Stichprobenverteilungen, Schätz– und Testtheorie

Begleitende Unterlagen zur Übung *Induktive Statistik*

Michael Westermann

Universität Essen

Inhaltsverzeichnis

Vorwort Diese Seiten sind als begleitender Text zur Übung *Induktive Statistik* gedacht. Dabei geht es –wie der Titel schon verrät– um die Themen Schätz– und Testtheorie. Der Text soll keinesfalls den Besuch der Lehrveranstaltungen bzw. die Lektüre der einschlägigen Literatur ersetzen, sondern lediglich dazu dienen, den in den relevanten Sitzungen behandelten Stoff nochmals grob zusammenzufassen und die Zusammenhänge zu verdeutlichen.

1 Grundzüge der Stichprobentheorie

Gegenstand der induktiven Statistik ist der sogenannte *Repräsentativschluß*. Mithilfe der Daten, welche aus einer Stichprobe gewonnen werden, sollen Rückschlüsse auf die Grundgesamtheit, aus der diese Stichprobe stammt, gezogen werden. Diese Rückschlüsse können auf zwei Arten geschehen:

Schätzen: Bestimmung eines Näherungswertes eines unbekannten Parameters der Grundgesamtheit aufgrund der Beobachtungen in der Stichprobe (**ohne** Vorkenntnisse). Geschieht durch Punkt– oder Intervallschätzung.

Testen: Überprüfen von statistischen Aussagen (*Hypothesen*) über die Verteilung einer Zufallsvariable oder über einen unbekannten Parameter (**mit** Vorkenntnissen).

Die Voraussetzung für die Durchführung dieser beiden Verfahren ist das Vorliegen einer Stichprobe.

1.1 Stichproben und Stichprobenfunktionen

Bei der Ziehung einer Stichprobe, werden n Elemente aus der N Elemente umfassenden Grundgesamtheit entnommen. Wenn $n = N$ ist , spricht man nicht mehr von einer Stichprobe, sondern von einer Totalerhebung. Das Ziehen einer Stichprobe besteht darin, n Elemente $x_1, \ldots, x_n$ aus der Grundgesamtheit auszuwählen und an ihnen die Ausprägungen des Merkmals zu beobachten. Dabei heißt der Quotient n/N *Auswahlsatz*.

Die Art, wie die Stichproben entnommen werden, heißt *Auswahlverfahren*. Dabei ist es von besonderer Bedeutung, daß das Prinzip der Zufallsauswahl eingehalten wird. **Nur dann** können die Methoden der schließenden Statistik auf diese Stichproben angewendet werden.

Deﬁnition 1 (Stichprobe) *Eine durch Zufallsauswahl gewonnene, endliche Menge von Beobachtungswerten $x_1, \ldots, x_n$ aus einer endlichen oder unendlichen Grundgesamtheit heißt*

– ***uneingeschränkte Zufallsauswahl**, wenn jedes Element der Grundgesamtheit die Möglichkeit besitzt, in die Stichprobe zu gelangen,*
– ***uneingeschränkte, einfache Zufallsauswahl**, wenn die Wahrscheinlichkeit, in die Stichprobe zu gelangen, bei jedem Element der Grundgesamtheit gleich groß ist.* ❑

Bei der Entnahme der Elemente aus der Grundgesamtheit ist weiterhin zu berücksichtigen, ob diese *mit* oder *ohne* Zurücklegen geschieht (*zmz, zoz*).

Unabhängig von dieser Fragestellung bleibt folgendes festzuhalten: In der Grundgesamtheit besitzt das Merkmal X die (unbekannte) Verteilungsfunktion $F(X)$ mit den (unbekannten) Parametern (Momenten) $E(X) = \mu$ und $V(X) = \sigma^2$. Die Elemente der Stichprobe können als Realisationen von n Zufallsvariablen $X_1, \ldots, X_n$ angesehen werden. Für alle diese Zufallsvariablen gilt:

– Sie stammen aus der gleichen Grundgesamtheit für die die oben genannten Eigenschaften gelten. ($E(X) = \mu$ und $V(X) = \sigma^2$)
– Die Realisationen aller X_i sind unabhängig voneinander.

Alle Elemente der Stichprobe sind somit *unabhängig, identisch verteilt*, oder kurz *i.i.d.* (für independent, identically distributed).

Auf der Basis der Stichprobe soll der unbekannte Parameter θ einer Grundgesamtheit geschätzt werden.[1] Diese Schätzung wird mittels einer *Schätz–* oder *Stichprobenfunktion* vorgenommen. Dabei handelt es sich um eine Funktion

$$\hat{\theta} = g(X_1, \ldots, X_n),$$

[1] Im weiteren Verlauf werden bestimmte *Parameter* der Grundgesamtheit, wie Mittel– und Anteilswerte, oder die Varianz allgemein mit θ gekennzeichnet. Diese Parameter der Grundgesamtheit werden durch die einzelnen Realisationen der Grundgesamtheit bestimmt.

welche lediglich von den Realisationen der Stichprobenvariablen abhängt. Deswegen stellt $g(\cdot)$ selbst ebenfalls eine Zufallsvariable dar.

Werden die konkreten Stichprobenrealisationen $x_1, \dots x_n$ in diese Funktion eingesetzt geht die Stichprobenfunktion in einen konkreten *Schätzwert* über. Dieser Wert heißt dann *Punktschätzer*.

1.2 Verteilungen von Stichprobenfunktionen

Wie eben beschrieben, handelt es sich bei der Funktion $g(\cdot)$ um eine Zufallsvariable. Die Verteilung dieser Zufallsvariablen wird *Stichprobenverteilung* des Schätzers $\hat{\theta}$ genannt. Diese Stichprobenverteilung besitzt einen Erwartungswert und eine Varianz, welche *Stichprobenfehler* oder *Standardfehler* genannt wird.

Beispiel 1 (Vgl. von der Lippe (1998), Aufgabe 7.2.7) *Die Grundgesamtheit ist zweipunktverteilt mit* $P(x = 1) = 0,5$ *und* $P(x = 2) = 0,5$. *Wie lautet die Stichprobenverteilung*

- *des arithmetischen Mittels* $\bar{X} = \frac{1}{n}\sum X_i$
- *des geometrischen Mittels* $\bar{X}_G = \sqrt[n]{X_1 \cdot \dots \cdot X_n}$

bei Stichproben vom Umfang $n = 2$ *und* $n = 3$ *(Ziehen mit Zurücklegen)? Was fällt bei der Betrachtung der Stichprobenverteilung von* $\bar{X}_G$ *auf?*

Es existieren insgesamt 2 Elementarereignisse und das Experiment wird $n = 2$ *Mal durchgeführt. Der Stichprobenraum* Ω *besteht aus* $2^2 = 4$ *Elementen, nämlich* $\Omega = \{(1,1);(1,2);(2,1);(2,2)\}$. *Das* ***arithmetische Mittel*** *kann bei diesen möglichen Stichproben die Werte 1, 1,5 und 2 annehmen. Für das* ***geometrische Mittel*** *erhält man die möglichen Realisationen 1,* $\sqrt{2}$*, und 2. Die entsprechenden Wahrscheinlichkeiten sind in der folgenden Tabelle dargestellt:*

Stichprobe	$\bar{x}$	$P(\bar{X} = \bar{x})$	$\bar{x}_G$	$P(\bar{X}_G = \bar{x}_G)$
(1,1)	*1*	$0,5 \cdot 0,5 = 0,25$	$\sqrt{1 \cdot 1} = 1$	*0,25*
(1,2)(2,1)	*1,5*	$2 \cdot 0,5 \cdot 0,5 = 0,5$	$\sqrt{1 \cdot 2} = \sqrt{2}$	*0,5*
(2,2)	*2*	$0,5 \cdot 0,5 = 0,25$	$\sqrt{2 \cdot 2} = 2$	*0,25*

Man erhält folgende Wahrscheinlichkeitsfunktion:

$$f(\bar{x}) = \begin{cases} 0,25 & \text{, für } \bar{x} = 1 \\ 0,5 & \text{, für } \bar{x} = 1,5 \\ 0,25 & \text{, für } \bar{x} = 2 \\ 0 & \text{sonst} \end{cases} \qquad f(\bar{x}_G) = \begin{cases} 0,25 & \text{, für } \bar{x}_G = 1 \\ 0,5 & \text{, für } \bar{x}_G = \sqrt{2} \\ 0,25 & \text{, für } \bar{x}_G = 2 \\ 0 & \text{sonst} \end{cases}$$

Die beiden Stichprobenverteilungen besitzen also die folgenden Parameter:

$$E(\bar{X}) = 1,5 \qquad E(\bar{X}_G) = 1,4571$$
$$V(\bar{X}) = 0,125 \qquad V(\bar{X}_G) = 0,1268$$

Es fällt auf, daß der Erwartungswert von $\bar{X}_G$ *nicht dem Erwartungswert der Grundgesamtheit entspricht* $(\mu = 1.5)$*. Weiterhin besitzt* $\bar{X}_G$ *einen größeren Stichprobenfehler als* $\bar{X}$ $(0,1268 > 0,125)$*.* ❏

Von Interesse ist die Verteilung dieser Stichprobenfunktionen. Bei bekannter Verteilung von X (z.B. der Normalverteilung) ist es möglich, die Verteilung der Stichprobenfunktion exakt zu bestimmen. Ist die Verteilung nicht bekannt, so lassen sich jedoch auf der Basis der Aussagen der (zentralen) Grenzwertsätze zumindest approximativ die Verteilungseigenschaften angeben.

1.3 Stichprobenverteilung des arithmetischen Mittels

Aus einer (heterograden) Grundgesamtheit mit $E(X) = \mu$ und $V(X) = \sigma^2$ werden alle Stichproben gezogen. Berechnet man für alle diese Stichproben das arithmetische Mittel $\bar{x}$, so gelangt man zur *Stichprobenverteilung* von $\bar{x}$. Die Schätzfunktion $g(\cdot)$ wird somit spezi£ziert und lautet nun

$$g(X_1, \ldots, X_n) = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Die Stichprobenverteilung lässt sich durch den Erwartungswert und die Varianz charakterisieren.

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \sum_{i=1}^{n} \mu = \frac{1}{n} n \cdot \mu = \mu.$$

Bei der Berechnung der Varianz muß berücksichtigt werden, ob das Experiment mit oder ohne Zurücklegen durchgeführt wird. Für den Fall unabhängiger Stichprobenvariablen (*zmz*) erhält man

$$V(\bar{X}) = V\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2 = \frac{1}{n^2} n \cdot \sigma^2 = \frac{\sigma^2}{n}.$$

Die Varianz für den Fall abhängiger Stichprobenvariablen (*zoz*) lautet (ohne Herleitung)

$$V(\bar{X}) = \frac{\sigma^2}{n} \frac{N-n}{N-1}.$$

1.4 Stichprobenverteilung des Anteilswertes

Ist das Merkmal X in der Grundgesamtheit Bernoulli–verteilt mit den Wahrscheinlichkeiten $P(X = 1) = \pi$ (Erfolg) und $P(X = 0) = 1 - \pi$ (Mißerfolg), so sind alle Elemente einer n Elemente umfassenden Stichprobe identisch Bernoulli–verteilt mit $E(X_i) = \pi$ und $V(X_i) = \pi(1 - \pi)$. Dann gibt die Summe $\sum_{i=1}^{n} X_i$ die *Anzahl* und der Ausdruck $\frac{1}{n} \sum_{i=1}^{n} X_i$ den *Anteil* der Erfolge an. Die Stichprobenfunktion für den Anteil der Erfolge soll mit

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

gekennzeichnet werden. Die Verteilungseigenschaften der Stichprobenfunktion können wegen der vorliegenden Verteilungsinformation über die Grundgesamtheit genau angegeben werden. Bei Ziehen mit Zurücklegen ist der Anteil relativiert binomialverteilt und bei Ziehen ohne Zurücklegen ist er relativiert hypergeometrisch verteilt. Der Erwartungswert der Stichprobenverteilung ist dann gegeben durch

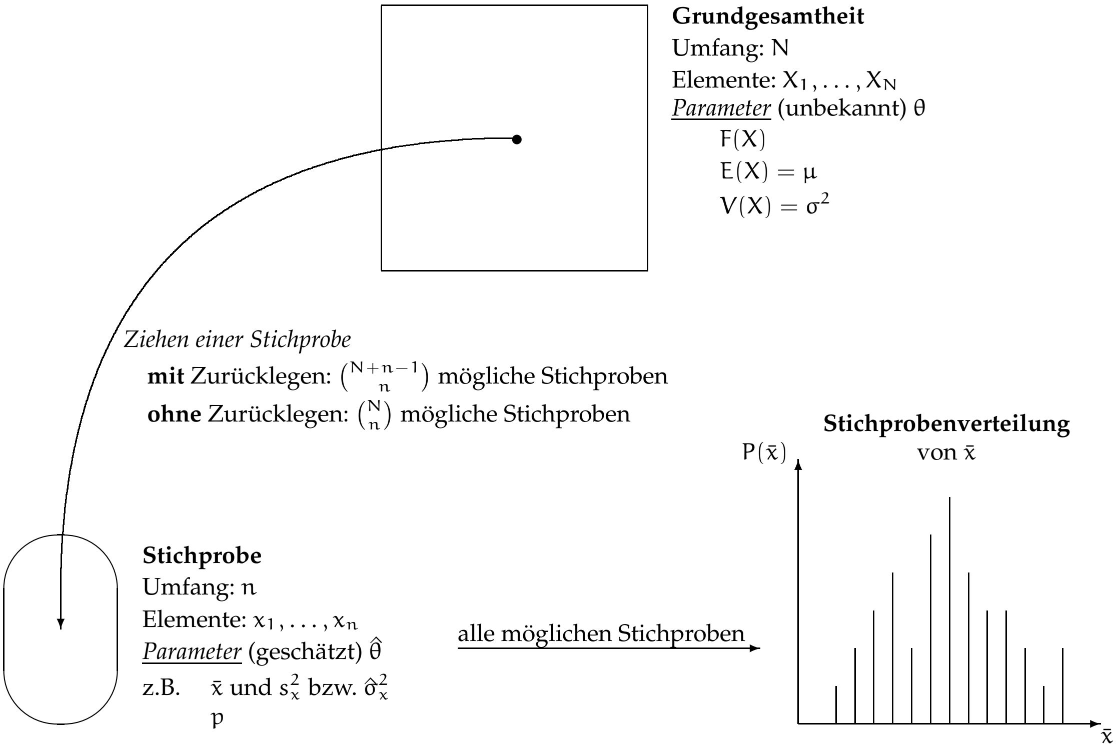
$$E(p) = \pi$$

und die Varianz ergibt sich als

$$V(p) = \frac{\pi(1-\pi)}{n} \quad (zmz) \qquad \text{bzw.} \qquad V(p) = \frac{\pi(1-\pi)}{n} \frac{N-n}{N-1} \quad (zoz).$$

Abbildung 1 faßt das bis hierher Beschriebene nochmal zusammen.

Abb. 1 Grundgesamtheit und Stichproben



2 Grundzüge der Schätztheorie

Wie bereits erwähnt wurde, handelt es sich bei der Schätzfunktion $\hat{\theta}$ um eine Zufallsvariable mit einer bestimmten Stichprobenverteilung. Das bedeutet aber, daß konkrete Schätzwerte mehr oder weniger stark von θ abweichen können. Es muß nun aus der Vielzahl der möglichen Schätzfunktionen $g(\cdot)$ diejenige ausgewählt werden, welche die Kriterien für *gute* Schätzer erfüllt. Konkret heißt das: Warum wird z.B. gerade das arithmetische Mittel zur Momentschätzung herangezogen und nicht irgendein anderer Schätzer?

2.1 *Erwartungstreue*

Die allgemeine Zielsetzung bei der Beurteilung von Schätzfunktionen soll sein, daß man mit der Schätzfunktion $\hat{\theta}$ Werte erhält, welche lediglich zufällig und nicht systematisch vom tatsächlichen Parameter θ abweichen. Ist dies der Fall, spricht man von einem *erwartungstreuen* Schätzer. Ist die Abweichung nicht zufällig, spricht man von *verzerrten* Schätzern. Der Erwartungswert des Schätzfehlers (der *Bias*) sollte für jeden Stichprobenumfang verschwinden:

$$B = E(\hat{\theta} - \theta) = 0 \qquad \text{bzw.} \qquad E(\hat{\theta}) = \theta.$$

2.2 *Ef£zienz*

Aus der Klasse der erwartungstreuen Schätzer, wird derjenige mit der geringsten Varianz ausgewählt. Sind also $\hat{\theta}_1$ und $\hat{\theta}_2$ zwei erwartungstreue Schätzer, dann ist der Schätzer $\hat{\theta}_1$ dann *ef£zienter* als $\hat{\theta}_2$, wenn bei gleichem Stichprobenumfang n gilt:

$$V(\hat{\theta}_{1,n}) < V(\hat{\theta}_{2,n}).$$

2.3 Asymptotische Erwartungstreue

Bei verzerrten Schätzern ist der Bias ungleich null. Diese Schätzer heißen jedoch dann *asymptotisch erwartungstreu*, wenn diese Verzerrung mit wachsendem Stichprobenumfang gegen Null strebt, also

$$\lim_{n\to\infty} E(\hat\theta - \theta) = 0 \qquad \text{bzw.} \qquad \lim_{n\to\infty} E(\hat\theta) = \theta.$$

2.4 Konsistenz

Konsistenz bedeutet, daß die Varianz von (erwartungstreuen) Schätzern mit zunehmendem Stichprobenumfang gegen null konvergiert. Für die praktische Durchführung von z.B. Stichprobenerhebungen bedeutet das, daß ein Mehraufwand bei der Erhebung (größerer Stichprobenumfang) auch ein besseres[2] Resultat bei der Schätzung liefert.

Damit ein Schätzer konsistent ist, muß also gelten

$$\lim_{n\to\infty} V(\hat\theta) = 0.$$

Bei verzerrten Schätzern geschieht die Überprüfung auf Konsistenz mittels des *Mean Square Errors* (MSE). Dieser ist de£niert als

$$MSE = E\left(\hat\theta - \theta\right)^2.$$

Durch eine Zerlegung kann dieser Ausdruck auf eine getrennte Betrachtung des Bias und der Varianz des Schätzers reduziert werden.

$$\begin{aligned}
E\left(\hat\theta - \theta\right)^2 &= E\left[(\hat\theta \underbrace{- E(\hat\theta)) + (E(\hat\theta)}_{=0} - \theta)\right]^2 \\
&= \underbrace{E\left[\hat\theta - E(\hat\theta)\right]^2}_{V(\hat\theta)} + 2E\left[\underbrace{(\hat\theta - E(\hat\theta)) \cdot (E(\hat\theta) - \theta)}_{=0}\right] + \underbrace{E\left[E(\hat\theta) - \theta\right]^2}_{=B^2} \\
&= V(\hat\theta) + B^2
\end{aligned}$$

Damit der verzerrte Schätzer konsistent ist, muß der Grenzwert des MSE null sein, also

$$\begin{aligned}
\lim_{n\to\infty} MSE &= \lim_{n\to\infty} E\left[\left(\hat\theta - \theta\right)^2\right] = 0 \\
&= \lim_{n\to\infty} \left[V(\hat\theta) + B^2\right] = 0 \\
&= \lim_{n\to\infty} V(\hat\theta) + \lim_{n\to\infty} B^2 = 0
\end{aligned}$$

Da beide Ausdrücke nicht negativ sein können, liegt Konsistenz dann vor, wenn sowohl die Varianz als auch der Bias mit zunehmendem Stichprobenumfang verschwinden. Ein konsistenter Schätzer muß also mindestens asymptotisch erwartungstreu sein.

Beispiel 2 *Gegeben sei die Schätzfunktion $\bar X = \frac{1}{n}\sum_{i=1}^n X_i$ als Schätzfunktion für den Erwartungswert μ der Grundgesamtheit. Es gilt dann*

$$E(\bar X) = \frac{1}{n} E\left(\sum X_i\right) = \frac{1}{n}\sum E(X_i) = \frac{1}{n} n\mu = \mu$$

Das bedeutet, daß das arithmetische Mittel ein erwartungstreuer Schätzer für den Erwartungswert ist.

[2] Besser in dem Sinn, daß die Wahrscheinlichkeit für große Schätzfehler durch eine Vergrößerung des Stichprobenumfangs deutlich verringert werden kann.

Die Varianz des arithmetischen Mittels wurde weiter oben bereits hergeleitet. Diese beträgt $V(\bar{X}) = \sigma^2/n$. *Man sieht, daß dieser Ausdruck mit zunehmendem Stichprobenumfang verschwindet*

$$\lim_{n\to\infty} V(\bar{X}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0,$$

und somit handelt es sich bei der Schätzfunktion $\bar{X}$ *um einen erwartungstreuen und konsistenten Schätzer für den Erwartungswert* μ *in der Grundgesamtheit.*

Betrachten wir nun die Stichprobenvarianz $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \overline{x^2} - \bar{x}^2$ *als Schätzer für die Varianz* σ^2 *der Grundgesamtheit. Als Erwartungswert erhält man (ohne Herleitung)*

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

Die Stichprobenvarianz ist somit kein erwartungstreuer Schätzer für die Varianz der Grundgesamtheit. Man erkennt jedoch, daß dieser Schätzer asymptotisch erwartungstreu ist, da gilt

$$\lim_{n\to\infty} E(s^2) = \lim_{n\to\infty} \frac{n-1}{n}\sigma^2 = \lim_{n\to\infty}\left(1 - \frac{1}{n}\right)\sigma^2 = \sigma^2$$ ❑

2.5 Methoden zur Gewinnung von Punktschätzungen

Methode der Momente: Zur Schätzung werden die Momente der empirischen Häu£gkeitsverteilung, wie $\bar{x}, s^2, \ldots$ herangezogen.

Methode der kleinsten Quadrate: Die Schätzfunktion $\hat{\theta}$ wird so bestimmt, daß die Summe S der Abweichungsquadrate $(X_i - \theta)^2$ minimal wird, also

$$S = \sum (X_i - \hat{\theta})^2 \to \text{Min!}$$

Beispiel 3 (Vgl. Assenmacher (2000), Beispiel 8.2) *Der Parameter* μ *soll mit der Methode der kleinsten Quadrate geschätzt werden. Die zu minimierende Größe lautet demnach*

$$S = \sum (X_i - \hat{\mu})^2 \to \textit{Min}!$$

Die Lösung der ersten Ableitung ergibt

$$\begin{aligned}
\frac{dS}{d\hat{\mu}} &= -2\sum (X_i - \hat{\mu}) \stackrel{!}{=} 0 \\
\Rightarrow &\quad \sum (X_i - \hat{\mu}) = 0 \quad \Rightarrow \sum X_i - \sum \hat{\mu} = 0 \\
\Rightarrow &\quad \sum X_i = n\hat{\mu} \qquad \Rightarrow \hat{\mu} = \frac{1}{n}\sum X_i = \bar{x}
\end{aligned}$$

Maximum–Likelihood–Methode: Die Wahrscheinlichkeit, eine bestimmte Stichprobe $x_1, \ldots, x_n$ und damit auch einen konkreten Schätzwert $\hat{\theta}$ zu erhalten, ist vom Parameter der Grundgesamtheit θ abhängig. Der Parameter wird nun so geschätzt, daß die Wahrscheinlichkeit für das Eintreten der konkret vorliegenden Stichprobe maximal wird. Diese Wahrscheinlichkeit kann für unabhängige Zufallsvariablen folgendermaßen geschrieben werden:

$$\begin{aligned}
f(x_1, \ldots, x_n|\theta) &= [f(x_1|\theta) \cdot f(x_2|\theta) \cdot \cdots \cdot f(x_n|\theta)] \\
&= L(x|\theta)
\end{aligned}$$

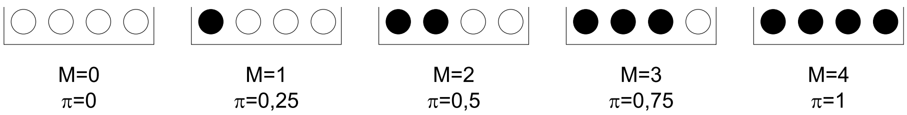
Die Funktion L nennt man *Likelihoodfunktion*. Die Maximum–Likelihood–Methode basiert darauf, den Parameter θ, so zu bestimmen, daß diese Funktion maximal wird:

$$L(\hat{\theta}) = \max_{\theta} L(\theta).$$

Die Lösung des Maximierungsproblems liefert dann den Maximum–Likelihood–Schätzer (ML) für θ.

Beispiel 4 (Vgl. von der Lippe (1998), Aufgabe 8.1.1) *Aus einer Urne mit* $N = 4$ *Kugeln werde eine Stichprobe vom Umfang* $n = 2$ *mit Zurücklegen gezogen. Die Urne enthält eine unbekannte Anzahl* M *von schwarzen und* $N - M$ *von weißen Kugeln. Eine Stichprobe ergab eine schwarze und eine weiße Kugel. Bei welchem Wert für* M *(also* $M = 1, 2, 3$ *oder* 4*) ist es am wahrscheinlichsten, daß gerade ein solches Stichprobenergebnis (also eine schwarze und eine weiße Kugel) entsteht? Man differenziere die Likelihood-Funktion nach* M *und ermittle so den unbekannten Wert* M*!*

Das Mischungsverhältnis in der Urne (d.h. die Anzahl schwarzer M *und weißer* $N - M$ *Kugeln) ist unbekannt. Es wurde eine Stichprobe vom Umfang* $n = 2$ *gezogen, die eine weiße und eine schwarze Kugel enthielt. Der Anteil* p *in der Stichprobe der schwarzen Kugeln beträgt* $p = 0,5$*. Man überlegt sich nun folgendes. Insgesamt sind 4 Kugeln in der Urne. Folglich können die folgenden Urnen existieren:*



Für welche dieser fünf Urnen ist nun die gezogene Stichprobe am wahrscheinlichksten? Dazu stellt man die Likelihoodfunktion auf. Diese lautet:

$$\begin{aligned} f(x) = L(x|n, M, N) &= \binom{n}{x} \left(\frac{M}{N}\right)^x \left(\frac{N-M}{N}\right)^{n-x} \\ &= \binom{n}{x} \pi^x (1-\pi)^{n-x} \\ &= L(\pi) \end{aligned}$$

Diese Funktion nimmt für die verschiedenen Werte von π *folgende Funktionswerte an:*

$$\begin{aligned} L(0) &= \binom{2}{1} 0^1 1^1 = 0 & L(0,25) &= \binom{2}{1} 0,25^1 0,75^1 = 0,375 \\ L(0,5) &= \binom{2}{1} 0,5^1 0,5^1 = 0,5 & L(0,75) &= \binom{2}{1} 0,75^1 0,25^1 = 0,375 \\ L(1) &= \binom{2}{1} 1^1 0^1 = 0 \end{aligned}$$

Diese Funktion nimmt für $\hat{\pi} = 0,5$ *den maximalen Wert an. Daraus folgt, die Anzahl der schwarzen Kugeln beträgt* $\hat{M} = 2$*.*

Die allgemeine Lösung lautet folgendermaßen:

$$L(x|n, M, N) = \binom{2}{1} \left(\frac{M}{4}\right)^1 \left(\frac{4-M}{4}\right)^1 = \frac{1}{8} M(4-M) = \frac{1}{2} M - \frac{1}{8} M^2$$

Man erhält dann

$$\begin{aligned} \frac{dL}{dM} &= \frac{1}{2} - \frac{1}{4} M \stackrel{!}{=} 0 && \textit{notw. Bed. für Maximum} \\ &\Rightarrow \hat{M} = 2 \\ \frac{d^2 L}{dM^2} &= -\frac{1}{4} < 0 && \textit{hinr. Bed. für Maximum} \end{aligned}$$

und somit $\hat{M} = 2$ *als ML–Schätzer für* M*.* ❑

2.6 Intervallschätzung von Anteils– und Erwartungswerten

Bei der Konstruktion von Konfidenzintervallen macht man sich die Schätz– und Verteilungseigenschaften von Punktschätzern zunutze. Das Stichprobenmittel $\bar{X}$ ist bekanntlich ein erwartungstreuer Schätzer für den Erwartungswert μ in der Grundgesamtheit. Weiterhin ist bekannt, daß die Stichprobenverteilung des arithmetischen Mittels der Normalverteilung gehorcht und zwar entweder auf der Basis der Grenzwertsätze für $n > 30$ oder wegen der Reproduktionseigenschaften der Normalverteilung, wenn alle X_i normalverteilt sind. Somit gilt (approximativ)

$$\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right) \qquad \rightsquigarrow z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \qquad \Rightarrow z \sim N(0;1)\,.$$

Aufgrund der Verteilungseigenschaften des arithmetischen Mittels kann man nun einen Bereich bestimmen, in dem sich mit einer bestimmten *Sicherheitswahrscheinlichkeit* eine konkrete Realisation von $\bar{X}$ ereignet. Dazu berechnet man

$$P\left(-z_\alpha \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq +z_\alpha\right) = 1 - \alpha.$$

Jeder konkrete Schätzwert liefert eine erwartungstreue Schätzung für den unbekannten Parameter. Löst man nun die Ungleichung in der Klammer nach μ auf, erhält man

$$P\left(\bar{X} - z_\alpha \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Das so konstruierte Intervall gibt also eine Bandbreite von Realisationen von $\bar{X}$ an, welche mit einer Wahrscheinlichkeit von $1 - \alpha$ den unbekannten Parameter einschließen.

Interpretation von Konfidenzintervallen: Beim Konfidenzintervall sind die bestimmenden Werte der Zufallsvariable nicht bekannt. Die Grenzen des Intervalls sind zufällig, da sie von von der Realisation der Stichprobe abhängen. Der wahre Parameter dagegen ist eine Konstante. Es kann folglich auch nicht gesagt werden, daß der Parameter zwischen x_u und x_o liegt. Die richtige Aussage bei Konfidenzintervallen lautet vielmehr:

> Mit einer Wahrscheinlichkeit von $1 - \alpha$ enthält das konstruierte Konfidenzintervall den wahren Parameter.

Oder anders:

> Würde das Zufallsexperiment beliebig oft wiederholt, so würden die konstruierten Konfidenzintervalle in 95 von 100 Fällen den wahren Parameter enthalten.

Allgemeine Form von Konfidenzintervallen: Das Konfidenzintervall in seiner allgemeinen Form lautet:

$$P\left(\hat{\theta} - z_\alpha \cdot \sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + z_\alpha \cdot \sigma_{\hat{\theta}}\right) = 1 - \alpha.$$

Die Grenzen des Konfidenzintervalls lauten demnach

$$\left[\hat{\theta} - z_\alpha \cdot \sigma_{\hat{\theta}}; \hat{\theta} + z_\alpha \cdot \sigma_{\hat{\theta}}\right].$$

Den Mittelpunkt des Konfidenzintervalls bildet dabei der *Schätzwert* $\hat{\theta}$ in der Stichprobe. Wenn die Grundgesamtheit heterograd ist, bildet das arithmetische Mittel $\bar{x}$ eine erwartungstreue Schätzung für den Parameter μ in der Grundgesamtheit. Bei homograden Grundgesamtheiten wird der Anteil der Erfolge in der Stichprobe $p = \sum X/n$ als erwartungstreue Schätzung für die Erfolgswahrscheinlichkeit π in der Grundgesamtheit herangezogen.

Die *Breite* des Konfidenzintervalls berechnet sich als

$$2 \cdot z_\alpha \cdot \sigma_{\hat{\theta}}$$

und wird durch die folgenden Größen beein¤ußt:

1. Vom *Stichprobenfehler* $\sigma_{\hat{\theta}}$. Dieser wird nach der unten folgenden Übersicht bestimmt (Vgl. von der Lippe (1998), Übersicht 8.9). Je größer der Stichprobenfehler ist, desto breiter wird das Konfidenzintervall.
2. Von der *Vertrauenswahrscheinlichkeit* $1-\alpha$. Diese Wahrscheinlichkeit legt das dazugehörige $1-\alpha$–Quantil der Normalverteilung bzw. der t–Verteilung fest. Mit steigender Sicherheit,wird der entsprechende z_α–Wert immer größer und das Intervall immer breiter. (Extremfall: Mit 100%iger Sicherheit be£ndet sich der Parameter in einem Intervall von $-\infty$ bis $+\infty$)
3. Vom Stichprobenumfang n: Bei sonst gleichen Parametern bewirkt eine Erhöhung des Stichprobenumfangs, daß der Stichprobenfehler abnimmt und somit das Konfidenzintervall nicht mehr so breit ist.

Bei der Ermittlung des Stichprobenfehlers und des der Sicherheitswahrscheinlichkeit $1-\alpha$ zugeordneten z–Wertes kann Abbildung 2 herangezogen werden.[3]

Beispiel 5 *Aus einer Gesamtheit von Männern und Frauen, die sich einer stationären Entwöhnungsbehandlung für Alkoholabhängige unterzogen hatten, wurde eine Stichprobe vom Umfang $n = 808$ gezogen. Nach 18 Monaten waren von diesen Personen 290 ungebessert.*

Zu bestimmen ist ein 99%iges Konfidenzintervall für die Mißerfolgswahrscheinlichkeit.

In diesem Fall handelt es sich um die Bestimmung eines Konfidenzintervalls für Anteilswerte. Der Umfang der Grundgesamtheit ist unbekannt, folglich wird die Endlichkeitskorrektur vernachlässigt. Die Varianz der Grundgesamtheit ist unbekannt, folglich wird der Stichprobenfehler gemäß der Übersicht bestimmt durch

$$\hat{\sigma}_q = \sqrt{\frac{pq}{n-1}}$$

Das Konfidenzintervall hat also folgende Gestalt:

$$P\left(q - z\hat{\sigma}_q \leq 1-\pi \leq q + z\hat{\sigma}_q\right) = 1-\alpha.$$

Das Vertrauensniveau determiniert den z–Wert, welcher hier eigentlich in der t–Verteilung abgelesen werden müsste. Da der Stichprobenumfang aber größer als 30 ist, kann wieder die Standardnormalverteilung herangezogen werden. Es ergeben sich also folgende Bestimmungsgrößen für das Konfidenzintervall:

$$\begin{aligned}
q &= \frac{290}{808} = 0,3589 \\
\hat{\sigma}_q &= \sqrt{\frac{0,6411 \cdot 0,3589}{807}} = 0,0169 \\
1-\alpha = 0,99 \quad &\Rightarrow \alpha = 0,01 \quad \Rightarrow z_\alpha = 2,5758
\end{aligned}$$

Das Konfidenzintervall lautet also

$$\begin{aligned}
0,3589 - 2,5758 \cdot 0,0169 \leq (1-\pi) &\leq 0,3589 + 2,5758 \cdot 0,0169 \\
0,3154 \leq (1-\pi) &\leq 0,4024
\end{aligned}$$

❑

[3] Vgl. auch von der Lippe (1998), Übersicht 8.9, S.88.

Abb. 2 Stichprobenfehler und $1-\alpha$ –Quantil

Bestimmung des Stichprobenfehlers $\sigma_{\hat{\theta}}$

σ^2 der GG bekannt? — ja →

	heterograd	homograd
zmz	$\frac{\sigma}{\sqrt{n}}$	$\sqrt{\frac{\pi(1-\pi)}{n}}$
zoz	$\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$	$\sqrt{\frac{\pi(1-\pi)}{n}\cdot\frac{N-n}{N-1}}$

nein →

	heterograd[a]	homograd
zmz	$\frac{\hat{\sigma}}{\sqrt{n}}$	$\sqrt{\frac{pq}{n-1}}$
zoz	$\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N}}$	$\sqrt{\frac{pq}{n-1}\cdot\frac{N-n}{N}}$

Bestimmung von z

σ^2 der GG bekannt? — ja →

Der z–Wert entspricht dem $1-\alpha$ –Quantil der Standardnormalverteilung
Ablesebeispiel: $1-\alpha = 0,95$
$\Rightarrow \Phi(z) = 0,95$
$\Rightarrow z = \pm 1,96$

nein — $n > 30$ →

Der z–Wert entspricht dem $1-\frac{\alpha}{2}$ –Quantil der t–Verteilung[b] mit $n-1$ Freiheitsgraden
Ablesebeispiel: $n = 20 \Rightarrow r = n-1 = 19$
$1-\alpha = 0,95 \Rightarrow \alpha = 0,05$
$\Rightarrow 1-\frac{\alpha}{2} = 0,975$
$\Rightarrow t_{19;0,975} = 2,09$

[a] Dabei stellt $\hat{\sigma}^2 = \frac{1}{n-1}\sum(x_i-\bar{x})^2$ die geschätzte Varianz der Grundgesamtheit dar.
[b] Vgl. von der Lippe (1998), Tabelle **T–4**, S. 120

3 Testtheorie

3.1 Einführung

In den vorherigen Abschnitten wurden Punktschätzer und Konfidenzintervalle behandelt, d.h., man hat versucht auf der Basis der Stichprobe einen konkreten Wert bzw. ein Intervall potentieller Werte für den unbekannten Parameter zu bestimmen.

Im folgenden werden Annahmen/Vermutungen (*Hypothesen*) über die Parameter der Verteilung einer Grundgesamtheit überprüft (*statistischer Test*). Dabei sind folgende Begriffe von Bedeutung:

Statistische Hypothese:	Annahme/Vermutung über die Verteilung der Grundgesamtheit.
Statistischer Test:	Auf der Basis einer Prüfgröße T ($\leadsto$ *Stichprobenfunktion*) wird mit einer vorgegebenen Irrtumswahrscheinlichkeit α (*Signi£kanzniveau*) über die Verwerfung oder Nichtverwerfung einer Hypothese entschieden.

3.2 Einstichprobentest

Unabhängig von dem konkret durchzuführenden Test (Parameter– oder Verteilungstest) verläuft die Durchführung eines statistischen Tests nach dem folgenden Schema:

Formulierung der Nullhypothese und der Alternativhypothese:

$$
\begin{array}{llll}
H_0 : \theta = \theta_0 & \text{vs.} & H_1 : \theta > \theta_0 & \text{einseitig } \textit{nach oben} \\
 & & < \theta_0 & \text{einseitig } \textit{nach unten} \\
 & & \neq \theta_0 & \text{zweiseitig}
\end{array}
$$

Bestimmung der Prüfgröße/Teststatistik: Aufgrund der Stichprobenverteilung der Schätzer wird eine Prüfgröße oder auch *Teststatistik* konstruiert. Das geschieht, indem man unter der Bedingung der Gültigkeit der Nullhypothese einen bestimmten Wert T ermittelt. Zur Ermittlung der Prüfgröße kann Abbildung 3 herangezogen werden.[4]

Festlegung des Signi£kanzniveaus α, Bestimmung des kritischen Wertes: Die Wahl des Signi£kanzniveaus bestimmt den kritischen Wert. Der kritische Wert entspricht dem dem Signi£kanzniveau zugeordneten Quantil der (theoretischen) Stichprobenverteilung der Prüfgröße. In unseren Fällen ist das die Standardnormal– bzw. t–Verteilung. Für den Fall der Normalverteilung und einem Signi£kanzniveau von $\alpha = 0,05$ erhält man folgende möglichen Werte:

$$
\begin{array}{ll}
c = +1,6449 & \text{einseitig } \textit{nach oben} \\
c = -1,6449 & \text{einseitig } \textit{nach unten} \\
c = \pm 1,96 & \text{zweiseitig}
\end{array}
$$

Vergleich der Realisation der Teststatistik mit dem kritischen Wert c: Die empirische Größe (Prüfgröße) wird mit der theoretischen Größe (kritischer Wert) verglichen. Wenn die Wahrscheinlichkeit für die Realisation des empirischen Wertes kleiner als das vorgeschriebene Signi£kanzniveau ist, emp£ehlt es sich, die Nullhypothese abzulehnen. Die Entscheidung über die Annahme bzw. Ablehnung (besser Nichtannahme) der Nullhypothese geschieht mit folgenden Entscheidungshilfen: Lehne H_0 ab, wenn gilt:

$$
\begin{array}{rl}
T > c & \text{einseitig } \textit{nach oben} \\
T < c & \text{einseitig } \textit{nach unten} \\
|T| > |c| & \text{zweiseitig}
\end{array}
$$

[4] Vgl. auch von der Lippe (1998), Übersicht 9.3, S.97.

Abb. 3 Prüfgröße T und kritischer Wert c im Einstichproben–Fall

Bestimmung von T

Heterograd

	σ^2 bekannt	σ^2 unbekannt[a]
zmz	$\dfrac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$	$\dfrac{\bar{X}-\mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}}$
zoz	$\dfrac{\bar{X}-\mu_0}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$	$\dfrac{\bar{X}-\mu_0}{\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$

Homograd

	σ^2 bekannt[b]
zmz	$\dfrac{p-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$
zoz	$\dfrac{p-\pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}\frac{N-n}{N}}}$

Bestimmung des kritischen Wertes c

σ^2 der GG bekannt?

ja

Der kritische Wert c entspricht dem $1-\alpha$ –Quantil der Standardnormalverteilung

Ablesebeispiel: $1-\alpha = 0,95$
zweiseitiger Test
$\Rightarrow \Phi(z) = 0,95$
$\Rightarrow z = \pm 1,96$

nein

$n > 30$

Der kritische Wert c entspricht dem $1-\alpha$ –Quantil (beim einseitigen Test) bzw. dem $1-\frac{\alpha}{2}$ –Quantil (beim zweiseitigen Test) der t–Verteilung[c] mit $n-1$ Freiheitsgraden

Ablesebeispiel: $n = 20 \Rightarrow r = n-1 = 19$
$1-\alpha = 0,95$
einseitiger Test
$\Rightarrow t_{19;0,95} = 1,73$

[a] Dabei stellt $\hat{\sigma}^2 = \dfrac{1}{n-1}\sum(x_i-\bar{x})^2$ die geschätzte Varianz der Grundgesamtheit dar.
[b] Die Varianz der Grundgesamtheit wird dabei hypothetisch angenommen.
[c] Vgl. von der Lippe (1998), Tabelle **T–4**, S. 120

3.3 Zweistichprobentest

Häu£g werden statistische Tests zum Vergleich zweier Populationen bzw. zweier Verteilungen herangezogen. Dies geschieht, um eindeutige Kausalaussagen machen zu können, wie z.B. über die Wirksamkeit von Medikamenten. Zur Durchführung ist eine Experimentgruppe und eine Kontrollgruppe notwendig. Diese beiden Gruppen werden dann hinsichtlich eines bestimmten

Merkmals untersucht und überprüft, ob sich die Merkmalswerte in den beiden Gruppen signi£kant unterscheiden.

Dabei wird angenommen, daß zwei unabhängige Stichproben vorliegen, d.h., über die Zugehörigkeit einer Person (oder eines Elements) zur einen oder anderen Gruppe wird nach dem Zufallsprinzip entschieden. Für die Elemente der Gruppe 1 sei angenommen, daß sie aus einer Grundgesamtheit $N(\mu_1; \sigma_1^2)$ bzw. $Z(\pi_1)$ stammen und für die Elemente der Gruppe 2 analog.

Heterograder Fall: ($\rightsquigarrow$ Vergleich zweier Mittelwerte)

Die Hypothesen lauten:

$$H_0 : \mu_1 = \mu_2 \qquad \text{vs.} \qquad H_1 : \mu_1 \neq \mu_2$$

bzw.

$$\begin{aligned} H_0 : \mu_1 - \mu_2 = \Delta = 0 \qquad \text{vs.} \qquad H_1 : \mu_1 - \mu_2 = & \Delta \neq 0 \\ & \Delta > 0 \\ & \Delta < 0 \end{aligned}$$

Das arithmetische Mittel ist ein erwartungstreuer Schätzer für den Parameterwert der Grundgesamtheit μ. Bei unabhängigen Stichproben gilt für die Differenz zweier Mittelwerte somit:

$$E(\overline{x}_1 - \overline{x}_2) = E(\overline{x}_1) - E(\overline{x}_2) = \mu_1 - \mu_2$$

$$V(\overline{x}_1 - \overline{x}_2) = V(\overline{x}_1) + V(\overline{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Eine Standardisierung der Mittelwertdifferenz führt zu

$$T = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Unter der Gültigkeit der Nullhypothese erhält man:

$$\begin{aligned} T_{|H_0} = \frac{(\overline{x}_1 - \overline{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} & \sim N(0;1) \text{ wenn } \sigma_1^2 \text{ und } \sigma_2^2 \text{ bekannt sind.} \rightsquigarrow \text{Fall 1 und 2!} \\ & \sim t_{n-2}, \text{ mit } n = n_1 + n_2, \text{ wenn } \sigma_1^2 \text{ und } \sigma_2^2 \text{ unbekannt aber} \\ & \quad \text{gleich sind.} \rightsquigarrow \text{Fall 3!} \\ & \sim t_\nu, \text{ wenn } \sigma_1^2 \text{ und } \sigma_2^2 \text{ unbekannt und ungleich sind.} \\ & \quad \rightsquigarrow \text{Fall 4!} \end{aligned}$$

Homograder Fall: ($\rightsquigarrow$ Vergleich von Anteilswerten)

$$\begin{aligned} H_0 : \Delta = \pi_1 - \pi_2 = 0 \qquad \text{vs.} \qquad H_1 : \Delta = \pi_1 - \pi_2 & \neq 0 \\ & > 0 \\ & < 0 \end{aligned}$$

Als Prüfgröße erhält man

$$T = \frac{p_1 - p_2}{\sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0;1)$$

mit

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}.$$

3.4 α– und β–Fehler

Bei der Entscheidung für eine der beiden Hypothesen können zwei Arten von Fehlern begangen werden:

Testentscheidung	wirklicher Zustand	
	H_0 ist wahr	H_1 ist wahr
H_0	kein Fehler ($\geq 1-\alpha$)	Fehler 2. Art (β–Fehler $\leq \beta$)
H_1	Fehler 1. Art (α–Fehler $\leq \alpha$)	kein Fehler ($\geq 1-\beta$)

Unter der Annahme, daß H_0 zutreffend ist, soll gelten

$$P(T \in K_\alpha | H_0) \leq \alpha,$$

wobei K_α den Ablehnungsbereich kennzeichnet. Daraus folgt unmittelbar

$$P(T \notin K_\alpha | H_0) \geq 1 - \alpha.$$

Die Wahrscheinlichkeit β, d.h. die Wahrscheinlichkeit dafür, H_0 fälschlicherweise anzunehmen, beträgt somit

$$P(T \notin K_\alpha | H_1) \leq \beta.$$

Die Prüfgröße fällt also in den Annahmebereich, welcher bei Gültigkeit von H_0 spezi£ziert ist, obwohl H_1 gültig ist.

Beispiel 6 (homograder Fall) *Ein Pharmakonzern hat ein neues Medikament auf den Markt gebracht. Der Konzern behauptet, daß die Heilungschancen bei der Verwendung dieses Medikaments bei 0.75 liegen.*

☞ *Diese Aussage ist eine statistische Hypothese, welche durch eine Stichprobe überprüft werden kann, z.B. durch $n = 40$ Patienten, die mit dem Medikament behandelt werden. Da das Medikament nicht an allen potentiellen Patienten getestet werden kann, sondern nur an diesen 40, kann man nicht sicher sein, ob die aufgestellte Hypothese richtig oder falsch ist.*

*In einem Versuch wurden von $n = 40$ Versuchspersonen $x = 25$ Patienten geheilt ($p = 25/40 = 0,625$). Das bestärkt die Krankenkassenverbände in der Annahme, daß das neue Medikament **nicht** die angegebene Heilungschance besitzt.*

☞ *Aufgrund der vorliegenden Stichprobe wird nun ein Hypothesentest durchgeführt, an dessen Ende die Entscheidung für eine der beiden Hypothesen steht.*

1. *Formulierung der Hypothese bzw. der Gegenhypothese:*

   $$H_0 : \pi = 0,75 \qquad vs. \qquad H_1 : \pi < 0,75$$

2. *Aufstellen der Teststatistik:*
   homograder Fall, zmz – da der Umfang der Grundgesamtheit nicht bekannt ist:

   $$T = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = \frac{0,625 - 0,75}{\sqrt{\frac{0,75(1-0,75)}{40}}} = -1,8257.$$

3. *Ermittlung des kritischen Wertes:*
   Bei der Wahl für ein Signi£kanzniveau (hier: $\alpha = 0,05$) gelangt man zu einem kritischen Wert, der in diesem Fall der Normalverteilung entnommen wird. ($c = -1,6449$)
4. *Entscheidung:*
   Da der Wert der Teststatistik kleiner als der kritische Wert ist ($T < c$), kann die Nullhypothese zum Niveau $\alpha = 0,05$ abgelehnt werden. Das bedeutet, daß sich das Stichprobenergebnis signi£kant vom hypothetischen Parameterwert unterscheidet.
   Die Aussage des Pharmakonzerns kann auf der Basis der vorliegenden Stichprobe nicht bestätigt werden.

5. *Berechnung des β–Fehlers:*
   Vorab wird der kritische Anteilswert berechnet, d.h., der Wert, bei dem keine Entscheidung getroffen werden kann. In diesem Fall muß gelten: $T = c$

$$T = c \qquad \Rightarrow \qquad \frac{p_c - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} = -1.6449 \qquad \Rightarrow \qquad p_c = 0.6374.$$

   Wenn nun der Anteilswert in der Stichprobe kleiner als p_c *ist, fällt die Entscheidung zugunsten von* H_1*, ist er größer als* p_c*, entscheidet man sich für* H_0*.*
   Der β*–Fehler bedeutet, daß man sich für* H_0 *entscheidet – d.h.* $p > p_c$ *–, obwohl* H_1 *gültig ist:*

$$P(T \notin K_\alpha | H_1) \leq \beta.$$

$$P\left(T_{H_1} > \frac{0.6374 - 0.5}{\sqrt{\frac{0.5 \cdot 0.5}{40}}}\right) = 1 - P(T \leq 1.7378) = 0.0406.$$ ❑

3.5 Zusammenhang von Konfidenzintervallen und Hypothesentests

Das Ergebnis eines statistischen Tests heißt signi£kant, wenn H_0 abgelehnt wird. Daraus läßt sich folgendes formulieren:

– Signi£kant bedeutet, daß das um den Stichprobenwert $\hat{\theta}$ konstruierte Konfidenzintervall den hypothetischen Parameterwert θ_0 nicht enthält.
– Die Grenzen des Konfidenzintervalls schließen den unbekannten Parameterwert mit einer Vertrauenswahrscheinlichkeit von $1 - \alpha$ ein. D.h., jeder Wert im Konfidenzintervall ist *verträglich* mit der Hypothese H_0.
– Das bedeutet aber **nicht**, daß H_0 auch wahr ist, sondern lediglich, daß das Stichprobenergebnis nicht gegen den hypothetischen Parameterwert spricht.
– Wenn der hypothetische Parameterwert nicht im Konfidenzintervall liegt, so bedeutet dies, daß sich der *wahre* und der hypothetische Parameterwert unterscheiden. Sie sind *statistisch unterscheidbar* oder eben *signi£kant*!

Literaturempfehlungen

ASSENMACHER, W. (2000): Induktive Statistik; Berlin u.a., Springer.
BAMBERG, G., BAUER, F. (1998): Statistik, 10. Au¤age; München u.a., Oldenbourg.
BLEYMÜLLER, J., GEHLERT, G. und GÜLICHER, H. (1996): Statistik für Wirtschaftswissenschaftler, 10. Au¤age; München, Vahlen.
GONICK, L, SMITH, W. (1993): The cartoon guide to Statistics; New York, Harper Perennial.
HARTUNG, J., HEINE, B. (1996): Statistik – Übungen: Induktive Statistik, 3. Au¤age; München u.a., Oldenbourg.
LINDGREN, B. W. (1976): Statistical Theory, 3rd Edition; New York, Macmillan.
LIPPE, P. VON DER (1998): Induktive Statistik: Formeln, Aufgaben, Klausurentraining, 5. Au¤age; München u.a., Oldenbourg.
SCHLITTGEN, R. (1998): Einführung in die Statistik: Analyse und Modellierung von Daten, 8. Au¤age; München u.a., Oldenbourg.
SCHWARZE, J. (1997): Grundlagen der Statistik II: Wahrscheinlichkeitsrechnung und induktive Statistik, 6. Au¤age; Herne, Berlin, Neue Wirtschafts–Briefe.
SCHWARZE, J. (1999): Aufgabensammlung zur Statistik, 3. Au¤age; Herne, Berlin, Neue Wirtschafts–Briefe.