraden[9] unter der **Voraussetzung**, dass für alle erwarteten Häufigkeiten $\boldsymbol{n\pi_j \geq 1, n\pi_j \geq 5}$ **für mindestens 80% der Klassen** gilt. Ansonsten sind in der Kontingenztabelle Spalten oder Zeilen **zusammenzufassen**.

Voraussetzung

Sind die Abweichungen zwischen den beobachteten und den erwarteten Häufigkeiten zu groß, so ist auch der Wert von χ^2 groß. Die Nullhypothese wird abgelehnt, falls χ^2 den **kritischen Wert** $c_o = \chi^2_{1-\alpha}(m-1)$ überschreitet ($\chi^2 > c_o$). Die Nullhypothese wird dagegen nicht abgelehnt, falls $\chi^2 \leq c_o$ gilt.

kritischer Wert

Der Wert $\chi^2_{1-\alpha}(m-1)$ ist der Tabelle der χ^2-Verteilung zur Wahrscheinlichkeit $1-\alpha$ bei $m-1$ Freiheitsgraden zu entnehmen, wobei α das vorzugebende Signifikanzniveau ist.

Beispiel 4.2.1:

Ein Würfel soll überprüft werden, ob alle Augenzahlen gleich wahrscheinlich sind, d.h., ob für die Augenzahl eine Gleichverteilung vorliegt. Eine Stichprobe vom Umfang $n = 300$ liefert die in der nachfolgenden Tabelle angegebenen Häufigkeiten.

Augen-zahl x_j	beobachtete Häufigkeit h_j	erwartete Häufigkeit $n\pi_j$	$h_j - n\pi_j$	$(h_j - n\pi_j)^2$	$\frac{(h_j-n\pi_j)^2}{n\pi_j}$
1	45	50	-5	25	0.5
2	60	50	10	100	2
3	55	50	5	25	0.5
4	40	50	-10	100	2
5	40	50	10	100	2
6	60	50	10	100	2
$\sum$					$\chi^2 = 9$

[9] Im Fall von Kontingenztabellen entspricht die Anzahl der Freiheitsgrade der Anzahl frei wählbarer Zellhäufigkeiten bei festgelegten Randhäufigkeiten. Die Anzahl der Freiheitsgrade wird hier aufgrund der Bedingung $\sum_{j=1}^{m} h_j = n$ um eins reduziert. In der Summe sind $m-1$ Summanden frei wählbar, der m-te muss so gewählt werden, dass die Bedingung $\sum_{j=1}^{m} h_j = n$ erfüllt ist.

Für den kritischen Wert ergibt sich bei einem Signifikanzniveau von $\alpha = 0.05$ *und mit* $m - 1 = 6 - 1 = 5$ *Freiheitsgraden* $\chi^2_{0.95}(5) = 11.07$. *Da* $\chi^2 = 9 \leq 11.07 = \chi^2_{0.95}(5)$ *kann die Nullhypothese „Es handelt sich um einen idealen Würfel" nicht verworfen werden.*

4.2.2 χ^2-Unabhängigkeitstest

Der χ^2-Test kann auch zur Überprüfung der Unabhängigkeit zweier klassierter Merkmale verwendet werden. In diesem Fall wird der Test als **χ^2-Unabhängigkeitstest** bezeichnet. Das Verfahren ist prinzipiell das gleiche wie in Abschnitt 4.2.1. Als **Nullhypothese** wird immer die **Unabhängigkeit** der Merkmale formuliert.

Nullhypothese

$$\begin{aligned} H_0: \quad & \boldsymbol{P}(X = x_j, Y = y_k) = \boldsymbol{P}(X = x_j) \cdot \boldsymbol{P}(Y = y_k) \\ & \text{für alle } j = 1, ..., m,\ k = 1, ..., r, \\ \\ H_1: \quad & \boldsymbol{P}(X = x_j, Y = y_k) \neq \boldsymbol{P}(X = x_j) \cdot \boldsymbol{P}(Y = y_k) \\ & \text{für mindestens ein } j, k. \end{aligned}$$

Mit den Bezeichnungen $\boldsymbol{P}(X = x_j, Y = y_k) = \pi_{jk}$ und $\boldsymbol{P}(X = x_j) = \pi_{j.}$ bzw. $\boldsymbol{P}(Y = y_k) = \pi_{.k}$ ergibt sich

$$H_0: \quad \pi_{jk} = \pi_{j.} \cdot \pi_{.k}$$

als Nullhypothese. Gesucht ist ein Schätzer für π_{jk}, der mit den beobachteten Häufigkeiten verglichen werden kann. Unter Gültigkeit von H_0 lässt sich der Schätzer $\hat{\pi}_{jk}$ als Produkt der Randhäufigkeiten $f_{j.} \cdot f_{.k}$ schreiben. Die geschätzte erwartete absolute Häufigkeit der Merkmalskombination (x_j, y_k) kann unter H_0 mit

$$n\hat{\pi}_{jk} = nf_{j.}f_{.k} = n\frac{h_{j.}}{n}\frac{h_{.k}}{n} = \frac{h_{j.}h_{.k}}{n}$$

angegeben werden.

Zur Veranschaulichung werden hier nochmals die zugrundegelegten Kontingenztabellen aufgeführt (vgl. KE 1 Abschnitt 3.1.3).

	y_1	y_2	...	y_r	$\sum$
x_1	h_{11}	h_{12}	...	h_{1r}	$h_{1.}$
x_2	h_{21}	h_{22}	...	h_{2r}	$h_{2.}$
...	...	...	...	...	...
x_m	h_{m1}	h_{m2}	...	h_{mr}	$h_{m.}$
$\sum$	$h_{.1}$	$h_{.2}$	...	$h_{.r}$	n

	y_1	y_2	...	y_r	$\sum$
x_1	f_{11}	f_{12}	...	f_{1r}	$f_{1.}$
x_2	f_{21}	f_{22}	...	f_{2r}	$f_{2.}$
...	...	...	...	...	...
x_m	f_{m1}	f_{m2}	...	f_{mr}	$f_{m.}$
$\sum$	$f_{.1}$	$f_{.2}$	...	$f_{.r}$	1

Als **Prüfgröße** wird

Prüfgröße χ^2

$$\chi^2 = \sum_{j=1}^{m} \sum_{k=1}^{r} \frac{(h_{jk} - n\hat{\pi}_{jk})^2}{n\hat{\pi}_{jk}}$$

verwendet, die approximativ χ^2-verteilt ist mit $(m-1)(r-1)$ Freiheitsgraden[10]. Die Nullhypothese wird abgelehnt, wenn die Abstände zwischen den beobachteten und den erwarteten Häufigkeiten zu groß sind, d.h. $\boldsymbol{\chi^2 > c_o = \chi^2_{1-\alpha}((m-1)(r-1))}$. H_0 wird nicht abgelehnt falls $\chi^2 \leq \chi^2_{1-\alpha}((m-1)(r-1))$ ist.

kritischer Wert

Auch hier müssen bei Nichtgültigkeit der **Voraussetzung $\boldsymbol{n\hat{\pi}_{jk} \geq 1, n\hat{\pi}_{jk} \geq 5}$ für mindestens 80% der Klassen** für alle j, k, Spalten bzw. Zeilen zusammengefasst werden.

Voraussetzung

Beispiel 4.2.2:

Eine Befragung von 200 Studenten nach der Religionszugehörigkeit X und der Kinderzahl Y hat das folgende Ergebnis geliefert.

Religion	Kinderzahl				
	1	2	3	4	$\sum$
ev.	50	30	15	5	100
kath.	15	10	16	19	60
sonst.	15	10	9	6	40
$\sum$	80	50	40	30	200

[10] Die Anzahl der Freiheitsgrade ergibt sich durch die Bedingungen $\sum_{j=1}^{m} h_{jk} = h_{.k}$ und $\sum_{k=1}^{r} h_{jk} = h_{j.}$. Es sind $m-1$ Zeilensummen und $r-1$ Spaltensummen frei wählbar, so dass insgesamt $(m-1)(r-1)$ Zellhäufigkeiten frei wählbar sind.

Für den Wert der Testgröße $\chi^2 = \sum_{j=1}^{m} \sum_{k=1}^{r} \frac{(h_{jk} - n\hat{\pi}_{jk})^2}{n\hat{\pi}_{jk}}$ *ergibt sich mit der Hilfstabelle*

	h_{j1}	$n\hat{\pi}_{j1}$	h_{j2}	$n\hat{\pi}_{j2}$	h_{j3}	$n\hat{\pi}_{j3}$	h_{j4}	$n\hat{\pi}_{j4}$
h_{1k}	50	*40*	30	*25*	15	*20*	5	*15*
h_{2k}	15	*24*	10	*15*	16	*12*	19	*9*
h_{3k}	15	*16*	10	*10*	9	*8*	6	*6*

der Wert

$$\begin{aligned} \chi^2 &= \frac{100}{40} + \frac{25}{25} + \frac{25}{20} + \frac{100}{15} + \frac{81}{24} + \frac{25}{15} + \frac{16}{12} \\ &\quad + \frac{100}{9} + \frac{1}{16} + \frac{0}{10} + \frac{1}{8} + \frac{0}{6} = 29.1\,. \end{aligned}$$

Der kritische Wert ergibt sich mit $(3-1)\cdot(4-1) = 6$ *Freiheitsgraden und* $\alpha = 0.01$ *zu 16.81 (s. im Glossar die Tabelle der* χ^2*-Verteilung).*

$$\chi^2_{0.99}(6) = 16.81 < \chi^2 = 29.1\,.$$

Die Nullhypothese „Unabhängigkeit der Merkmale“ wird daher abgelehnt. Bei einer Irrtumswahrscheinlichkeit von weniger als 0.01 *ist somit die Abhängigkeit der Merkmale „Religionszugehörigkeit“ und „Kinderzahl“ statistisch nachgewiesen.*

4.2.3 Vorzeichentest

Ein nichtparametrischer Test über die Lage einer Verteilung **(Einstichprobenfall)** ist der **Vorzeichentest**, auch **Mediantest** genannt. Unter der Annahme unabhängiger Stichprobenvariablen $X_1, ..., X_n$ und einer stetigen Verteilung mit mindestens einer Ordinalskalierung wird das **zweiseitige Testproblem**

Nullhypothese

$$H_0: \; X_{med} = \delta_0 \quad \text{vs.} \quad H_1: \; X_{med} \neq \delta_0$$

zugrundegelegt, wobei X_{med} den unbekannten Median der Zufallsvariablen X bezeichnet. Die unabhängigen Stichprobenvariablen werden mit dem vorgegebenen Wert δ_0 verglichen. Die Prüfgröße Z_n mit

$$Z_n = \sum_{i=1}^{n} Z_i \qquad (\text{Anzahl der} X_i < \delta_0)$$

Prüfgröße Z_n

$$Z_i = \begin{cases} 0 & \text{falls } X_i > \delta_0 \\ & \qquad\qquad i = 1, ..., n \\ 1 & \text{falls } X_i < \delta_0 \end{cases}$$

ist unter Gültigkeit von H_0 und unter der Annahme, dass $X_i \neq \delta_0$ für alle i gilt, binomialverteilt mit den Parametern $\pi = 0.5$ und n.

Wegen der Stetigkeit gilt hier unter Gültigkeit der Nullhypothese $\boldsymbol{P}(X_i < \delta_0) = \boldsymbol{P}(X_i > \delta_0) = \frac{1}{2}$ und $\boldsymbol{P}(X_i = \delta_0) = 0$. Aufgrund von Rundungen können in der Stichprobe jedoch Werte $X_i = \delta_0$ auftreten. In diesem Fall bleiben die Werte unberücksichtigt und n wird um die entsprechende Anzahl verringert.

Die Nullhypothese wird abgelehnt, falls $Z_n < c_u$ oder $Z_n > c_o$ ist, mit den **kritischen Werten**:

kritische Werte

$$\boldsymbol{c_u}: \quad F_X(c_u - 1) \leq \frac{\alpha}{2} \quad F_X(c_u) > \frac{\alpha}{2}$$

$$\boldsymbol{c_o}: \quad F_X(c_o) \geq 1 - \frac{\alpha}{2} \quad F_X(c_o - 1) < 1 - \frac{\alpha}{2}$$

Für das **einseitige Testproblem** lassen sich die Überlegungen leicht übertragen.

Ist die Veränderung eines Merkmals von Interesse, wird das Merkmal an denselben Untersuchungseinheiten unter den gegebenen Bedingungen zweimal gemessen z.B. Gewinn einer Kaufhauskette vor und nach Renovierung der Filialen, Preis mehrerer Güter in zwei verschiedenen Supermärkten. In dem Fall wird von einer **verbundenen Stichprobe** gesprochen, d.h. von jeder Untersuchungseinheit liegen zwei Beobachtungen vor **(gepaarte Beobachtungen**). Durch Differenzenbildung kann jedoch auf die Vorgehensweise im Einstichprobenfall zurückgegriffen werden.

verbundene Stichprobe

Beispiel 4.2.3:

Eine Firma möchte den Erfolg zweier Werbekampagnen A und B untersuchen. Dazu wird die Anzahl der Neukunden vor und nach Durchführung der Werbekampagnen in je 20 Städten ermittelt. Es liegen 20 Beobachtungspaare vor, wobei die zugrundeliegenden Städte innerhalb eines jeden Beobachtungspaares miteinander vergleichbar sind. Unter dieser Annahme liegt eine verbundene Stichprobe mit den Wertepaaren (X_i, Y_i) vor $(i = 1, ..., 20)$.

Stadt	1	2	3	4	5	6	7	8	9	10
Werbekampagne A	46	58	50	50	52	46	46	58	55	45
Werbekampagne B	48	49	49	48	45	47	42	56	56	50
Differenz	-2	9	1	2	7	-1	4	2	-1	-5

Stadt	11	12	13	14	15	16	17	18	19	20
Werbekampagne A	48	60	52	40	44	50	50	56	44	60
Werbekampagne B	40	55	49	38	47	45	49	54	42	50
Differenz	8	5	3	2	-3	5	1	2	2	10

Es ist zu testen, ob die Werbekampagnen signifikant unterschiedliche Ergebnisse liefern. Die Nullhypothese lautet, dass beide Werbekampagnen die gleiche durchschnittliche Neukundenanzahl liefern, d.h.

$$H_0: \; D_{med} = 0 \textit{ mit den Zufallsvariablen } D_i = X_i - Y_i.$$

Es sei

$$Z_i = \begin{cases} 1 & \text{falls } D_i > 0 \\ & \\ 0 & \text{falls } D_i < 0 \end{cases} \qquad i = 1, ..., n$$

definiert. Ist die Nullhypothese richtig, dann muss die Anzahl der positiven Differenzen genau so groß sein wie die der negativen Differenzen. Für den Fall $D_i = 0$ bleibt das Wertepaar (X_i, Y_i) unberücksichtigt, und die Anzahl n wird entsprechend verringert. Die Summe

Prüfgröße D_n

$$D_n = \sum_{i=1}^{n} Z_i$$

*ist also $B(n, 0.5)$-verteilt und wird als **Prüfgröße** des Vorzeichentests bei **verbundenen Messungen** verwendet. D_n entspricht der Anzahl der positiven Differenzen. Konkret ergibt sich hier für die Anzahl D_n der positiven Vorzeichen*

$$d_n = 15.$$

D_n ist $B(20, 0.5)$-verteilt, so dass sich bei einem Signifikanzniveau von 0.05 *die kritischen Werte zu*

$$c_u = 6 \quad \textit{und} \quad c_o = 14$$

ergeben. Da $d_n = 15 > 14 = c_o$ ist, wird die Nullhypothese abgelehnt.

Zu beachten ist, dass der Vorzeichentest nur Relationen der Art „größer/kleiner“ bzw. „ja/nein“ verwendet. Weitere Informationen z.B. die Rangordnung beim Vorliegen einer Ordinalskala werden nicht genutzt.

4.2.4 Wilcoxon-Vorzeichen-Rang-Test

Ein Test, der nicht nur die Richtung der Beobachtungen in Form des Vorzeichens, sondern auch die Größe der Beobachtungen in Form des Betrages berücksichtigt, ist der **Wilcoxon-Vorzeichen-Rang-Test**. Vorausgesetzt wird hier neben einer stetigen symmetrischen Verteilung, dass die Stichprobenvariablen $X_1, ..., X_n$ unabhängig identisch verteilt sind und X metrisch skaliert ist. Im **Einstichprobenfall** wird wie beim einfachen Vorzeichentest das **zweiseitige Testproblem**

$$H_0: \; X_{med} = \delta_0 \quad \text{vs.} \quad H_1: \; X_{med} \neq \delta_0$$

Nullhypothese

zugrundegelegt. Betrachtet werden die Ränge der Beträge der Differenzen $D_i = X_i - \delta_0$, d.h. $\text{rg}|D_i|$. Als Prüfgrößen können

$$\begin{aligned} W^+ &= \sum_{i=1}^{n} \text{rg}|D_i| Z_i \\ W^- &= \sum_{i=1}^{n} \text{rg}|D_i|(1 - Z_i) \end{aligned}$$

$$Z_i = \begin{cases} 1 & \text{falls } D_i > \delta_0 \\ 0 & \text{falls } D_i < \delta_0 \end{cases} \qquad i = 1, ..., n$$

Prüfgröße W^+, W^-

verwendet werden. Die Summe W^+ (W^-) berücksichtigt alle Ränge, für dessen Beobachtungen $X_i > \delta_0$ ($X_i < \delta_0$) gilt. Oft wird als Prüfgröße die jeweils kleinere Realisation der beiden Summen ausgewählt ($\min(W^+, W^-)$). An dieser Stelle wird W^+ als Prüfgröße verwendet, so dass die Nullhypothese abgelehnt wird, wenn W^+ zu groß oder zu klein ist bzw. formal wenn

$$W^+ < w_{\frac{\alpha}{2}}(n) \text{ oder } W^+ > w_{1-\frac{\alpha}{2}}(n) = \frac{n(n+1)}{2} - w_{\frac{\alpha}{2}}(n)$$

gilt. Die kritischen Werte $w_\alpha(n)$ können der folgenden Tabelle entnommen werden.

n	$w_{0.01}(n)$	$w_{0.025}(n)$	$w_{0.05}(n)$	$w_{0.1}(n)$	$w_{0.9}(n)$	$w_{0.95}(n)$	$w_{0.975}(n)$	$w_{0.99}(n)$
4	0	0	0	1	9	10	10	10
5	0	0	1	3	12	14	15	15
6	0	1	3	4	17	18	20	21
7	1	3	4	6	22	24	25	27
8	2	4	6	9	27	30	32	34
9	4	6	9	11	34	36	39	41
10	6	9	11	15	40	44	46	49
11	8	11	14	18	48	52	55	58
12	10	14	18	22	56	60	64	68
13	13	18	22	27	64	69	73	78
14	16	22	26	32	73	79	83	89
15	20	26	31	37	83	89	94	100
16	24	30	36	43	93	100	106	112
17	28	35	42	49	104	111	118	125
18	33	41	48	56	115	123	130	138
19	38	47	54	63	127	136	143	152
20	44	53	61	70	140	149	157	166

Tabelle 4.2.1: *Kritische Werte für den Wilcoxon-Vorzeichen-Rang-Test*

Für große Stichprobenumfänge ($n > 20$) kann die Approximation durch die Standardnormalverteilung verwendet werden. Unter der Gültigkeit von H_0 ist zu erwarten, dass W^+ und W^- gleich groß sind. Es gilt unter der Nullhypothese für den Erwartungswert $E(W^+) = \frac{n(n+1)}{4}$ und für die Varianz $\text{Var}(W^+) = \frac{n(n+1)(2n+1)}{24}$[11]. Die Teststatistik

$$W^{+*} = \frac{W^+ - E(W^+)}{\sqrt{\text{Var}(W^+)}} = \frac{W^+ - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

ist für $n > 20$ approximativ standardnormalverteilt, so dass die Hypothese H_0 abgelehnt wird, falls $W^{+*} < -z_{1-\frac{\alpha}{2}}$ oder $W^{+*} > z_{1-\frac{\alpha}{2}}$ gilt. Für das **einseitige Testproblem** lassen sich die Überlegungen leicht übertragen.

[11] $W^+ = \sum_{i=1}^n \text{rg}|D_i| Z_i$, wobei Z_i bernoulliverteilt ist mit Erwartungswert $\frac{1}{2}$ und Varianz $\frac{1}{4}$. Somit ist $E(W^+) = \sum_{i=1}^n \text{rg}|D_i| E(Z_i) = \sum_{i=1}^n iE(Z_i) = \frac{1}{2}\sum_{i=1}^n i = \frac{1}{2} \cdot \frac{n(n+1)}{2}$ und $\text{Var}(W^+) = \sum_{i=1}^n i^2 \text{Var}(Z_i) = \frac{1}{4}\sum_{i=1}^n i^2 = \frac{1}{4} \cdot \frac{n(n+1)(2n+1)}{6}$.

Wie der Vorzeichentest, kann auch der Wilcoxon-Vorzeichen-Rang-Test für zwei verbundene Stichproben $X_1, ..., X_n$ und $Y_1, ...Y_n$ verwendet werden. Analog dazu werden anstelle der Differenzen $D_i = X_i - \delta_0$ die Differenzen $D_i = X_i - Y_i$ und die Nullhypothese

$$H_0: \ D_{med} = 0$$

betrachtet. Ansonsten wird die Überprüfung der Nullhypothese in gleicher Weise durchgeführt.

Beispiel 4.2.4:

Es wird auf die Angaben aus Beispiel 4.2.3 (Beispiel Werbekampagne in Stadt A und Stadt B) zurückgegriffen.

Stadt-Nr.	Differenz #Kunden (D_i)	$rg\|D_i\|$	$D_i > 0$	$D_i < 0$
1	-2	8		8
2	9	19	19	
3	-1	2.5		2.5
4	2	8	8	
5	7	17	17	
6	-1	2.5		2.5
7	4	14	14	
8	2	8	8	
9	-1	2.5		2.5
10	-2	8		8
11	8	18	18	
12	5	15.5	15.5	
13	3	12.5	12.5	
14	2	8	8	
15	-3	12.5		12.5
16	5	15.5	15.5	
17	-1	2.5		2.5
18	2	8	8	
19	2	8	8	
20	10	20	20	
			171.5	38.5
			W^+	W^-

Die Testgröße W^+ lautet 171.5. Zum Signifikanzniveau $\alpha = 0.05$ werden aus der Tabelle 4.2.1 die kritischen Werte $c_u = w_{\frac{\alpha}{2}}(n) = 53$ und $c_o = w_{1-\frac{\alpha}{2}}(n) = 157$ entnommen. Es gilt $171.5 > 157$, so dass die Nullhypothese abgelehnt wird.

Hinweis:
Wird nur das Minimum der beiden Rangsummen W^+ und W^- zur Entscheidung herangezogen, reicht es aus, den unteren kritischen Wert $w_{\frac{\alpha}{2}}(n)$ zu betrachten. Liegt W^+ unter $w_{\frac{\alpha}{2}}(n)$, dann liegt W^- über der oberen kritischen Grenze $w_{1-\frac{\alpha}{2}}(n)$ und umgekehrt.

4.2.5 Wilcoxon-Rangsummen-Test

Liegen zwei unabhängige Stichproben $X_1, ..., X_n$ und $Y_1, ..., Y_m$ vor, so dass die Stichprobenvariablen $X_1, ..., X_n, Y_1, ..., Y_m$ unabhängig sind, kann mittels des **Wilcoxon-Rangsummen-Tests** ein Vergleich der Mittelwerte durchgeführt werden. Unter der Annahme X und Y besitzen stetige Verteilungsfunktionen, wird das **zweiseitige Testproblem**

Nullhypothese

$$H_0: \quad X_{med} = Y_{med} \quad \text{vs.} \quad H_1: \quad X_{med} \neq Y_{med}$$

zugrundegelegt. Unter der Gültigkeit von H_0 müssen beide Stichproben ähnliche Werte aufweisen, d.h in keiner Stichprobe befinden sich mehr kleinere bzw. größere Werte. Wie schon beim Wilcoxon-Vorzeichen-Rang-Test werden anstelle der Stichprobenwerte die Ränge betrachtet. Dazu werden hier beide Stichproben zu einer Gesamtstichprobe **(gepoolte Stichprobe)** zusammengefasst und die Ränge $\text{rg}(X_1), ..., \text{rg}(Y_m)$ gebildet. Treten Bindungen zwischen den X und Y Werten auf, werden Durchschnittsränge gebildet. Bindungen innerhalb der gleichen Stichprobe sind nicht relevant, so dass die Rangzuweisung in dem Fall zufällig erfolgen kann. Als Prüfgröße wird

gepoolte Stichprobe

Prüfgröße $W_{n,m}$

$$W_{n,m} = \sum_{i=1}^{n} \text{rg}(X_i) = \sum_{i=1}^{n+m} iV_i$$

$$V_i = \begin{cases} 1 & \text{falls die } i\text{-te Beobachtung} \\ & \text{eine } X\text{-Variable ist} \\ 0 & \text{sonst} \end{cases}$$

verwendet. Ist $W_{n,m}$ zu groß oder zu klein, wird die Nullhypothese abgelehnt. Die kritischen Werte $w_{\frac{\alpha}{2}}(n,m)$ und $w_{1-\frac{\alpha}{2}}(n,m)$ können aus den Tabellen im Glossar entnommen werden, wobei die Beziehung

$$w_{1-\frac{\alpha}{2}}(n,m) = n(n+m+1) - w_{\frac{\alpha}{2}}(n,m)$$

gilt. Die Nullhypothese wird abgelehnt, falls $W_{n,m} < w_{\frac{\alpha}{2}}(n,m)$ oder $W_{n,m} > w_{1-\frac{\alpha}{2}}(n,m)$ ist.

Auch hier kann die Teststatistik durch die Normalverteilung approximiert werden. Für große Stichproben (m oder $n > 25$) ist die Teststatistik

$$W^*_{n,m} = \frac{W_{n,m} - E(W_{n,m})}{\sqrt{\mathrm{Var}(W_{n,m})}} = \frac{W_{n,m} - \frac{n(n+m+1)}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}}$$

approximativ standardnormalverteilt.[12]

Beispiel 4.2.5:

Ein Berufsschullehrer vergleicht die Ergebnisse einer Lernstandserhebung zweier Klassen zum Niveau $\alpha = 0.1$. Die Rangzahlen sind für die gepoolte Stichprobe vergeben worden ($n + m = 26$).

	Punktzahl														$\sum$
X	78	41	55	82	67	28	45	73	39	31	46	68	53	57	
rg($\boldsymbol{X}$)	**24**	**6**	**13**	**25**	**19**	**1**	**8**	**22**	**5**	**2**	**9**	**20**	**12**	**15**	**181**
Y	56	44	65	92	77	51	64	33	59	48	69	38			
rg($\boldsymbol{Y}$)	**14**	**7**	**18**	**26**	**23**	**11**	**17**	**3**	**16**	**10**	**21**	**4**			**170**

Somit ist $W_{14,12} = 181$ und die kritischen Werte ergeben sich laut Tabelle zu $w_{0.95}(14,12) = 248$ und $w_{0.05}(14,12) = 130$. Da $130 \le W_{14,12} \le 248$ gilt, wird die Nullhypothese, beide Klassen weisen den gleichen Leistungsstand auf, nicht abgelehnt.

[12] Unter Gültigkeit der Nullhypothese gilt $P(V_i = 1) = \frac{n}{n+m}$, $P(V_i = 0) = \frac{m}{n+m}$, $E(V_iV_j) = P(V_i = 1, V_j = 1) = \frac{n(n-1)}{(n+m)(n+m-1)}$ und somit $E(V_i) = E(V_i^2) = \frac{n}{n+m}$, $\mathrm{Var}(V_i) = E(V_i^2) - (E(V_i))^2 = \frac{nm}{(n+m)^2}$ und $\mathrm{Cov}(V_i, V_j) = -\frac{nm}{(n+m)^2(n+m-1)}$ (vgl. [3]).

Übungsaufgaben

Aufgaben zu Kapitel 1

Übungsaufgabe 1.1:

Zeigen Sie, dass die Stichprobenvarianz

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

eine erwartungstreue Schätzfunktion für die Varianz σ^2 *einer normalverteilten Grundgesamtheit ist.*

Übungsaufgabe 1.2:

Zeigen Sie, dass der Stichprobenanteilswert

$$P = \frac{1}{n}\sum_{i=1}^{n} X_i$$

eine erwartungstreue Schätzfunktion für den unbekannten Anteilswert π *einer binomialverteilten Grundgesamtheit ist.*

Übungsaufgabe 1.3:

Zeigen Sie, dass der Stichprobenmittelwert

$$\overline{X} = \sum_{i=1}^{n} X_i$$

eine effiziente Schätzfunktion für den unbekannten Mittelwert μ *einer normalverteilten Grundgesamtheit ist.*

Die beiden folgenden Aufgaben sind erst nach Bearbeitung des Kapitels „Konfidenzintervall für den Parameter σ^2 eines normalverteilten Merkmals“ zu lösen.

Übungsaufgabe 1.4:

Gegeben sei eine normalverteilte Grundgesamtheit mit bekanntem Erwartungswert μ und unbekannter Varianz σ^2. Vergleichen Sie die Schätzer $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ und $S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ bezüglich ihrer Effizienz.

Übungsaufgabe 1.5:

Zeigen Sie, dass für einfache Zufallsstichproben

a) *der Stichprobenmittelwert $\overline{X}$ für den unbekannten Parameter μ einer normalverteilten Grundgesamtheit,*

b) *der Stichprobenanteilswert P für den unbekannten Anteilswert π einer dichotomen Grundgesamtheit,*

c) *die Stichprobenvarianz $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ für die unbekannte Varianz σ^2 einer normalverteilten Grundgesamtheit*

eine im quadratischen Mittel konsistente Schätzfunktion ist.

Aufgaben zu Kapitel 2

Übungsaufgabe 2.1:

Aus früheren Untersuchungen ist bekannt, dass das verfügbare Monatseinkommen von Studenten normalverteilt ist, mit einer Standardabweichung von $\sigma = 150$.

In einer Zufallsstichprobe vom Umfang $n = 225$ *aus den rund* 40000 *Studenten einer großen Universität, wird ein durchschnittliches verfügbares Monatseinkommen der Studenten von* $\overline{x} = 710$ *ermittelt. Bestimmen Sie ein* 95%*-Konfidenzintervall für das Durchschnittseinkommen aller Studierenden.*

Übungsaufgabe 2.2:

a) *Von* 1000 *Studenten werden* 36 *zufällig ausgewählte Studenten nach dem Alter befragt. Es wird ein Durchschnittsalter von* $\overline{x} = 23$ *Jahren mit einer Stichprobenstandardabweichung von* $s = 0.6$ *Jahren ermittelt. Bestimmen Sie ein* 90%(95%)*-Konfidenzintervall für das Durchschnittsalter. Es kann von einer normalverteilten Grundgesamtheit ausgegangen werden.*

b) *Welches Konfidenzintervall ergibt sich, wenn aus* 250 *Studenten* 25 *befragt werden?*

Übungsaufgabe 2.3:

Ein Pharmaunternehmen, das sich auf Vitaminpräparate spezialisiert hat, möchte die monatlichen Durchschnittsausgaben der 1000 *Beschäftigten für firmeneigene Produkte bestimmen, wobei von der Normalverteilungsannahme ausgegangen wird. Dazu wird eine Stichprobe mit Zurücklegen vom Umfang* $n = 64$ *genommen. Es wird ein Durchschnittswert von* 30 € *bei einer Standardabweichung von* 1.6 € *festgestellt.*

a) *Bestimmen Sie ein* 95%*-Konfidenzintervall für die monatlichen Durchschnittsausgaben sämtlicher Beschäftigten.*

b) *Nehmen Sie an, der Stichprobenumfang beträgt nur* $n = 36$*. Wie groß ist das gesuchte* 95%*-Konfidenzintervall?*

c) *Nach einem Jahr sollen bei einem anderen Standort des Pharmaunternehmens mit* 1500 *Beschäftigten wiederum die monatlichen Durchschnittsausgaben durch eine Stichprobe mit Zurücklegen ermittelt werden. Aus früheren Messungen kann von einer Standardabweichung von* 2.1 € *ausgegangen werden. Die Stichprobe vom Umfang* $n = 49$ *ergibt einen Durchschnittswert von* 28 €*. Bestimmen Sie ein* 90%*-Konfidenzintervall für die monatlichen Durchschnittsausgaben aller Beschäftigten.*

Übungsaufgabe 2.4:

Aus einer Produktionsserie werden durch eine einfache Zufallsstichprobe mit Zurücklegen 200 *Stück entnommen. Es wird festgestellt, dass* 80 *Stück den Anforderungen nicht entsprechen. Geben Sie an, innerhalb welcher Grenzen der Anteil mangelhafter Stücke mit einer Wahrscheinlichkeit von* 0.9 *erwartet werden kann.*

Übungsaufgabe 2.5:

Ein Baustoffhändler bezieht Fliesen. Unter 150 *Fliesen, die er der Lieferung entnimmt, befinden sich* 30 *Fliesen 2. Wahl.*

Bestimmen Sie ein 90%*-Konfidenzintervall für den Anteil der Fliesen 2. Wahl in der Lieferung.*

Übungsaufgabe 2.6:

Untersucht wird das Einkommen von Studenten eines Fachbereichs einer bestimmten Universität. Dazu wird eine Stichprobe vom Umfang $n = 50$ *betrachtet. Als empirische Varianz ergibt sich* $s^2 = 191844$. *Bestimmen Sie unter der Normalverteilungsannahme das* 90 %*-Konfidenzintervall für die Varianz.*

Aufgaben zu Kapitel 3

Übungsaufgabe 3.1:

Beantworten Sie die oben angeführten Fragen a), b) und c) in Abschnitt 3.1.1 für die folgenden Probleme:

1) *Ein Händler will feststellen, ob das mittlere Gewicht einer Lieferung von* 3000 *Eiern tatsächlich dem Sollwert* 60 *g entspricht.*

2) *Ein Schausteller betreibt ein Glücksrad. Ein „notorischer“ Spieler behauptet, dass von den* 20 *Zahlen die „*13*“ mit der Wahrscheinlichkeit* 0.1 *auftritt.*

Übungsaufgabe 3.2:

An einer Schule soll eine neue Methode zum Vokabellernen getestet werden. Nach der alten Methode wurden im Wochenmittel $\mu_0 = 25$ *neue Vokabeln gelernt. Lehrer A und Lehrer B sollen unabhängig voneinander die neue Methode überprüfen. Formulieren Sie für die zwei nachstehenden Testprobleme eine sinnvolle Nullhypothese und Alternativhypothese.*

a) Lehrer A möchte die neue Methode in Zukunft anwenden, falls sie nicht nachweisbar schlechter ist.

b) Lehrer B würde nie eine neue Methode einführen, ohne sicher zu sein, dass sie die alte deutlich übertrifft.

Übungsaufgabe 3.3:

Formulieren Sie zu jedem der nachstehenden Testprobleme eine sinnvolle Nullhypothese und Alternativhypothese.

a) Es soll überprüft werden, ob der durchschnittliche Intelligenzquotient einer Probandengruppe A ($\overline{IQ}_A$) größer ist als der einer Probandengruppe B ($\overline{IQ}_B$).

b) Ein Hersteller von Motorblöcken möchte wissen, ob der zugesagte mittlere Bohrungsdurchmesser von 78.65 mm in der laufenden Produktion noch eingehalten wird.

c) Die Betonmischer einer bestimmten Firma haben nach Herstellerangaben einen Benzinverbrauch von nur $\mu_0 = 1.2\ l/h$ bei einer Standardabweichung von $\sigma = 0.05\ l/h$. Ein Konkurrent der Firma würde sich freuen, wenn er mit Hilfe einer Stichprobe einen höheren Durchschnittsverbrauch nachweisen könnte.

Übungsaufgabe 3.4:

Die Bauern A und B wollen durch einen statistischen Test prüfen, ob ihre Äpfel das Durchschnittsgewicht der Klasse „K“ erfüllen. Dazu entnehmen sie ihrer Apfelernte je eine Stichprobe und wiegen sie. Das Mittelgewicht von Äpfeln der Handelsklasse „K“ beträgt $\mu_0 = 150\ g$ bei $\sigma = 10\ g$.

Formulieren Sie für die beiden folgenden Fälle Nullhypothese und Alternativhypothese:

1) A will die Äpfel, um seinem guten Ruf nicht zu schaden, nur dann verkaufen, wenn er sicher ist, dass sie im Mittel das Mindestgewicht überschreiten.

2) B's Ruf ist schon längst ruiniert, er befürchtet aber strafrechtliche Konsequenzen, wenn er Äpfel verkauft, deren mittleres Gewicht unter der Qualitätsnorm liegt.

Übungsaufgabe 3.5:

Durch einen Test soll die Hypothese „Der Anteil π *der durch Krankheit ausgefallenen Arbeitsstunden in der Bundesrepublik Deutschland ist im April des laufenden Jahres niedriger als im April des Vorjahres (V)“ überprüft werden. Die Nullhypothese „Es ist keine Verringerung eingetreten* $(H_0: \ \pi \geq \pi_0 = \pi(V))$*“ kann bei einem Signifikanzniveau von* $\alpha = 0.05$ *nicht verworfen werden.*

Das bedeutet,

A) *die Nullhypothese ist damit statistisch widerlegt.*

B) *bei einem Signifikanzniveau von* $\alpha = 0.05$ *ist eine Verringerung nicht statistisch nachweisbar.*

C) *mit einer Wahrscheinlichkeit von* 0.05 *kann dennoch eine Verringerung vorliegen.*

D) *die Anzahl der durch Krankheit ausgefallenen Arbeitsstunden hat sich nicht verringert.*

Welche der Aussagen sind zutreffend?

Übungsaufgabe 3.6:

Eine Mensa bezieht Brötchen von einer Großbäckerei. Die Großbäckerei garantiert ein mittleres Gewicht von mindestens $\mu_0 = 45$ *g bei einer Standardabweichung von* $\sigma = 2$ *g.*

Die Mensa unterzieht die tägliche große Lieferung einer Abnahmeprüfung.

a) *Wie lauten Null- und Alternativhypothese?*

b) *Bestimmen Sie den Ablehnungsbereich für* $\alpha = 0.05$ *und den Stichprobenumfang* $n = 25$ *unter Verwendung der Normalverteilung.*

c) *Eine Stichprobe liefert* $\overline{x} = 44$ *g. Wie wird entschieden?*

d) *Wie groß ist das Risiko des Herstellers, dass eine Lieferung mit* $\mu_0 = 45$*, die also gerade noch den Anforderungen entspricht, zurückgewiesen wird (*α*-Fehler)?*

e) *Bestimmen Sie den hier vorliegenden p-Wert* $(\overline{x} = 44)$.

Aufgaben zu Kapitel 4

Übungsaufgabe 4.1:

Ein Automobilproduzent A behauptet, dass die Laufleistung eines bestimmten Models mindestens $\mu_0 = 100000\,km$ *beträgt, bei einer Standardabweichung von* $\sigma = 30000$ *km. Ein konkurrierender Automobilproduzent B möchte diese Behauptung widerlegen* ($\alpha = 0.05$). *Der Automobilproduzent B ermittelt in einer Stichprobe mit Zurücklegen vom Umfang* $n = 36$ *eine durchschnittliche Laufleistung von* $\overline{x} = 90000$ *km.*

Übungsaufgabe 4.2:

Überprüft werden soll, ob die mittlere Füllmenge μ *von Flaschen bei einer laufenden Produktion von* 500 *ml abweicht. Eine Stichprobe vom Umfang* $n = 25$ *liefert* $\overline{x} = 490$ *und* $s = 25$. *Die Füllmenge* X *ist approximativ normalverteilt. Bestimmen Sie den Ablehnungsbereich* ($\alpha = 0.05$).

Wie entscheiden Sie sich?

- ☐ *Die Nullhypothese wird verworfen.*
- ☐ *Die Nullhypothese wird nicht verworfen.*

Welche Interpretation Ihrer Entscheidung ist richtig?

- ☐ *Die Nullhypothese ist statistisch abgesichert.*
- ☐ *Die Alternativhypothese ist statistisch abgesichert.*
- ☐ *Dass die Alternativhypothese zutrifft, kann ausgeschlossen werden.*
- ☐ *Dass die Nullhypothese zutrifft, ist nicht statistisch abgesichert.*

Übungsaufgabe 4.3:

Bei einer im Vorjahr durchgeführten Vollerhebung der Altersstruktur der Bevölkerung einer Großstadt wurde ein Durchschnittsalter von 38 *Jahren bei einer Standardabweichung von* 8 *Jahren festgestellt. Durch eine im laufenden Jahr durchgeführte Stichprobenerhebung vom Umfang* $n = 100$ *soll geprüft werden, ob eine signifikante Änderung des Durchschnittsalters eingetreten ist. In der Stichprobe wird ein Durchschnittsalter von* $\overline{x} = 39$ *Jahren ermittelt. Es ist ein Signifikanzniveau von* 5% *zugrunde zu legen. Die bei der Erhebung im Vorjahr ermittelte Standardabweichung von* $\sigma = 8$ *kann als gleichbleibend vorausgesetzt werden.*

Übungsaufgabe 4.4:

Der Hersteller einer Drehmaschine gibt an, dass seine Maschine sehr genau arbeitet. Er belegt diese Behauptung mit der Angabe, dass die approximativ normalverteilten Durchmesser der gedrehten Teile eine Varianz von $\sigma^2 = 0.01$ *haben. Eine Versuchsserie des Käufers vom Umfang* 32 *ergab eine empirische Varianz von* $s^2 = 0.012$. *Kann die Angabe des Herstellers als falsch angesehen werden* ($\alpha = 0.05$)*?*

Übungsaufgabe 4.5:

Eine neue Abfüllanlage, mit der eine Flüssigkeit in 0.5 *l Flaschen abgefüllt wird, soll eine geringe Abweichung der Füllmenge aufweisen. Die Füllmenge sei normalverteilt. Überprüft werden soll die Nullhypothese, dass eine durchschnittliche Abweichung vom Sollwert höchstens* 10 *ml beträgt* ($\alpha = 0.1$). *Eine Stichprobe von* 50 *Flaschen ergab einen Mittelwert von* 498 *ml und eine empirische Varianz von* 125.

Übungsaufgabe 4.6:

Zwei Wäschetrockner vom Typ X und Typ Y werden bezüglich ihres Energieverbrauchs miteinander verglichen. Untersucht werden soll zum Niveau 0.05 *ob sich signifikante Unterschiede nachweisen lassen. Von Typ X (Y) wurden* $n = 15$ $(m = 10)$ *Trockner bei einem durchschnittlichen Verbrauch von* $\overline{x}_x = 3.15$ $(\overline{x}_y = 2.55)$ *kWh und einer empirischen Standardabweichung von* $s_x^2 = 0.6$ $(s_y^2 = 0.4)$ *kWh betrachtet. Es kann von der Normalverteilungsannahme und der Varianzhomogenität* $(\sigma_X = \sigma_Y = \sigma)$ *ausgegangen werden.*

Übungsaufgabe 4.7:

Ein Mentor des Fachs Statistik hat in zwei verschiedenen Städten Probeklausuren durchgeführt. Dabei wurde die durchschnittliche Zeit bestimmt, in der die Studenten eine Aufgabe gelöst haben. Zum Niveau 0.1 *soll die Nullhypothese, die Varianzen beider Gruppen sei gleich, überprüft werden* $(n = 15, m = 20, s_x = 5.3, s_y = 3.5)$.

Übungsaufgabe 4.8:

Aufgrund von erhöhten Kreditverlusten, möchte eine Bank die Bewertungskriterien bei der Kreditvergabe verbessern. Ungeachtet des Risikos, keine Veränderung der Bewertungskriterien vorzunehmen, obwohl der Anteil über 15% *liegt (Fehler 2. Art), möchte die Bank nur in eine Verbesserung der Bewertungskriterien investieren, wenn statistisch nachgewiesen werden kann, dass der Anteil von gewährten Krediten mit Rückzahlungsschwierigkeiten mehr als* 15% *beträgt. In einer Stichprobe ohne Zurücklegen vom Umfang* $n = 20$ *aus* $N = 500$ *Kreditnehmern sind* $x = 7$ *Kreditnehmer mit Rückzahlungsschwierigkeiten (Signifikanzniveau* $\alpha = 0.05$*).*

Übungsaufgabe 4.9:

In einer Großstadt soll der Anteil der Beamten unter den Erwerbstätigen mindestens 20 % *betragen. In einer Stichprobe vom Umfang* $n = 100$ *befinden sich* 15 *Beamte. Kann die Behauptung bei einem Signifikanzniveau von* $\alpha = 0.05$ *widerlegt werden?*

Übungsaufgabe 4.10:

Ein Hersteller von Sicherungsautomaten behauptet, der Anteil der funktionsuntüchtigen Sicherungsautomaten betrage höchstens 0.1*. Diese Behauptung soll in einer Abnahmeprüfung mit einer Stichprobe des Umfangs* $n = 144$ *geprüft werden. Als Nullhypothese wird* $H_0: \ \pi \leq \pi_0 = 0.1$ *angegeben* ($\alpha = 0.05$)*. In der Stichprobe werden* 25 *defekte Sicherungsautomaten entdeckt.*

Übungsaufgabe 4.11:

Ein Würfel wird 120*mal mit folgenden Häufigkeiten für die verschiedenen Augenzahlen geworfen:*

Augenzahl	1	2	3	4	5	6
Häufigkeit	20	22	17	18	19	24

Überprüfen Sie mit Hilfe des χ^2*-Tests, ob es sich um einen gefälschten Würfel handelt* ($\alpha = 0.1$)*.*

Übungsaufgabe 4.12:

Es besteht eine Vermutung (Nullhypothese), dass die Körpergröße von Studenten approximativ normalverteilt ist mit Mittelwert 170 *cm und Standardabweichung* 10 *cm. Eine bei* 500 *Studenten durchgeführte Messung der Körpergröße führte zu folgendem Ergebnis:*

Körpergröße (in cm)	< 140	$[140; 170)$	$[170; 190)$	$[190; 200)$	≥ 200
Anzahl der Studenten	2	200	265	32	1

Kann aus diesem Ergebnis die aufgestellte Nullhypothese bei einem Signifikanzniveau von $\alpha = 0.05$ *abgelehnt werden?*

Übungsaufgabe 4.13:

Fassadenverkleidungen aus 3 *verschiedenen Materialien* (A, B, C) *wurden auf ihre Lebensdauer untersucht. Prüfen Sie, ob bei einem Signifikanzniveau von* $\alpha = 0.1$ *ein Zusammenhang zwischen Material und Lebensdauer besteht.*

Lebensdauer	Verkleidung A	B	C	$\sum$
über 5 Jahre	10	20	20	50
unter 5 Jahre	10	10	30	50
$\sum$	20	30	50	100

Übungsaufgabe 4.14:

Um den laufenden Gerüchten entgegenzutreten, dass die Mitglieder einer bestimmten Partei es einfacher hätten, höhere öffentliche Ämter zu besetzen, hat der Bürgermeister einer deutschen Großstadt ein unabhängiges Institut mit einer Untersuchung beauftragt. Das Institut erstellt daraufhin eine Statistik über die Einstellung von insgesamt 150 *Kommunalbeamten in den letzten* 5 *Jahren:*

Posten	Parteizugehörigkeit: Partei A	andere Partei bzw. ohne Parteibuch	$\sum$
mittlere Posten	15	60	75
gehobene Posten	15	30	45
höchste Posten	19	10	29
Stadtdirektor	1	0	1
$\sum$	50	100	150

Kann aus den zur Verfügung stehenden Daten bei einem Signifikanzniveau von $\alpha = 0.01$ *auf eine Abhängigkeit zwischen Ämtervergabe und Parteizugehörigkeit geschlossen werden?*

Übungsaufgabe 4.15:

An 16 Testpersonen soll überprüft werden, welches von zwei Schlafmitteln (X und Y) wirksamer ist. Dazu ist bei den 16 Testpersonen nur festgehalten worden, mit welchem Medikament ein längerer Schlaf herbeigeführt worden ist. In der nachfolgenden Tabelle ist jeweils durch ein Kreuz vermerkt, welches Medikament wirksamer gewesen ist ($\alpha = 0.05$).

Testperson	1	2	3	4	5	6	7	8
Medikament X	x	x		x	x	x	x	
Medikament Y			x					x
Vorzeichen der Differenz	+	+	-	+	+	+	+	-

Testperson	9	10	11	12	13	14	15	16
Medikament X	x	x	x			x	x	x
Medikament Y				x	x			
Vorzeichen der Differenz	+	+	+	-	-	+	+	+

Untersuchen Sie mittels eines geeigneten Tests, ob die Schlafmittel unterschiedlich wirken.

Übungsaufgabe 4.16:

Ein Erdölkonzern möchte feststellen, ob ein neu entwickelter Kraftstoffzusatz zur Verschleißminderung in Ottomotoren den Benzinverbrauch beeinflusst.

Bei zehn Testfahrzeugen wurden folgende Werte beobachtet:

Verbrauch l/100 km	1	2	3	4	5
mit Zusatz (X)	8.3	10.2	8.1	7.4	10.8
ohne Zusatz (Y)	8.8	10.6	8.7	7.9	10.7

Verbrauch l/100 km	6	7	8	9	10
mit Zusatz (X)	12.5	12.2	9.8	10.2	8.4
ohne Zusatz (Y)	13.2	12.9	9.6	10.1	8.9

a) *Führen Sie für die Nullhypothese H_0 : „Der Zusatz beeinflusst den Benzinverbrauch nicht“ den Wilcoxon-Vorzeichen-Rang-Test durch* $(\alpha = 0.05)$.

b) *Führen Sie den Vorzeichentest für die Nullhypothese aus a) durch* $(\alpha = 0.05)$.

Übungsaufgabe 4.17:

Ein Schuhhändler möchte zwei seiner Angestellten bezüglich ihrer Anzahl verkaufter Paar Schuhe in den letzten 10 *Tagen miteinander vergleichen. Dazu verwendet er den Wilcoxon-Rangsummen-Test zum Signifikanzniveau* $\alpha = 0.1$.

	# Schuhpaare pro Tag									
X	14	13	18	17	19	25	21	11	8	14
Y	24	15	20	17	12	24	26	22	25	16

Lösung der Übungsaufgaben

Lösung der Aufgaben zu Kapitel 1

Lösung 1.1:

Es ist $\sigma^2 = \mathrm{Var}(X_i) = E(X_i^2) - [E(X_i)]^2 = E(X_i^2) - \mu^2$*, d.h.* $E(X_i^2) = \sigma^2 + \mu^2$*. Weiter ist*

$$\begin{aligned}
\mathrm{Var}(\overline{X}) &= \sigma_{\overline{X}}^2 = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}\ n\ \sigma^2 = \frac{\sigma^2}{n}.
\end{aligned}$$

Mit diesen Beziehungen ergibt sich:

$$\begin{aligned}
E(S^2) &= E\left(\frac{1}{n-1}\sum(X_i - \overline{X})^2\right) \\
&= E\left(\frac{1}{n-1}\sum X_i^2 - \frac{2}{n-1}\overline{X}\sum X_i + \frac{1}{n-1}\sum \overline{X}^2\right) \\
&= E\left(\frac{1}{n-1}\sum X_i^2 - \frac{2}{n-1}\overline{X}\cdot n\overline{X} + \frac{1}{n-1}n\overline{X}^2\right) \\
&= \frac{1}{n-1}\left[\sum E(X_i^2) - nE(\overline{X}^2)\right] \\
&= \frac{1}{n-1}\left[n(\sigma^2+\mu^2) - n(\sigma_{\overline{X}}^2 + \mu^2)\right] \\
&= \frac{1}{n-1}\left(n\sigma^2 + n\mu^2 - n\frac{\sigma^2}{n} - n\mu^2\right) \\
&= \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2
\end{aligned}$$

Lösung 1.2:

Betrachtet wird ein Ereignis A. Die Stichprobenvariable X_i für die

$$X_i = \begin{cases} 0 & \text{für „}\overline{A}\text{ tritt ein“} \\ 1 & \text{für „}A\text{ tritt ein“} \end{cases}$$

gilt, sei bernoulli verteilt mit dem Parameter π.

$$\begin{aligned} E(P) &= E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}(0 \cdot (1-\pi) + 1 \cdot \pi) \\ &= \frac{1}{n}\sum_{i=1}^{n} \pi = \frac{1}{n} \cdot n\pi = \pi \end{aligned}$$

Lösung 1.3:

Die Erwartungstreue von $\overline{X}$ wurde bereits nachgewiesen. Zu zeigen bleibt, dass die Schätzfunktion $\overline{X}$ minimale Varianz besitzt, d.h. zu zeigen ist, dass $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \hat{\mu})^2$ minimal wird für $\overline{X}$.

$$\frac{dS^2}{d\hat{\mu}} = \frac{1}{n}\sum_{i=1}^{n} -2(X_i - \hat{\mu}) = \frac{-2}{n}\sum_{i=1}^{n} X_i + 2\hat{\mu} \stackrel{!}{=} 0.$$

Daraus folgt:

$$\hat{\mu} = \frac{\sum X_i}{n} = \overline{X}.$$

Da außerdem gilt $\frac{d^2S^2}{d\hat{\mu}^2} = 2 > 0$, liegt tatsächlich ein Minimum vor.

Lösung 1.4:

Gezeigt werden muss lediglich die Erwartungstreue der Schätzfunktion S^{*2}. Die Erwartungstreue von S^2 wurde bereits in Aufgabe 1.1 gezeigt.

$$\begin{aligned}
E(S^{*2}) &= \frac{1}{n}\sum_{i=1}^{n} E(X_i-\mu)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(E(X_i^2)-2\mu E(X_i)+\mu^2) \\
&= \frac{1}{n}(n\sigma^2+n\mu^2-2n\mu^2+n\mu^2) \\
&= \frac{1}{n}n\sigma^2=\sigma^2
\end{aligned}$$

Aufgrund der Normalverteilungsannahme gilt, dass die Statistik $\frac{(n-1)S^2}{\sigma^2}$ χ^2-verteilt ist mit $(n-1)$ Freiheitsgraden. Die Statistik $\frac{nS^{*2}}{\sigma^2}$ ist dagegen χ^2-verteilt mit n Freiheitsgraden. Somit gilt

$$\begin{aligned}
\operatorname{Var}\left(\frac{(n-1)S^2}{\sigma^2}\right) &= 2(n-1) \\
\Leftrightarrow \quad \frac{(n-1)^2}{\sigma^4}\operatorname{Var}(S^2) &= 2(n-1) \\
\Leftrightarrow \quad \operatorname{Var}(S^2) &= \frac{2\sigma^4}{(n-1)}
\end{aligned}$$

$$\begin{aligned}
\operatorname{Var}\left(\frac{nS^{*2}}{\sigma^2}\right) &= 2n \\
\Leftrightarrow \quad \frac{n^2}{\sigma^4}\operatorname{Var}(S^{*2}) &= 2n \\
\Leftrightarrow \quad \operatorname{Var}(S^{*2}) &= \frac{2\sigma^4}{n}
\end{aligned}$$

Der Schätzer S^{*2} ist aufgrund der geringeren Varianz effizienter als S^2.

Lösung 1.5:

Liegt eine erwartungstreue Schätzfunktion vor, d.h. der Bias nimmt den Wert Null an, so entspricht die mittlere quadratische Abweichung der Varianz. Das Kriterium $\lim_{n\to\infty} \mathrm{MSE}(T) = 0$ *vereinfacht sich zu dem Kriterium* $\lim_{n\to\infty} \mathrm{Var}(T) = 0$.

a) *Die Erwartungstreue von* $\overline{X}$ *wurde bereits gezeigt. Für die Varianz* $\mathrm{Var}(\overline{X}) = \sigma^2/n$ gilt

$$\lim_{n\to\infty} \frac{\sigma^2}{n} = 0.$$

b) *Auch für* P *wurde bereits die Erwartungstreue nachgewiesen. Für* $\mathrm{Var}(P) = \pi(1-\pi)/n$ *gilt*

$$\lim_{n\to\infty} \frac{\pi(1-\pi)}{n} = 0.$$

c) *Ebenso gilt für* S^2 *aufgrund der Erwartungstreue, dass der Bias dem Wert Null entspricht. Die Statistik* $\frac{(n-1)S^2}{\sigma^2}$ *ist* χ^2*-verteilt mit* $(n-1)$ *Freiheitsgraden, so dass* $\mathrm{Var}(S^2) = \frac{2\sigma^4}{(n-1)}$ *entspricht. Daher gilt*

$$\lim_{n\to\infty} \frac{2\sigma^4}{n-1} = 0.$$

Lösung der Aufgaben zu Kapitel 2

Lösung 2.1:

$n = 225 \quad \sigma = 150 \quad \sigma_{\overline{X}} = \frac{150}{\sqrt{225}} = 10 \quad z_{0.975} = 1.96$

$$\begin{array}{ll} \mu_u = \overline{x} - z_{1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{\overline{X}} & \mu_o = \overline{x} + z_{1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{\overline{X}} \\ \mu_u = 710 - 1.96 \cdot 10 = 710 - 19.6 & \mu_o = 710 + 1.96 \cdot 10 = 710 + 19.6 \\ \mu_u = 690.4 & \mu_o = 729.6 \end{array}$$

Lösung 2.2:

a) $N = 1000$, $n = 36$, $\overline{x} = 23$, $s = 0.6$, σ *unbekannt, Stichprobe ohne Zurücklegen.*

$\frac{n}{N} = 0.037 < 0.05 \quad \hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{0.6}{\sqrt{36}} = 0.1 \quad n - 1 = 35 \geq 30$

$$\begin{array}{lll} \mu_u = \overline{x} - z_{1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{\overline{X}} & \mu_o = \overline{x} + z_{1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{\overline{X}} & \\ \mu_u = 22.835 & \mu_o = 23.165 & z_{0.95} = 1.65 \\ (\mu_u = 22.804 & \mu_o = 23.196 & z_{0.975} = 1.96) \end{array}$$

b) $N = 250$, $n = 25$, $\overline{x} = 23$, $s = 0.6$, σ *unbekannt, Stichprobe ohne Zurücklegen.*

$\frac{n}{N} = \frac{25}{250} = 0.1 \geq 0.05 \quad \hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N}} = \frac{0.6}{5} \cdot \sqrt{\frac{225}{250}} = 0.1138$
$n - 1 = 24 < 30$

$$\begin{array}{lll} \mu_u = \overline{x} - t_{1-\frac{\alpha}{2}}(n-1) \cdot \hat{\sigma}_{\overline{X}} & \mu_o = \overline{x} + t_{1-\frac{\alpha}{2}}(n-1) \cdot \hat{\sigma}_{\overline{X}} & \\ \mu_u = 22.805 & \mu_o = 23.195 & t_{0.95}\,(24) = 1.711 \\ (\mu_u = 22.765 & \mu_o = 23.235 & t_{0.975}(24) = 2.064) \end{array}$$

Lösung 2.3:

a) $N = 1000,\ n = 64,\ \overline{x} = 30,\ s = 1.6,\ \sigma$ *unbekannt, Stichprobe mit Zurücklegen.*

$\hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{1.6}{8} = 0.2 \qquad n - 1 = 63 \geq 30$

$$\begin{array}{lll} \mu_u = \overline{x} - z_{0.975} \cdot \hat{\sigma}_{\overline{X}} & \mu_o = \overline{x} + z_{0.975} \cdot \hat{\sigma}_{\overline{X}} & \\ \mu_u = 29.608 & \mu_o = 30.392 & z_{0.975} = 1.96 \end{array}$$

b) $N = 1000,\ n = 36,\ \overline{x} = 30,\ s = 1.6,\ \sigma$ *unbekannt, Stichprobe mit Zurücklegen.*

$\hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{1.6}{6} \approx 0.267 \qquad n - 1 = 35 \geq 30$

$$\mu_u = 29.477 \qquad \mu_o = 30.523 \qquad z_{0.975} = 1.96$$

c) $N = 1500,\ n = 49,\ \overline{x} = 28,$ *Stichprobe mit Zurücklegen.*
Als Schätzung des Parameters σ wird nun der Wert aus früheren Messungen benutzt.

$\sigma = 2.1$ *(bekannt)* $\qquad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2.1}{7} = 0.3$

$$\begin{array}{lll} \mu_u = \overline{x} - z_{0.95} \cdot \sigma_{\overline{X}} & \mu_o = \overline{x} + z_{0.95} \cdot \sigma_{\overline{X}} & \\ \mu_u = 27.505 & \mu_o = 28.495 & z_{0.95} = 1.65 \end{array}$$

Lösung 2.4:

$N =$„∞“, $n = 200,\ x = 80,$ *Stichprobe mit Zurücklegen.*
$p = \frac{x}{n} = \frac{80}{200} = \frac{2}{5} = 0.4 \quad (1-p) = 0.6 \quad np = 80 \geq 5, n(1-p) = 120 \geq 5$
$n = 200 \geq 30 \quad \alpha = 0.1$

$$\begin{array}{lll} \pi_u = p - z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} & \pi_o = p + z_{1-\frac{\alpha}{2}}\sqrt{\frac{p(1-p)}{n}} & \\ \pi_u = 0.34 & \pi_o = 0.46 & z_{0.95} = 1.65 \end{array}$$

Lösung 2.5:

$n = 150 \geq 30 \quad x = 30 \quad p = 0.2 \quad \alpha = 0.10$
$np = 30 \geq 5, n(1-p) = 120 \geq 5$

$$\pi_u = p - z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{p(1-p)}{n}} \qquad \pi_o = p + z_{1-\frac{\alpha}{2}}\sqrt{\tfrac{p(1-p)}{n}}$$
$$\pi_u = 0.15 \qquad \pi_o = 0.25 \qquad z_{0.95} = 1.65$$

Lösung 2.6:

Mit $n = 50$, $s^2 = 191844$, $\alpha = 0.1$ *und den Quantilen* $\chi^2_{0.05}(49) = 33.93$ *und* $\chi^2_{0.95}(49) = 66.339$ *ergeben sich die Grenzen:*

$$\sigma_o^2 = \frac{49 \cdot 191844}{33.93} = 277051.459$$
$$\sigma_u^2 = \frac{49 \cdot 191844}{66.339} = 141701.804$$

Lösung der Aufgaben zu Kapitel 3

Lösung 3.1:

1. a) *Quantitatives Merkmal.*

   b) *Endliche Grundgesamtheit mit* $N = 3000$*.*

   c) *Es kann für* X *approximativ Normalverteilung angenommen werden.*

2. a) *qualitatives Merkmal (dichotome Grundgesamtheit).*

   b) *Die Grundgesamtheit besteht aus allen theoretisch möglichen Versuchen, ist also unendlich.*

   c) *Bernoulliverteilung mit der Wahrscheinlichkeit für „13" bzw. für „keine* 13*". Das führt bei* n*-maliger Wiederholung zur Binomialverteilung.*

Lösung 3.2:

a) *Falls die neue Methode nicht nachweisbar schlechter ist, möchte Lehrer A diese anwenden. Die Nullhypothese lautet dann „Die neue Methode ist nicht schlechter" oder formal:* $H_0: \ \mu \geq \mu_0 = 25$*. Sie wird erst dann abgelehnt, wenn in einer Erprobung die mittlere Anzahl* $\overline{X}$ *der pro Woche erlernten Vokabeln den Wert* 25 *signifikant unterschreitet, falls also* $\overline{x}$ *kleiner als* c_u *ausfällt* ($\overline{x} < c_u$)*. Der Lehrer ist sich dann (bei einer Irrtumswahrscheinlichkeit von* α*) sicher, dass die neue Methode schlechter ist, und wird sie nicht verwenden.*

b) *Lehrer B möchte abgesichert haben, dass die neue Methode die alte deutlich übertrifft. Er legt daher die Nullhypothese* $H_0: \ \mu \leq \mu_0 = 25$ *zugrunde oder* H_0 *: „Die neue Methode ist nicht besser als die alte".* H_0 *wird erst dann abgelehnt, wenn* $\overline{x}$ *oberhalb des oberen kritischen Wertes liegt* ($\overline{x} > c_o$)*. Führt der Test zu einer Ablehnung der Nullhypothese, dann wird er annehmen, dass die neue Methode die alte bei einer Irrtumswahrscheinlichkeit von höchstens* α *statistisch nachweisbar übertrifft.*

Lösung 3.3:

a) $H_0: \ \overline{IQ}_A \leq \overline{IQ}_B$ vs. $H_1: \ \overline{IQ}_A > \overline{IQ}_B$

Falls H_0 *abgelehnt werden kann, wäre statistisch gesichert, dass* $\overline{IQ}_A > \overline{IQ}_B$ *gilt.*

b) $H_0: \ \mu = \mu_0 = 78.65$ vs. $H_1: \ \mu \neq \mu_0$

Die Produktion wird im Allgemeinen nur dann gestoppt, wenn „sicher" ist, dass die Norm nicht mehr eingehalten wird. Dabei ist sowohl eine Abweichung nach oben als auch nach unten von Interesse.

c) $H_0: \ \mu \leq \mu_0 = 1.2 \ l/h$ vs. $H_1: \ \mu > \mu_0$

Die Alternativhypothese muss enthalten, was nachgewiesen werden soll.

Lösung 3.4:

1) $H_0: \ \mu \leq \mu_0 = 150 \ g$ vs. $H_1: \ \mu > \mu_0$

Da A „sicher" sein will, dass die Äpfel im Mittel schwerer sind als 150 *g, muss die Alternativhypothese diese Forderung enthalten.*

2) $H_0: \ \mu \geq \mu_0 = 150 \ g$ vs. $H_1: \ \mu < \mu_0$

B möchte nur „gewarnt" werden, wenn statistisch gesichert ist, dass die Äpfel im Mittel unter 150 *g wiegen. Entsprechend muss die Alternativhypothese formuliert werden.*

Lösung 3.5:

B) *Ist richtig.*

A) *Wäre nur richtig, wenn* H_0 *verworfen wird.*

C) *Die Wahrscheinlichkeit, dass doch eine Verringerung vorliegt, ist unbekannt. Die Wahrscheinlichkeit des Fehlers 2. Art beträgt höchstens* $1 - \alpha = 0.95$.

D) H_0 *wurde nicht abgelehnt, ist deshalb aber auch nicht bewiesen, daher ist Formulierung B) richtig.*

Lösung 3.6:

a) $H_0: \ \mu \geq \mu_0 = 45 \ g \qquad \text{vs.} \qquad H_1: \ \mu < \mu_0$

b) *Es ergibt sich für*

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{5} = 0.4$$

und $z_{1-\alpha} = 1.65$ *der untere kritische Wert*

$$c_u = \mu_0 - z_{1-\alpha}\sigma_{\overline{X}} = 45 - 1.65 \cdot 0.4 = 44.34.$$

Ablehnungsbereich: $\{\overline{x} \mid \overline{x} < 44.34\}$

c) *Mit* $\overline{x} = 44$ *wird* H_0 *abgelehnt. Das mittlere Gewicht liegt also mit „Sicherheit" unter 45 g.*

d) *Das „Risiko" beträgt definitionsgemäß* $\alpha = 0.05$ *(wurde in b) vorgegeben).*

e) *Gesucht ist die Wahrscheinlichkeit* $p = \boldsymbol{P}(X \leq 44)$.

$$\begin{aligned} \boldsymbol{P}(X \leq 44) &= \boldsymbol{P}(Z \leq \tfrac{44-45}{0.4}) = \boldsymbol{P}(Z \leq -2.5) \\ &= 1 - \boldsymbol{P}(Z \leq 2.5) = 1 - 0.9938 = 0.0062 \end{aligned}$$

Da $p < \alpha = 0.05$ *gilt, wird* H_0 *abgelehnt.*

Lösung der Aufgaben zu Kapitel 4

Lösung 4.1:

$$H_0: \ \mu \geq \mu_0 = 100000 \ (m/h) \qquad \text{vs.} \qquad H_1: \ \mu < \mu_0$$

Die Standardabweichung σ ist bekannt und $\overline{X}$ ist wegen $n > 30$ approximativ normalverteilt. Wenn von einer Stichprobe mit Zurücklegen ausgegangen wird, kann als Prüfgröße $Z = \frac{\overline{X}-\mu_0}{\sigma_{\overline{X}}}$ verwendet werden, welche approximativ standardnormalverteilt ist ($n = 36, \alpha = 0.05$). Mit $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{30000}{6} = 5000$ ergibt sich:

$$z = \frac{90000 - 100000}{5000} = -2.$$

Wegen $-2 < -z_{0.95} = -1.65$ ist die Behauptung des Automobilherstellers A statistisch widerlegt.

Lösung 4.2:

$$H_0: \ \mu = \mu_0 = 500 \ (ml) \qquad \text{vs.} \qquad H_1: \ \mu \neq \mu_0$$

Die Standardabweichung σ ist unbekannt und X ist approximativ normalverteilt. Es liegt eine Stichprobe ohne Zurücklegen vor ($n \leq 30$), doch aufgrund der laufenden Produktion gilt $\frac{n}{N} < 0.05$. Als Teststatistik kann $T = \frac{\overline{X}-\mu_0}{\hat{\sigma}_{\overline{X}}} = \sqrt{n}\frac{\overline{X}-\mu_0}{S}$ verwendet werden, welche approximativ t-verteilt ist mit $n-1$ Freiheitsgraden ($n = 25, \alpha = 0.05$).

$$t \ = \ \sqrt{25}\,\frac{490 - 500}{5} = -2$$

Die Nullhypothese wird nicht verworfen, da $t_{0.025}(24) = -2.064 \leq -2 \leq t_{0.975}(24) = 2.064$ gilt.

Dass die Nullhypothese zutrifft ist jedoch nicht statistisch gesichert.

Lösung 4.3:

$$H_0: \ \mu = \mu_0 = 38 \qquad \text{vs.} \qquad H_1: \ \mu \neq \mu_0$$

Die Standardabweichung σ ist bekannt und wegen $n > 30$ ist $\overline{X}$ approximativ $N(38,\ \sigma^2_{\overline{X}})$-verteilt. Es liegt eine Stichprobe ohne Zurücklegen vor, aber da es sich um eine Großstadt handelt gilt $\frac{n}{N} < 0.05$. Es wird die approximativ standardnormalverteilte Prüfgröße $Z = \frac{\overline{X}-\mu_0}{\sigma_{\overline{X}}}$ verwendet mit $\sigma_{\overline{X}} = \frac{8}{\sqrt{100}} = 0.8$ $(n = 100, \alpha = 0.05)$.

$$z \ = \ \frac{39 - 38}{0.8} = 1.25$$

Es ist $z_{0.025} = -z_{0.975} = -1.96 \leq 1.25 = z \leq z_{0.975} = 1.96$, d.h. die Realisation z der Prüfgröße Z liegt zwischen den gegebenen kritischen Werte $c_u = -z_{0.975}$ und $c_o = z_{0.975}$. Die Nullhypothese kann somit nicht abgelehnt werden.

Lösung 4.4:

$$H_0: \ \sigma^2 \leq \sigma_0^2 = 0.01 \qquad \text{vs.} \qquad H_1: \ \sigma^2 > \sigma_0^2 = 0.01$$

Die Prüfgröße $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ ist χ^2-verteilt mit $n-1 = 31$ Freiheitsgraden. Mit $s^2 = 0.012$ ergibt sich die Realisation

$$\chi^2 = \frac{31 \cdot 0.012}{0.01} = 37.2.$$

Der obere kritische Wert c_o entspricht dem Quantil $\chi^2_{1-\alpha}(31)$ mit $\alpha = 0.05$. Da $\chi^2 = 37.2 \leq \chi^2_{0.95}(31) = 44.99$, kann H_0 nicht abgelehnt werden. Es besteht kein hinreichender Grund, an der Aussage des Herstellers zu zweifeln.

Lösung 4.5:

$$H_0: \ \sigma^2 \leq \sigma_0^2 = 100 \quad \text{vs.} \quad H_1: \ \sigma^2 > \sigma_0^2$$

Die Prüfgröße $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ *ist* χ^2*-verteilt mit* $n-1 = 49$ *Freiheitsgraden. Mit* $s^2 = 125$ *ergibt sich die Realisation*

$$\chi^2 = \frac{125 \cdot 49}{100} = 61.25.$$

Als kritischer Wert ergibt sich zum Signifikanzniveau $\alpha = 0.1$ *das Quantil* $c_o = \chi^2_{0.9}(49) = 62.038$, *so dass die Nullhypothese nicht abgelehnt werden kann* $(61.25 < 62.038)$.

Lösung 4.6:

$$H_0: \ \mu_X - \mu_Y = 0 \quad \text{vs.} \quad H_1: \ \mu_X - \mu_Y \neq 0$$

Unter den getroffenen Annahmen (Normalverteilungsannahme, Varianzhomogenität, σ_X *und* σ_Y *unbekannt,* $n, m < 30$, $\alpha = 0.05$) *wird zur Berechnung der Prüfgröße* $T = \frac{\overline{X}_X - \overline{X}_Y}{S_D}$ *die Varianz*

$$s_d^2 = \left(\frac{1}{n} + \frac{1}{m}\right) \frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2} = \frac{1}{6} \cdot \frac{14 \cdot 0.6 + 9 \cdot 0.4}{23} = \frac{2}{23}$$

benötigt. Da sich der Wert der Prüfgröße T *zu*

$$t = \frac{\overline{x}_x - \overline{x}_y}{s_d} = \frac{3.15 - 2.55}{\sqrt{\frac{2}{23}}} = 2.035$$

ergibt mit $-2.069 = t_{0.025}(23) \leq t \leq t_{0.975}(23) = 2.069$, *kann die Nullhypothese nicht abgelehnt werden.*

Lösung 4.7:

$$H_0: \frac{\sigma_Y^2}{\sigma_X^2} = 1 \quad \text{vs.} \quad H_1: \frac{\sigma_Y^2}{\sigma_X^2} \neq 1$$

.
Für die Prüfgröße $F = \frac{S_Y^2}{S_X^2}$ *ergibt sich der Wert* $\frac{12.25}{28.09} = 0.436$. *Die kritischen Werte lauten* ($\alpha = 0.1$)*:*

$$\begin{aligned} c_o &= F_{0.95}(19, 14) = 2.4 \\ c_u &= F_{0.5}(19, 14) = \frac{1}{F_{0.95}(14, 19)} = \frac{1}{2.26} = 0.442 \end{aligned}$$

Da $0.436 < F_{0.5}(19, 14) = 0.442$ *gilt, wird* H_0 *abgelehnt.*

Lösung 4.8:

$$H_0: \pi \leq \pi_0 = 0.15 \quad \text{vs.} \quad H_1: \pi > \pi_0$$

Betrachtet wird die Zufallsvariable X *(Anzahl der Kreditnehmer mit Rückzahlungsschwierigkeiten). In der Stichprobe befinden sich* $x = 7$ *Kreditnehmer mit Rückzahlungsschierigkeiten.*

Es gilt $n\pi_0 = 1.05 < 5, n(1 - \pi_0) = 5.95 < 5$.

Eine Approximation durch die Normalverteilung ist also nicht möglich.

Die Zufallsvariable X *ist* $B(20, 0.15)$*-verteilt und* c_o *wird zum Signifikanzniveau* $\alpha = 0.05$ *mittels der Bedingungen*

$$F_X(c_0) \geq 0.95 \quad \textit{und} \quad F_X(c_0 - 1) < 0.95.$$

bestimmt. Aus der Tabelle der Binomialverteilung ergibt sich mit $F_X(6) = 0.9781$ *und* $F_X(5) = 0.9327$ *der kritische Wert* $c_o = 6$. *Die Hypothese* H_0 *ist wegen* $x = 7 > c_o$ *abzulehnen* ($\alpha^* = 0.0219$).

Lösung 4.9:

$$H_0: \ \pi \geq \pi_0 = 0.2 \qquad \text{vs.} \qquad H_1: \ \pi < \pi_0$$

Mit $n = 100$ *gilt* $n\pi_0 = 20 \geq 5, n(1-\pi_0) = 80 \geq 5$. *Es handelt sich um eine Stichprobe ohne Zurücklegen.*
Wegen $\frac{n}{N} < 0.05$ (*Großstadt*) *und* $\sigma_P = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = 0.04$ *ist*

$$Z = \frac{P - \pi_0}{\sigma_P} = \frac{P - 0.2}{0.04}$$

approximativ standardnormalverteilt. Mit $p = \frac{15}{100} = 0.15$ *ergibt sich für die Prüfgröße der Wert* $z = -1.25$. *Die Nullhypothese wird zum Signifikanzniveau* $\alpha = 0.05$ *nicht abgelehnt, da* $z = -1.25 \geq -1.65 = z_{0.05}$ *gilt.*

Lösung 4.10:

$$H_0: \ \pi \leq \pi_0 = 0.1 \qquad \text{vs.} \qquad H_1: \ \pi > \pi_0 = 0.1$$

Mit $n\pi_0 = 144 \cdot 0.1 = 14.4 \geq 5, n(1-\pi_0) = 129.6 \geq 5$ *und der Annahme einer unendlichen Grundgesamtheit, kann die Normalverteilung als Näherung verwendet werden. Es ergibt sich für den einseitigen Test zum Signifikanzniveau* $\alpha = 0.05$ *der kritische Wert* $z_{0.95} = 1.65$. *Es ist* $\sigma_P = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = 0.025$. *Die Nullhypothese wird abgelehnt, falls*

$$Z = \frac{P - 0.1}{0.025} > 1.65$$

gilt. Ein Ausschuss von 25 Stück entspricht einem Anteil von $p = \frac{25}{144} = 0.174$, *so dass sich* $z = 2.96$ *ergibt.*

Wegen $2.96 > 1.65 = z_{0.95}$ *muss die Behauptung des Herstellers als statistisch widerlegt gelten.*

Lösung 4.11:

$$H_0: \quad \boldsymbol{P}(X = x_j) = \tfrac{1}{6} \text{ für alle } j = 1, ..., 6$$
$$H_1: \quad \boldsymbol{P}(X = x_j) \neq \tfrac{1}{6} \text{ für mindestens ein } j = 1, ..., 6$$

Augenzahl x_j	beobachtete Häufigkeit h_j	erwartete Häufigkeit $n\pi_j$	$(h_j - n\pi_j)^2$	$\frac{(h_j - n\pi_j)^2}{n\pi_j}$
1	20	20	0	0
2	22	20	4	$\frac{4}{20}$
3	17	20	9	$\frac{9}{20}$
4	18	20	4	$\frac{4}{20}$
5	19	20	1	$\frac{1}{20}$
6	24	20	16	$\frac{16}{20}$
				$\chi^2 = \frac{34}{20} = 1.7$

Nach dem χ^2-Anpassungstest mit der Prüfgröße $\chi^2 = \sum_{j=1}^{m} \frac{(h_j - n\pi_j)^2}{n\pi_j}$ ergibt sich zum Signifikanzniveau $\alpha = 0.1$ das Quantil $\chi^2_{1-\alpha}(m-1) = \chi^2_{0.9}(5) = 9.236$. Mit $9.236 \geq 1.7 = \chi^2$ kann die Hypothese „Der Würfel ist ideal" nicht abgelehnt werden.

Lösung 4.12:

Als Prüfgröße für den χ^2-Anpassungstest wird $\chi^2 = \sum_{j=1}^{m} \frac{(h_j - n\pi_j)^2}{n\pi_j}$ verwendet. Die erwarteten absoluten Häufigkeiten werden hier über die Wahrscheinlichkeiten der Normalverteilung π_j für die entsprechenden Intervalle berechnet. Für die erste und die letzte Klasse gilt

$$\pi = \pi_1 = \boldsymbol{P}(X < 140) = \pi_5 = \boldsymbol{P}(X \geq 200) = \boldsymbol{P}(Z < -3) = 0.00135$$

und somit $n\pi = 0.675 < 5$. Die ersten beiden und die letzten beiden Klassen werden daher zusammengefasst, um die Bedingung $n\pi_j \geq 1, n\pi_j \geq 5$ für mindestens 80% der Klassen zu erfüllen.

$$H_0: \quad \boldsymbol{P}(X = x_j) = \pi_j \text{ für alle } j = 1, 2, 3$$
$$H_1: \quad \boldsymbol{P}(X = x_j) \neq \pi_j \text{ für mindestens ein } j = 1, 2, 3$$

Für die zusammengefassten Klassen gilt

$$\begin{aligned}
\pi_1 &= \boldsymbol{P}(X < 170) = \boldsymbol{P}\big(Z < \tfrac{170-\mu}{\sigma}\big) = \boldsymbol{P}(Z < 0) = 0.5,\\
\pi_2 &= \boldsymbol{P}(170 \le X < 190) = \boldsymbol{P}(0 \le Z < 2) = 0.4772,\\
\pi_3 &= \boldsymbol{P}(190 \le X) = \boldsymbol{P}(2 \le Z) = 0.0228.
\end{aligned}$$

j	Körpergröße X in Klasse j	π_j	$n\pi_j$	h_j	$h_j - n\pi_j$	$(h_j - n\pi_j)^2$	$\frac{(h_j - n\pi_j)^2}{n\pi_j}$
1	$x < 170$	0.5	250	202	-48	2304	9.216
2	$170 \le x < 190$	0.4772	238.6	265	26.4	696.96	2.921
3	$190 \le x$	0.0228	11.4	33	21.6	466.56	40.926
		1.0000	500	500	0		$53.063 = \chi^2$

Zum Signifikanzniveau $\alpha = 0.05$ *ergibt sich das Quantil* $\chi^2_{1-\alpha}(m-1) = \chi^2_{0.95}(2) = 5.99$. *Mit* $\chi^2_{0.95}(2) < 53.063$ *wird die Nullhypothese abgelehnt.*

Lösung 4.13:

$$\begin{aligned}
H_0: &\quad \pi_{jk} = \pi_{j.} \cdot \pi_{.k} \text{ für alle } j = 1,2,\ k = 1,2,3\\
H_1: &\quad \pi_{jk} \ne \pi_{j.} \cdot \pi_{.k} \text{ für mindestens ein } j,k
\end{aligned}$$

Für den χ^2*-Unabhängigkeitstest wird* $\chi^2 = \sum_{j=1}^{m}\sum_{k=1}^{r}\frac{(h_{jk}-n\hat{\pi}_{jk})^2}{n\hat{\pi}_{jk}}$ *als Prüfgröße verwendet mit* $n\hat{\pi}_{jk} = \frac{h_{j.}h_{.k}}{n}$ $(n = 100)$.

	h_{j1}	$n\hat{\pi}_{j1}$	h_{j2}	$n\hat{\pi}_{j2}$	h_{j3}	$n\hat{\pi}_{j3}$
h_{1k}	10	*10*	20	*15*	20	*25*
h_{2k}	10	*10*	10	*15*	30	*25*

$$\chi^2 = \frac{25}{15} + \frac{25}{25} + \frac{25}{15} + \frac{25}{25} = 5\frac{1}{3}.$$

Zum Signifikanzniveau $\alpha = 0.1$ *ergibt sich der obere kritische Wert zu* $c_o = \chi^2_{1-\alpha}((m-1)(r-1)) = \chi^2_{0.9}(2) = 4.605$. *Mit* $\chi^2 = 5.333 > 4.605$ *kann die Nullhypothese abgelehnt und somit auf eine Abhängigkeit geschlossen werden.*

Lösung 4.14:

Für den χ^2-Unabhängigkeitstest wird $\chi^2 = \sum_{j=1}^{m}\sum_{k=1}^{r}\frac{(h_{jk}-n\hat{\pi}_{jk})^2}{n\hat{\pi}_{jk}}$ *als Prüfgröße verwendet mit* $n\hat{\pi}_{jk} = \frac{h_{j.}h_{.k}}{n}$ $(n = 150)$*. Die 3. und 4. Zeile werden zusammengefasst, da sonst die Bedingung* $n\hat{\pi}_{jk} \geq 1, n\hat{\pi}_{jk} \geq 5$ *für mindestens* 80% *der Klassen nicht erfüllt ist* $(n\hat{\pi}_{41} = 0.333, n\hat{\pi}_{42} = 0.667)$*. Es ergeben sich somit* $(3-1)(2-1) = 2$ *Freiheitsgrade.*

$$\begin{aligned} H_0: &\quad \pi_{jk} = \pi_{j.}\cdot\pi_{.k} \text{ für alle } j = 1,2,3,\ k = 1,2 \\ H_1: &\quad \pi_{jk} \neq \pi_{j.}\cdot\pi_{.k} \text{ für mindestens ein } j,k \end{aligned}$$

	h_{j1}	$n\hat{\pi}_{j1}$	h_{j2}	$n\hat{\pi}_{j2}$
h_{1k}	15	*25*	60	*50*
h_{2k}	15	*15*	30	*30*
h_{3k}	20	*10*	10	*20*

$$\chi^2 = \frac{100}{25}+\frac{100}{50}+\frac{100}{10}+\frac{100}{20} = 4+2+10+5 = 21$$

Zum Signifikanzniveau $\alpha = 0.01$ *ergibt sich der obere kritische Wert* c_o *zu* $\chi^2_{1-\alpha}((m-1)(r-1)) = \chi^2_{0.99}(2) = 9.21$ *Wegen* $\chi^2 > 9.21$ *wird die Nullhypothese (Unabhängigkeit der Merkmale „Parteizugehörigkeit" und „Posten") abgelehnt.*

Lösung 4.15:

$$H_0: \ D_{med} = 0 \ \text{ vs. } \ H_1: \ D_{med} \neq 0 \ mit\ D_i = X_i - Y_i$$

Als Prüfgröße des Vorzeichentests (Signifikanzniveau $\alpha = 0.05$*) wird*

$$D_n = \textit{Anzahl der positiven Differenzen}$$

verwendet, welche an dieser Stelle $B(16,\ 0.5)$*-verteilt ist. Aus der Tabelle der Binomialverteilung ergibt sich mit* $F_X(12) = 0.9894 \geq 0.975$ *und* $F_X(11) = 0.9616 < 0.975$ *der kritische Wert* $c_o = 12$ *und mit* $F_X(3) = 0.0106 \leq 0.025$ *und* $F_X(4) = 0.0384 > 0.025$ *der kritische Wert* $c_u = 4$*. Für* $d_n = 12$ *gilt* $c_u \leq d_n \leq c_o$*, so dass die Nullhypothese nicht abgelehnt werden kann.*

Lösung 4.16:

$$H_0: \ D_{med} = 0 \ \text{ vs. } \ H_1: \ D_{med} \neq 0 \ mit \ D_i = X_i - Y_i$$

Alternativ sollen jeweils zum Signifikanzniveau $\alpha = 0.05$ der Wilcoxon-Vorzeichen-Rang-Test a) und der Vorzeichentest b) durchgeführt werden.

i	1	2	3	4	5	6	7	8	9	10	$\sum$
x_i	8.3	10.2	8.1	7.4	10.8	12.5	12.2	9.8	10.2	8.4	
y_i	8.8	10.6	8.7	7.9	10.7	13.2	12.9	9.6	10.1	8.9	
$x_i - y_i$	-0.5	-0.4	-0.6	-0.5	0.1	-0.7	-0.7	0.2	0.1	-0.5	
rg_i^+					1.5			3	1.5		6
rg_i^-	6	4	8	6		9.5	9.5			6	49

a) *Die Testgröße des Wilcoxon-Vorzeichen-Rang-Tests*

$$W^+ = \sum_{i=1}^{n} \text{rg}|D_i|Z_i \quad mit \quad Z_i = \begin{cases} 1 & \text{falls } D_i > \delta_0 \\ \\ 0 & \text{falls } D_i < \delta_0 \end{cases} \qquad i = 1, ..., n$$

nimmt den Wert 6 an. Für $n = 10$ und $\alpha = 0.05$ ergibt sich der untere kritische Wert zu $c_u = w_{\frac{\alpha}{2}}(n) = w_{0.025}(10) = 9$. Die Nullhypothese muss abgelehnt werden, denn es gilt $W^+ = 6 < 9 = c_u$.

b) *Als Testgröße des Vorzeichentests wird*

$$D_n = \textit{Anzahl der positiven Differenzen}$$

verwendet, welche hier $B(10,\ 0.5)$-verteilt ist. Aus der Tabelle der Binomialverteilung ergibt sich mit $F_X(8) = 0.9893 \geq 0.975$ und $F_X(7) = 0.9453 < 0.975$ der kritische Wert $c_o = 8$ und mit $F_X(1) = 0.0107 \leq 0.025$ und $F_X(2) = 0.0547 > 0.025$ der kritische Wert $c_u = 2$. Wegen $c_u = 2 \leq d_n = 3 \leq 8 = c_o$ kann H_0 nicht abgelehnt werden.

<u>*Lösung 4.17:*</u>

$$H_0: \; X_{med} = Y_{med} \quad \text{vs.} \quad H_1: \; X_{med} \neq Y_{med}$$

Als Prüfgröße des Wilcoxon-Rangsummen-Tests wird

$$W_{n,m} = \sum_{i=1}^{n} \mathrm{rg}(X_i) = \sum_{i=1}^{n+m} iV_i$$

$$V_i = \begin{cases} 1 & \text{falls die } i\text{-te Beobachtung} \\ & \text{eine } X\text{-Variable ist} \\ 0 & \text{sonst} \end{cases}$$

verwendet. Die Rangzahlen werden für die gepoolte Stichprobe vergeben, wobei Bindungen innerhalb derselben Stichprobe nicht von Interesse sind. Liegen Bindungen zwischen beiden Stichproben vor, werden Durchschnittsränge gebildet $(n + m = 20)$.

	# Schuhpaare pro Tag										$\sum$
X_i	14	13	18	17	19	25	21	11	8	14	
$\mathrm{rg}(X_i)$	5	4	11	9.5	12	18.5	14	2	1	6	**83**
Y_i	24	15	20	17	12	24	26	22	25	16	
$\mathrm{rg}(Y_i)$	16	7	13	9.5	3	17	20	15	18.5	8	127

Mit $\alpha = 0.1$ *ergeben sich die kritischen Werte zu* $w_{0.05}(10, 10) = 83$ *und* $w_{0.95}(10, 10) = 127$. *Da* $83 \leq W_{10,10} = 83 \leq 127$ *gilt, kann die Nullhypothese, beide Verkäufer weisen die gleichen Umsatzzahlen auf, soeben nicht abgelehnt werden.*