

# 13 Konfidenzintervalle

## 13.1 Schätzgenauigkeit

„**Schätzen heißt Fehler machen**“ sagt der Volksmund. Aber damit ist mehr das Raten als das Schätzen im statistischen Sinne gemeint. **Schätzen in der Statistik** heißt, Kenngrößen aus einer Stichprobe anstatt aus der Grundgesamtheit zu ermitteln.

**Schätzfehler.** Schätzwerte haben immer einen zufälligen Fehler, jedoch – so wird gehofft – keinen systematischen Fehler.

**Unbekannter Zufallsfehler.** Die Größe des Zufallsfehlers eines einzelnen Schätzwertes ist prinzipiell unbekannt – sonst könnte man den Schätzwert um den Schätzfehler korrigieren und hätte dann den wahren Wert. Aus Stichproben können die wahren Werte prinzipiell nicht ermittelt, sondern nur geschätzt werden.

**Stichprobenumfang (sample size).** Die Größe des Schätzfehlers hängt ab vom Stichprobenumfang, d.h. von der Anzahl der Fälle in der Stichprobe. Je mehr Fälle die Stichprobe umfasst, desto genauer wird die Schätzung. Wird der Stichprobenumfang immer größer und erreicht schließlich die Fallzahl der Grundgesamtheit (die endlich oder unendlich sein kann), so geht der Schätzwert in den wahren Wert (Parameter) über.

**Homogenität der Grundgesamtheit.** Die Größe des Schätzfehlers hängt auch von der Streuung innerhalb der Grundgesamtheit ab. Bei homogenen Grundgesamtheiten sind die Schätzungen genauer als bei inhomogenen, einfach weil die Streuung innerhalb der Grundgesamtheit an die Stichprobe weitergegeben wird.

## 13.2 Prinzip des Konfidenzintervalls

**Abschätzung, wie groß der Zufallsfehler mit einer gewählten Wahrscheinlichkeit höchstens ist.** Wie erwähnt, kann die Größe des Zufallsfehlers nicht bestimmt werden. Mit statistischen Verfahren lässt sich aber abschätzen, welche Größe des Zufallsfehlers mit welcher Wahrscheinlichkeit auftritt. Dann muss man sich entscheiden, welches Risiko man bereit ist zu tragen, und eine Vertrauenswahrscheinlichkeit wählen. Damit lässt sich abschätzen, wie groß der Zufallsfehler mit der gewählten Wahrscheinlichkeit höchstens ist.

Das **Konfidenzintervall** wird errechnet aus dem Schätzwert plus maximaler Zufallsfehler nach oben und dem Schätzwert minus maximaler Zufallsfehler nach unten. Damit ist ein Bereich angegeben, in dem der unbekannte wahre Wert mit der gewählten Wahrscheinlichkeit liegt.

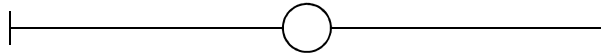
**Namen.** Das Konfidenzintervall heißt auch **Konfidenzbereich**, **Vertrauensbereich** oder **Mutungsintervall**, englisch **confidence interval**. Lateinisch heißt *confidere* vertrauen.

**Konfidenzwahrscheinlichkeit.** Die Wahrscheinlichkeit, mit der der wahre Wert irgendwo im Konfidenzbereich liegt, heißt Konfidenzwahrscheinlichkeit oder **Vertrauenswahrscheinlichkeit**. Die Konfidenzwahrscheinlichkeit wird gewählt. Meist wird 0.95 gewählt, gelegentlich aber auch 0.90 oder 0.99. Die Wahrscheinlichkeit, dass der wahre Wert nicht im Konfidenzbereich liegt, heißt Irrtumswahrscheinlichkeit und wird oft mit  $\alpha$  bezeichnet. Es gilt

$$\text{Irrtumswahrscheinlichkeit} = 1 - \text{Konfidenzwahrscheinlichkeit}$$

$$\text{Konfidenzwahrscheinlichkeit} = 1 - \text{Irrtumswahrscheinlichkeit}$$

**Darstellung von Konfidenzintervallen.** Der wahre Wert, der Parameter, ist fest, aber unbekannt. Der Schätzwert und sein Konfidenzbereich enthalten Zufallsfehler. Dies wird meist folgendermaßen dargestellt:



Wobei der Kreis der Schätzwert ist und der Konfidenzbereich durch den Strich dargestellt wird. Nach wie vor ist die Position des wahren Werts unbekannt. Er liegt aber mit der gewählten Konfidenzwahrscheinlichkeit innerhalb des Intervalls.

**Formelschreibweise.**  $\text{Wsk} \{uG \text{ KI} \leq \text{wahrer Wert} \leq oG \text{ KI}\} = 1 - \alpha$

In Worten: Die Wahrscheinlichkeit, dass der wahre Wert zwischen der unteren Grenze des Konfidenzintervalls (uG KI) und der oberen Grenze des Konfidenzintervalls (oG KI) liegt, ist die gewählte Konfidenzwahrscheinlichkeit  $(1 - \alpha)$ . Das Zeichen  $\alpha$  wird auch an anderen Stellen für Irrtumswahrscheinlichkeit verwendet.

**Beispiel für Konfidenzintervall des Medians.** An einer Stichprobe vom Umfang  $n = 100$  wurde eine mediane Überlebenszeit von 27.5 Monaten berechnet. Mit 95% Konfidenzwahrscheinlichkeit reicht das Konfidenzintervall von 20.2 bis 37.0 Monaten. Damit ist die wahre mediane Überlebenszeit in der Grundgesamtheit weiterhin unbekannt, jedoch liegt sie mit der gewählten Konfidenzwahrscheinlichkeit von 95% irgendwo zwischen 20.2 und 37.0 Monaten. Mit 95% Wahrscheinlichkeit hat die Stichprobe die mediane Überlebenszeit um nicht mehr als 7.3 Monate überschätzt und um nicht mehr als 9.5 Monate unterschätzt.

Die **Größe des Konfidenzintervalls** hängt ab von der Anzahl der Fälle in der Stichprobe (Stichprobenumfang), der Streuung innerhalb der Grundgesamtheit und von der gewählten Konfidenzwahrscheinlichkeit. Je größer die Konfidenzwahrscheinlichkeit gewählt wird, desto größer wird auch der Konfidenzbereich. Möchte man ganz sicher gehen und wählt als Konfidenzwahrscheinlichkeit 1.0, so erstreckt sich der Konfidenzbereich z.B. des Mittelwerts von minus unendlich bis plus unendlich. Es hilft somit nicht weiter, auf völlige Sicherheit zu gehen.

**Kenngößen.** Ein Konfidenzintervall kann für nahezu jede statistische Maßzahl berechnet werden. In diesem Buch werden Konfidenzintervalle für Mittelwert, Standardabweichung, Anteile und relative Häufigkeiten und für den Korrelationskoeffizienten vorgestellt. Für die anderen statistischen Kenngrößen sind auch Konfidenzintervalle bekannt und können in der Literatur nachgeschlagen oder mit einer Statistiksoftware berechnet werden.

**Symmetrie des Konfidenzintervalls.** Bei Merkmalen, die im Prinzip von  $-\infty$  bis  $+\infty$  definiert sind, ist das Konfidenzintervall meistens symmetrisch und lässt sich dann darstellen als Schätzwert  $\pm$  maximaler Zufallsfehler, z.B.  $25.3 \pm 1.8$ . Bei Merkmalen, die nur einen bestimmten Definitionsbereich haben, wie z.B. ein Anteil, der nur zwischen 0 und 1 variieren kann, ist das Konfidenzintervall meist asymmetrisch. Vor allem an den Definitionsgrenzen des Merkmals, bei einem Anteil also knapp über 0 und knapp unter 1, wird das Konfidenzintervall stark asymmetrisch.

**Punkt- und Intervallschätzer.** Die Größe des Konfidenzintervalls wird aus der Stichprobe berechnet und ist deshalb auch nur ein Schätzwert. Der eigentliche Schätzwert wird als Punktschätzer und sein Konfidenzintervall als Intervallschätzer bezeichnet.

**Kurzschreibweise.** Konfidenzintervalle werden sehr häufig verwendet, deshalb gibt es die Kurzschreibweise mit z.B. 95% KI = 10.5 bis 20.1. Dabei geben die 95% die Konfidenzwahrscheinlichkeit an und 10.5 bis 20.1 ist das Konfidenzintervall. Bei symmetrischen Konfidenzintervallen ist auch die Kurzschreibweise 95% KI =  $15.3 \pm 4.8$  üblich.

**Grenzen des Konfidenzintervalls.** Der Konfidenzbereich berücksichtigt nur den Zufallsfehler der Stichprobe – sonst nichts! Der Konfidenzbereich deckt nicht ab die systematischen Fehler der Stichprobe, eine falsche Formel, Rechenfehler und sonstige Fehler. Der Konfidenzbereich kann für nahezu alle statistischen Kenngrößen berechnet werden, z.B. für Prävalenzschätzer, Sensitivitätsschätzer, Risikoschätzer, Regressionskoeffizienten usw.

### 13.3 Konfidenzintervall des Mittelwerts

**Gedankenversuch zur Verteilung der Mittelwerte aus vielen Stichproben.** Man nehme eine unendlich große Grundgesamtheit, ziehe daraus viele Stichproben jeweils vom Umfang  $n$ , z.B.  $n = 100$  Fälle und berechne aus jeder der vielen Stichproben den Mittelwert. Natürlich werden die vielen Stichproben leicht unterschiedliche Mittelwerte liefern. Nun betrachten wir die Häufigkeitsverteilung dieser Stichproben-Mittelwerte. Nach dem zentralen Grenzwertsatz (Kap.11.3, S.211) ist es eine Normalverteilung. Der Mittelwert dieser Verteilung der Mittelwerte ist der beste verfügbare Schätzer für den wahren Mittelwert. Die Standardabweichung der Verteilung der Mittelwerte ist aber viel kleiner als die Standardabweichung der Daten. Mit zunehmender Fallzahl in den Stichproben werden die Mittelwerte der einzelnen Stichproben immer genauer geschätzt, sie variieren mit zunehmendem Stichprobenumfang immer weniger. Demzufolge ist die Standardabweichung der Verteilung der Mittelwerte viel kleiner als die Standardabweichung der Daten.

**Standardfehler des Mittelwerts = standard error of the mean = SEM.** Mathematiker können zeigen, dass sich bei diesem Gedankenversuch die Standardabweichung der Verteilung der Mittelwerte nach folgender Formel berechnet:

$$\text{SEM} = \frac{s}{\sqrt{n}}$$

Der SEM lässt sich nach dieser Formel auch berechnen, wenn nicht viele, sondern nur eine einzige Stichprobe vorliegt. Von dieser Stichprobe lässt sich die Standardabweichung  $s$  der Daten bestimmen, die Fallzahl  $n$  dieser Stichprobe ist ohnehin gegeben, und damit kann der SEM berechnet werden.

**Bedeutung des SEM.** Der SEM ist viel kleiner als die Standardabweichung  $s$  der Daten. Eitle Forscher und Autoren geben deshalb lieber den SEM an als die Standardabweichung  $s$ , einfach weil die Streuungsbalken kleiner und damit schöner sind. Der Standardfehler eines Schätzers ist eine wichtige theoretische Größe, aber nicht ganz einfach zu interpretieren (siehe obigen Gedankenversuch), da er sowohl von  $s$  als auch von  $n$  abhängt. Wenn Sie in medizinischer (nicht in mathematischer!) Fachliteratur den SEM finden – und Sie finden ihn häufig – und daraus schließen, dass der Autor eitel ist und herzlich wenig von Statistik versteht, so ist Ihre Schlussfolgerung fast immer richtig. Im medizinischen Kontext ist die Angabe des SEM fast nie sinnvoll. Ein Spaßvogel ist, wer einen medizinischen Vortragenden, der in seinen Abbildungen den SEM verwendet, bittet, den SEM zu erläutern und zu interpretieren. Erklärt der Vortragende nur, der SEM werde oft verwendet, dann kann der Spaßvogel anmerken, dass Vortragende ihre eigenen Abbildungen doch verstehen sollten.

**Verteilung der Mittelwerte.** Nachdem bekannt ist, dass die Verteilung der Mittelwerte eine Normalverteilung ist und Schätzwerte für deren Mittelwert und Streuung (SEM) vorliegen, lassen sich die Wahrscheinlichkeiten für die einzelne Mittelwerte ermitteln. Am einfachsten nimmt man die Tabelle der Standard-Normalverteilung (Box 11.7, S.219) vom Typ  $(1 - 2\alpha)$  und ermittelt z.B., dass 95% der Werte (hier der Mittelwerte) im Intervall  $\mu \pm 1.96 \sigma$  liegen, weil 95% aller Werte zwischen  $z = -1.96$  und  $z = +1.96$  liegen. Der Schätzwert für  $\mu$  ist der Stichprobenmittelwert, der Schätzwert für  $\sigma$  ist der SEM. Das bedeutet, 95% aller Stichprobenmittelwerte liegen zwischen  $M \pm 1.96 * SEM$ .

**t-Verteilung.** Aus theoretischen Überlegungen, die hier nicht ausgeführt sind und die damit zusammenhängen, dass  $M$  und  $SEM$  aus der gleichen Stichprobe geschätzt werden und deshalb nicht unabhängig voneinander sind, ergibt sich, dass für kleinere Fallzahlen anstatt der Normalverteilung eine andere theoretische Verteilung verwendet werden muss, die **Student t-Verteilung**, kurz **t-Verteilung**. Die t-Verteilung wurde von **William Gosset** (1876 bis 1937) unter dem Pseudonym „Student“ eingeführt. Die t-Verteilung hat nur einen Parameter, der als Anzahl Freiheitsgrade (Fg) bezeichnet wird, wobei  $Fg = n - 1$  ist. Für größere Fallzahlen geht die t-Verteilung in die Standard-Normalverteilung über. Für das Konfidenzintervall des Mittelwerts wird sowohl bei kleineren als auch bei größeren Fallzahlen die t-Verteilung verwendet. Die t-Verteilung ist in **Box 13.1** tabelliert.

**Vertrauens- versus Irrtumswahrscheinlichkeit.** Meist wird bei der t-Verteilung nicht die hier benötigte Vertrauens- oder **Konfidenzwahrscheinlichkeit**, sondern die mit  $\alpha$  bezeichnete Irrtumswahrscheinlichkeit angegeben. Die Umrechnung ist aber einfach Vertrauenswahrscheinlichkeit =  $1 - 2\alpha$ . Wurde als Vertrauenswahrscheinlichkeit z.B. 95% gewählt, so ist die Irrtumswahrscheinlichkeit  $1 - 0.95 = 0.05$ . Die Zahl 2 bei der Irrtumswahrscheinlichkeit  $\alpha$  bedeutet „zweiseitig“ und kommt daher, dass das Konfidenzintervall sowohl Unterschätzungen als auch Überschätzungen abdeckt.

**Konfidenzintervall des Mittelwerts** ist

$$\text{Konfidenzintervall} = M \pm t * SEM \quad \text{wobei} \quad \begin{array}{l} M = \text{Stichprobenmittelwert} \\ s = \text{Standardabweichung der Stichprobe} \\ n = \text{Fallzahl, Stichprobenumfang} \\ t \text{ kommt aus der t-Verteilung (Box 13.1)} \end{array}$$

Wird der SEM nicht als eigene Größe betrachtet, sondern in diese Formel eingesetzt, so ergibt sich:

$$\text{Konfidenzintervall} = M \pm t * \frac{s}{\sqrt{n}} \quad \text{wobei} \quad \begin{array}{l} M = \text{Stichprobenmittelwert} \\ s = \text{Standardabweichung der Stichprobe} \\ n = \text{Fallzahl, Stichprobenumfang} \\ t \text{ kommt aus der t-Verteilung (Box 13.1)} \end{array}$$

Für den t-Wert wird die gewählte Konfidenzwahrscheinlichkeit (z.B. 95%) beziehungsweise die zweiseitige Irrtumswahrscheinlichkeit (im Bsp. 5%) benötigt und  $Fg = n - 1$ . Dabei ist  $n$  die Fallzahl. Mit diesen zwei Angaben kann der t-Wert aus Tabelle in Box 13.1 entnommen oder von einem Programm errechnet werden. Wird ein Programm verwendet, so wird man sich nicht den t-Wert, sondern gleich das Konfidenzintervall des Mittelwerts errechnen lassen. Wird das  $\pm$  aufgelöst, so ergibt sich

$$\begin{aligned} \text{untere Grenze des Konfidenzintervalls} &= M - t * \frac{s}{\sqrt{n}} \\ \text{obere Grenze des Konfidenzintervalls} &= M + t * \frac{s}{\sqrt{n}} \end{aligned}$$

Dabei ist  $M$  der (empirische) Mittelwert,  $s$  die (empirische) Standardabweichung,  $n$  die Fallzahl und  $t$  kommt aus der t-Tabelle mit der gewählten Konfidenzwahrscheinlichkeit und den  $(n-1)$  Freiheitsgraden.