

8. GRUNDBEGRIFFE STATISTIK, SCHÄTZEN VON PARAMETERN UND KONFIDENZINTERVALLE

8.1. GRUNDBEGRIFFE STATISTIK

(8.1.1) Überblick

In diesem Kapitel werden zunächst einige Fragestellungen aus dem Bereich der beurteilenden Statistik vorgestellt. Anschliessend wird erklärt, wie diese unter Verwendung von früher eingeführten Begriffen (vor allem dem der Zufallsgrösse) einer mathematischen Behandlung zugänglich gemacht werden können. (8.1.2)

Ferner werden einige neue Fachausdrücke erläutert und mit von früher her bekannten in Verbindung gebracht: (8.1.3)

- Die *Population* ist die Menge der Untersuchungsobjekte, sie entspricht dem Ergebnisraum Ω .
- Die Beobachtung wird durch eine Zufallsgrösse X beschrieben; der Wert einer effektiv durchgeführten Beobachtung heisst *Realisierung* dieser Zufallsgrösse.
- Unter der *Grundgesamtheit* versteht man die Menge aller denkbaren Beobachtungswerte, zusammen mit ihrer Verteilung.
- Eine *Stichprobe vom Umfang n* ist eine zufällig gewählte Folge von n Objekten aus der Population; sie kann durch eine Folge von n Zufallsgrössen X_i beschrieben werden. (8.1.4)

Da es in diesem Kapitel um allgemeine Begriffsbildungen geht, dürfte es zweckmässig sein, es noch einmal sorgfältig durchzulesen, nachdem Sie einige Erfahrungen mit der beurteilenden Statistik gesammelt haben.

(8.1.2) Womit befasst sich die beurteilende Statistik?

In Kapitel 2 haben wir uns mit der beschreibenden (oder deskriptiven) Statistik befasst. Dort ging es darum, Datenmengen auf übersichtliche Weise darzustellen (Tabellen, Graphiken) und durch geeignete Masszahlen (Lagemasse wie Durchschnitt oder Median, Streuungsmasse wie Varianz oder Standardabweichung) zu charakterisieren.

In Kapitel 8-11 soll nun eine Einführung in die beurteilende (oder schliessende) Statistik gegeben werden. Deren Aufgabe besteht — zusammenfassend gesagt — darin, aus empirisch (durch Untersuchungen) ermittelten Stichproben Rückschlüsse auf eine Gesamtheit zu ziehen. Diese etwas trockene Formulierung erläutern wir zunächst an einigen Beispielen.

Beispiel 8.1.2.A

In (2.2.2.3) waren die Gewichte von 50 zweiwöchigen, aus einem grossen Hühnerhof ausgewählten Küken angegeben worden. Das durchschnittliche Gewicht betrug 102.96 g (2.2.3.3.b), und wir würden gerne wissen, was wir über das Durchschnittsgewicht *aller* Küken aus unserm Hühnerhof sagen können. Antworten dazu werden in (8.2.3) und (8.4.1) gegeben. \boxtimes

Beispiel 8.1.2.B

Bei der Züchtung einer neuen Kartoffelsorte fand man in 7 bzw. 6 Versuchsäckern die folgenden Erträge (in kg pro Are):

Alte Sorte	410	420	430	440	450	450	480
Neue Sorte	440	450	455	480	490	505	

Die Frage ist natürlich die, ob die neue Sorte gesamthaft gesehen (nicht nur auf die Versuchsäcker bezogen) grössere Erträge als die alte Sorte ergibt. Es ist ja immerhin so, dass auch ein Feld der alten Sorte einen recht hohen Ertrag lieferte (480 kg/a). Die Antwort wird in (9.3.2) ermittelt. \boxtimes

Beispiel 8.1.2.C

Wir wollen prüfen, ob eine Münze “ausgewogen” ist. Dazu werfen wir sie einige Male, z.B. 16-mal. Auch wenn die Münze ausgewogen ist, wird man nun nicht gerade 8-mal Kopf und 8-mal Zahl erwarten. Erscheint aber z.B. 15-mal Kopf und nur einmal Zahl, wird man überzeugt sein, dass etwas faul ist. Wie aber soll man sich verhalten, wenn 10-mal Kopf und 6-mal Zahl auftritt? Siehe dazu (9.1.4). \boxtimes

Beispiel 8.1.2.D

Im Jahre 1910 zählten RUTHERFORD und GEIGER während 326 Minuten die Zerfälle bei einem radioaktiven Poloniumpräparat, und zwar wurde diese Zeitspanne in Intervalle von 7.5 Sekunden Länge aufgeteilt. In der folgenden Tabelle ist die Anzahl der Intervalle angegeben, in denen 0, 1, 2, ... Zerfälle erfolgten:

Anzahl Zerfälle	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	≥ 15
Anzahl Intervalle	57	203	383	525	532	408	273	139	45	27	10	4	0	1	1	0

Kann man aufgrund dieser Daten annehmen, dass die Zahl der Zerfälle einer Poisson-Verteilung folgt? Die Antwort finden Sie in Beispiel 9.4.4.B. \boxtimes

(8.1.3) Wie kann man diese Probleme mathematisch erfassen?

Es geht nun darum, Fragestellungen wie die oben erwähnten einer mathematischen Behandlung zugänglich zu machen. Da unsere Stichproben (z.B. die Küken oder die Versuchsäcker) zufällig ausgewählt worden sind, ist es klar, dass die “Wissenschaft von

den zufälligen Ereignissen”, also die Wahrscheinlichkeitsrechnung, eine zentrale Rolle spielen wird. Ihr Einsatz wird möglich, wenn man erkennt, dass in allen Fragestellungen Zufallsgrößen stecken. Diese wollen wir nun aus den Beispielen von (8.1.2) herauslesen.

Beispiel A: Hier geht es offensichtlich um die (stetige) Zufallsgröße

$$X = \text{Gewicht von zweiwöchigen Küken.}$$

Speziell interessiert uns hier das durchschnittliche Gewicht aller solcher Küken, d.h., der Erwartungswert $E(X)$.

Beispiel B: Hier kommen gleich zwei (stetige) Zufallsgrößen vor, nämlich

$$X = \text{Ertrag pro Are der alten Sorte,}$$

$$Y = \text{Ertrag pro Are der neuen Sorte.}$$

Beispiel C: Hier geht es um ein Zufallsexperiment, das im 16-maligen Wurf einer Münze besteht. Da wir uns für die Anzahl Köpfe interessieren, drängt sich hier die diskrete Zufallsgröße

$$X = \text{Anzahl Köpfe in 16 Münzenwürfen}$$

auf.

Beispiel D: In diesem Beispiel wird die Frage gestellt, ob die Zufallsgröße

$$X = \text{Anzahl Zerfälle pro Zeitintervall von 7.5 Sekunden Dauer}$$

Poisson-verteilt sei. Diese Zufallsgröße ist diskret.

Die Zufallsgröße X beschreibt somit immer das Ergebnis einer Beobachtung, also einer Messung (wie etwa des Gewichts) oder einer Zählung (wie etwa die der radioaktiven Zerfälle). In diesem Zusammenhang nennt man das Ergebnis einer solchen Beobachtung auch eine *Realisierung* der Zufallsgröße X . Ferner haben sich in der beurteilenden Statistik weitere Fachausdrücke wie *Population* und *Grundgesamtheit* eingebürgert. Diese Ausdrücke, die wir nachher oft verwenden werden, sollen nun erläutert werden. Wir werden sehen, dass es sich bloss um neue — und in diesem Zusammenhang zweckmässige — Namen für bekannte Dinge handelt.

Eine Zufallsgröße X ist, wie Sie sich erinnern werden, streng genommen eine Abbildung von einem Ergebnisraum Ω in die Menge \mathbb{R} der reellen Zahlen (vgl. (4.1.3)):

$$X : \Omega \rightarrow \mathbb{R} .$$

Wir haben aber damals schon gesehen, dass man meistens darauf verzichtet, den Ergebnisraum Ω effektiv anzugeben (vgl. Bemerkung 8) in (4.1.3)), dass man ihn aber bei Bedarf rekonstruieren könnte. In unserer jetzigen Situation ist es nun ganz vernünftig, sich unter Ω von vornherein etwas Konkretes vorzustellen und zwar ist Ω nichts anderes als die Menge der zur Untersuchung stehenden Objekte. Diese Menge nennen wir kurz die *Population*.

Die Zufallsgrösse entspricht dem zu untersuchenden Merkmal. Sie kann verschiedene Werte annehmen (nämlich gerade alle möglichen Realisierungen); es sind aber nicht nur diese Werte von Interesse, sondern auch vor allem auch ihre Verteilung, denn man möchte natürlich wissen, welche Werte eine kleine und welche eine grosse Wahrscheinlichkeit haben. Die Menge der Werte, zusammen mit ihrer Verteilung, nennt man in der beurteilenden Statistik die *Grundgesamtheit*. Die gesamte Information über die Grundgesamtheit steckt in der dem Zufallsexperiment zugrunde liegenden Zufallsgrösse X mit der zugehörigen Verteilungsfunktion F . So gesehen ist der Gebrauch des Wortes “Grundgesamtheit” einfach eine anschauliche Umschreibung dieser Zufallsgrösse. Im Hinblick darauf sagt man dann auch beispielsweise, die Grundgesamtheit sei Poissonverteilt mit dem Parameter μ oder sie sei rechtsschief (2.2.2.5). Da alle theoretisch möglichen Werte berücksichtigt werden, ist diese Grundgesamtheit i. Allg. unendlich und somit kein reales Objekt, sondern ein solches unseres Denkens, eine Idealisierung.

Als Beispiel betrachten wir den Intelligenzquotienten (IQ). Die Population besteht hier aus allen Menschen (bzw. je nach dem Ziel der Untersuchung aus einer bestimmten Klasse von Menschen). Die Zufallsgrösse X ordnet jedem Menschen seinen IQ zu. Wenn also der Mensch ω einen IQ von 150 hat, dann ist $X(\omega) = 150$ eine Realisierung von X . Wie wir in Beispiel 5.10.5.A gesehen haben, ist X normal verteilt. Diese Tatsache drücken wir nun auch so aus, dass wir sagen, die Grundgesamtheit beim Intelligenzquotienten (oder, noch kürzer, der IQ selbst) sei normal verteilt mit $\mu = 100$, $\sigma = 15$.

Zur weiteren Illustration gehen wir unsere Beispiele von (8.1.2) nochmals durch:

Beispiel A: Hier besteht die Population (in unserer bisherigen Redeweise also der Ergebnisraum Ω) aus der Menge “aller” zweiwöchigen Küken. (Was das “aller” genau bedeuten soll, hängt von den Umständen ab: Alle zweiwöchigen Küken einer bestimmten Rasse, in einer bestimmten geographischen Region, in einem bestimmten Zeitpunkt ...). Die Zufallsgrösse X lässt sich nun wie folgt etwas genauer beschreiben: Wenn $\omega \in \Omega$ ein Küken ist, dann ist $x = X(\omega)$ sein Gewicht. Wiegt ein ausgewähltes Küken 100 Gramm, so ist dieser Wert $x = 100$ eine Realisierung von X . Die Grundgesamtheit besteht aus den möglichen Gewichten der Küken, ihre Verteilung ist nicht genau bekannt.

Beispiel B: Hier haben wir zwei Zufallsgrössen und somit auch zwei Populationen, nämlich die Menge aller Ackerstücke, die mit der alten bzw. die Menge aller Ackerstücke, die mit der neuen Sorte bepflanzt wurden und werden. Die zu untersuchenden Merkmale, dargestellt durch die Zufallsgrössen, sind die Erträge pro Hektare, was natürlich auch zwei Grundgesamtheiten liefert. Die möglichen Erträge sind zwar dieselben (rein theoretisch alle positiven Zahlen), doch interessiert uns, ob ihre Verteilungen verschieden seien und zwar insbesondere, ob bei der neuen Sorte der mittlere Ertrag (also der Erwartungswert der Grundgesamtheit) höher sei als bei der alten Sorte.

Beispiel C: Hier muss man sich unter der Population die Menge aller Serien von je 16 Münzenwürfen vorstellen; die Zufallsgrösse X gibt dann die Anzahl der Köpfe pro Serie an. Die Grundgesamtheit ist hier endlich und wird durch die Binomialverteilung mit den Parametern $n = 16$, $p = 0.5$ beschrieben, ihre möglichen Werte sind $0, 1, \dots, 16$.

Beispiel D: Hier fällt die Population etwas abstrakt aus. Sie ist nämlich die Menge aller Zeitintervalle von 7.5 Sekunden Dauer; das uns interessierende Merkmal (die Zufallsgrösse X) ist die Anzahl Zerfälle in einem solchen Zeitintervall. Wenn also in einem bestimmten Intervall 7 Zerfälle stattgefunden haben, dann ist “7” eine Realisierung der Zufallsgrösse. Die Grundgesamtheit besteht hier aus der Menge $\mathbb{N} = \{0, 1, 2, \dots\}$ aller natürlichen Zahlen mit einer unbekanntem Verteilung, von der man allerdings vermutet, es handle sich um eine Poisson-Verteilung.

Es sei noch erwähnt, dass die oben erläuterten Begriffe “Population” und “Grundgesamtheit” in der Literatur oft überhaupt nicht präzise definiert und im intuitiven Sinn verwendet werden. Manchmal findet man auch abweichende Definitionen. Oft versteht man dort unter “Grundgesamtheit” das, was hier “Population” heisst. Dies ist aber nicht weiter tragisch.

Das Ziel der statistischen Untersuchungen ist es nun — allgemein ausgedrückt — stets, Aussagen über die Grundgesamtheit zu gewinnen. In Beispiel 8.1.2.A wäre dies eine Aussage über den Erwartungswert, in Beispiel 8.1.2.D eine solche über den Typ der Verteilung. Da es in der Praxis (etwa aus Zeit- oder Kostengründen) nicht möglich sein wird, alle Objekte aus der Population zu untersuchen, greift man sich aus dieser eine Stichprobe heraus, z.B. 50 Küken, und untersucht bloss diese. Es gibt ja auch Versuche, die mit der Zerstörung des Objekts enden, wie z.B. eine Prüfung auf Bruch. Hier ist es sowieso klar, dass nur ein relativ kleiner Teil der Population untersucht werden kann.

Eine *Stichprobe* vom *Umfang* n ist also eine Folge von n Objekten aus der Population, gegeben z.B. durch die Auswahl von $n = 50$ Küken. Es ist oft zweckmässig, auch die entsprechenden Mess- oder Zählwerte (also Elemente der Grundgesamtheit) als Stichprobe zu bezeichnen, so dass man im Beispiel 8.1.2.A nicht nur die 50 ausgewählten Küken, sondern auch ihre in (2.2.2.3) angegebenen Gewichte

$$x_1 = 100, x_2 = 87, \dots, x_{50} = 106$$

als Stichprobe bezeichnet. Bei einer Stichprobe vom Umfang n wird also das der Zufallsgrösse X zugrunde liegende Experiment n -mal wiederholt.

Wenn wir von einer Stichprobe sprechen, dann meinen wir automatisch eine so genannte *Zufallsstichprobe*. Dies bedeutet, dass die Untersuchungsobjekte (Küken, Zeitintervall etc.) unabhängig voneinander ausgewählt werden und dass jedes Objekt dieselbe Wahrscheinlichkeit hat, in die Stichprobe zu gelangen.

Noch eine kleine Erläuterung zu einem etwas subtilen, aber für die Praxis nicht sonderlich wichtigen Punkt: Eine Stichprobe ist weiter oben als eine *Folge* und nicht als eine *Menge* definiert worden.

Worin besteht der Unterschied? In einer Folge darf ein Element mehrfach vorkommen. So sind z.B. (1,1,2) und (1,2,2) verschiedene Folgen, die zugehörigen Mengen sind jedoch in beiden Fällen gleich $\{1, 2\}$. Weil wir nun aber für die Stichprobe die Unabhängigkeit der Auswahl der Objekte fordern, so muss man zulassen, dass der Auswahlmechanismus (z.B. ein Zufallszahlengenerator) dasselbe Objekt mehrfach auswählt. Wenn wir uns die Population als Kugeln in einer Urne denken, heisst dies, dass die Auswahl “mit Zurücklegen” (vgl. (4.2.5)) erfolgt. Dieser Prozess liefert eine Folge, während die Auswahl “ohne Zurücklegen” eine Menge bestimmt. Wie wir aber in (4.2.5) gesehen haben, spielt der Unterschied zwischen den beiden Varianten bei einer grossen Population praktisch keine Rolle und noch viel weniger in den statistischen Anwendungen, wo man sich die Population oft sogar als unendlich gross vorstellt. So gesehen darf man ohne grosse Skrupel eine Stichprobe auch als eine *Teilmenge* der Population auffassen.

Wie bereits erwähnt, geht es in der beurteilenden Statistik darum, aus der Stichprobe Aussagen über die Grundgesamtheit zu erhalten, mit andern Worten Aussagen über die im Problem steckende Zufallsgrösse. Dies kann z.B. die Frage nach dem Typ der Verteilung dieser Zufallsgrösse (wie im Beispiel 8.1.2.D) oder nach ihrem Erwartungswert sein, wobei man hier auch von der Verteilung bzw. vom Erwartungswert der Grundgesamtheit (statt “der Zufallsgrösse”) zu sprechen pflegt. Entsprechende Beispiele werden wir später zur Genüge kennen lernen.

Was jetzt noch fehlt, ist eine wahrscheinlichkeitstheoretische Deutung des Begriffs der “Stichprobe”. Diese folgt im nächsten Abschnitt.

(8.1.4) Wie kann man eine Stichprobe beschreiben?

Wir illustrieren die Überlegung am vielbeanspruchten Beispiel der Küken. Wie in Beispiel 8.1.2.A sei

X = Gewicht eines zweiwöchigen Kükens.

Eine Stichprobe vom Umfang n (z.B. $n = 50$) wird durch Wägung von n dieser Tierchen erhalten. Das Resultat ist eine endliche Folge von n Zahlen (ein “ n -Tupel”, wie die Mathematiker zu sagen belieben, siehe (22.3) im ersten Band), konkret (Beispiel 2.2.2.3.A)

100, 87, 101, ..., 100, 106

bzw. allgemein

$x_1, x_2, x_3, \dots, x_n$.

Dabei ist x_i das Gewicht des Kükens Nummer i , d.h., des Kükens, das als “ i -tes” gewogen wurde ($i = 1, 2, \dots, n$). Da das Herausgreifen eines Kükens ein zufälliger Vorgang ist, erhalten wir so für jeden Index i ($i = 1, 2, \dots, n$) eine Zufallsgrösse X_i

X_i = Gewicht des Kükens Nummer i

oder allgemeiner

X_i = Wert der i -ten Beobachtung.

Im Sinne von (8.1.3) ist also die Zahl x_i eine Realisierung der Zufallsgrösse X_i (beachten Sie den Unterschied zwischen Klein- und Grossbuchstaben!). Der Index i dient dazu, die einzelnen Beobachtungen voneinander zu unterscheiden.

Die Stichprobe selbst wird daher durch die Folge von n Zufallsgrössen beschrieben:

$$X_1, X_2, \dots, X_n .$$

Wer will, kann diese n Zufallsgrössen zu einem “ n -dimensionalen Zufallsvektor”

$$\vec{X} = (X_1, X_2, \dots, X_n)$$

zusammenfassen, doch wollen wir die Abstraktion an dieser Stelle nicht zu weit treiben.

Wir formulieren nochmals den Unterschied zwischen X und X_i :

- X beschreibt das Gewicht eines beliebig herausgegriffenen einzelnen Kükens.
- X_i beschreibt das Gewicht des i -ten Kükens in einer beliebigen, aber gesamthaft betrachteten Stichprobe vom Umfang n .

Es besteht aber auch eine wichtige Gemeinsamkeit zwischen X und den einzelnen X_i . Da alle Beobachtungen in derselben Grundgesamtheit liegen (sie stammen ja alle aus demselben Hühnerhof) haben alle X_i dieselbe Verteilung wie X . Es gilt also für alle x :

$$P(X \leq x) = P(X_1 \leq x) = P(X_2 \leq x) = \dots = P(X_n \leq x) .$$

Man sagt, die Zufallsgrössen X_i seien identisch wie X verteilt. Diese Zufallsgrössen X_i werden wir verschiedentlich verwenden, um gewisse Sachverhalte mittels Zufallsvariablen auszudrücken, z.B. in (8.2.5).

Zum Schluss geben wir noch eine etwas präzisere, aber dafür auch etwas abstraktere Definition der Stichprobe: Eine Stichprobe vom Umfang n aus der durch X beschriebenen Grundgesamtheit ist eine Folge

$$(X_1, X_2, \dots, X_n)$$

von n unabhängigen Zufallsgrössen, welche identisch verteilt sind. Auf Englisch nennt man das independent and identically distributed - und die Abkürzung iid hat sich hierzu auch im Deutschen eingebürgert.

Dabei heissen die Zufallsgrössen X_1, \dots, X_n *unabhängig*, wenn für ihre Verteilungsfunktionen für alle Werte x_1, \dots, x_n gilt:

$$P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot P(X_2 \leq x_2) \cdot \dots \cdot P(X_n \leq x_n) .$$

(8.1.5) Nähere Informationen zu den X_i

Wir haben in (8.1.3) und auch schon früher erwähnt, dass jede Zufallsgrösse eigentlich eine Abbildung eines passenden Ergebnisraums in die Menge der reellen Zahlen

ist, dass aber dieser Ergebnisraum meist nicht explizit angegeben wird. Man kann sich nun fragen, auf welchem Ergebnisraum die eben eingeführten Zufallsgrößen X_i definiert sind. Dazu muss man ein klein wenig ausholen. Damit unsere Küken sich etwas erholen können, betrachten wir einmal Hühnereier, die wir auch auf ihr Gewicht untersuchen wollen. Die Population besteht hier aus allen Hühnereiern, die Zufallsgrösse

$$X : \Omega \rightarrow \mathbb{R}$$

stellt das Gewicht dar. Wenn ω ein Ei ist, dann ist $x = X(\omega)$ das Gewicht dieses Eis. In unserer neuen Sprache ist x eine Realisierung von X .

Eine Stichprobe vom Umfang n wird erhalten, indem wir n Eier aus unserer Population Ω auswählen. Ist z.B. $n = 4$, so besteht eine Stichprobe also aus einer "Vier-erliste"

$$(\omega_1, \omega_2, \omega_3, \omega_4)$$

von Eiern. Die Menge aller solcher Viererlisten bezeichnet man mit Ω^4 (in völliger Analogie zu Bezeichnungen wie \mathbb{R}^3 , (22.3) im ersten Band). Entsprechend gehört zu einer Stichprobe vom Umfang n der Ergebnisraum

$$\Omega^n = \{(\omega_1, \dots, \omega_n) \mid \omega_i \in \Omega\} .$$

Die Zufallsgrößen X_i sind nun alle auf Ω^n definiert, und X_i ordnet der Stichprobe $(\omega_1, \dots, \omega_i, \dots, \omega_n)$ den Wert des i -ten Merkmals, also z.B. das Gewicht des i -ten Eis zu. In Formeln

$$\begin{aligned} X_1 : \Omega^n &\rightarrow \mathbb{R}, & (\omega_1, \dots, \omega_n) &\mapsto x_1 = \text{Gewicht von } \omega_1 \\ & & &\vdots \\ X_n : \Omega^n &\rightarrow \mathbb{R}, & (\omega_1, \dots, \omega_n) &\mapsto x_n = \text{Gewicht von } \omega_n . \end{aligned}$$

Wir treiben diesen Formalismus noch etwas weiter, mit der Absicht, eine in (8.2.5) einzuführende Begriffsbildung auch abstrakt zu erklären. Wir bilden dazu eine neue Zufallsgrösse, die wir mit \bar{X} bezeichnen:

$$\bar{X} = \frac{1}{n}(X_1 + \dots + X_n) .$$

Da jede der Zufallsgrößen X_i auf Ω^n definiert ist, trifft dies auch für \bar{X} zu. Wir berechnen jetzt (ganz stur nach den Formeln) ihren Wert auf einem Element

$$(\omega_1, \dots, \omega_n) \in \Omega^n .$$

Nach Bemerkung 7) von (4.1.3) ist

$$\begin{aligned}\bar{X}(\omega_1, \dots, \omega_n) &= \frac{1}{n}(X_1 + \dots + X_n)(\omega_1, \dots, \omega_n) \\ &= \frac{1}{n}\left(X_1(\omega_1, \dots, \omega_n) + X_2(\omega_1, \dots, \omega_n) + \dots + X_n(\omega_1, \dots, \omega_n)\right) \\ &= \frac{1}{n}(x_1 + x_2 + \dots + x_n) \\ &= \bar{x}.\end{aligned}$$

(Bei der zweitletzten Gleichheit wurde die Beziehung $X_i(\omega_1, \dots, \omega_i, \dots, \omega_n) = x_i$ benutzt.) In Worten ausgedrückt: $\bar{X}(\omega_1, \dots, \omega_n)$ ist der Durchschnitt (das arithmetische Mittel) der Realisierungen x_1, \dots, x_n der Zufallsgrößen X_1, \dots, X_n , was natürlich die Bezeichnung rechtfertigt.

8.2. SCHÄTZEN VON PARAMETERN

(8.2.1) Überblick

Wenn man aufgrund einer Stichprobe auf die Grundgesamtheit schliessen will, dann wird man sich unter anderm für den Erwartungswert und die Varianz (oder die Standardabweichung) interessieren. Da diese Grössen nicht genau bekannt sind, wird man sie durch geeignete Werte *schätzen*, und zwar schätzt man

- μ durch \bar{x} , (8.2.3)

- σ^2 durch s^2 . (8.2.4)

Man spricht in solchen Fällen von einer *Punktschätzung*. Daneben gibt es noch so genannte *Intervallschätzungen*. Man gibt hier ein Intervall an, das *Vertrauensintervall* oder *Konfidenzintervall*, von dem man mit einer bestimmten vorgegebenen Sicherheit sagen kann, dass es einen gesuchten, aber unbekanntem Parameter enthält. Wir werden hier Vertrauensintervalle für den Erwartungswert und für die Wahrscheinlichkeit eines Ereignisses angeben. (8.2.9)

Im ersten Fall spielt eine gewisse stetige Verteilung, die so genannte *t-Verteilung*, eine wichtige Rolle. Diese wird auch im Zusammenhang mit statistischen Tests von Bedeutung sein. (8.2.7)
Kap. 9

Ferner wird der *Standardfehler* eingeführt, der von nun an sehr häufig gebraucht wird. Dabei handelt es sich einfach um die Standardabweichung des Durchschnitts \bar{X} . (8.2.5)

(8.2.2) Worum geht es bei der Parameterschätzung?

Wie wir in Kapitel 8.1 (speziell am Schluss von (8.1.3)) gesehen haben, besteht die Hauptaufgabe der Statistik darin, aus einer Stichprobe Rückschlüsse auf die Grundgesamtheit zu ziehen, oder — etwas gelehrter formuliert — Informationen über die zur Grundgesamtheit gehörende Zufallsgrösse X zu erlangen.

In diesem Zusammenhang ist sicherlich der Erwartungswert

$$\mu = E(X)$$

speziell interessant, denn dieser entspricht aufgrund seiner üblichen Interpretation dem “Durchschnittswert” der Zufallsvariablen X , konkret etwa:

In 8.1.2.A: X = Gewicht von Küken,
 μ = Durchschnittsgewicht aller Küken.

In 8.1.2.D: X = Anzahl Zerfälle pro Zeiteinheit,
 μ = durchschnittliche Anzahl Zerfälle pro Zeiteinheit.

Zwar ist μ nicht genau bekannt, doch lässt sich diese Zahl aufgrund der Stichprobe auf einfache Weise schätzen, wie wir in (8.2.3) sehen werden. Da der Erwartungswert $\mu = E(X)$ ein so genannter *Parameter* der Zufallsgrösse X ist, spricht man hier auch von einer *Parameterschätzung*.

Ebenfalls von Bedeutung ist ein weiterer Parameter, die Varianz

$$\sigma^2 = V(X),$$

für die wir in (8.2.4) ein Schätzverfahren kennen lernen werden.

Wenn, wie in den beiden angegebenen Fällen, ein Parameter durch eine bestimmte Zahl geschätzt wird, dann spricht man auch von einer *Punktschätzung*. Daneben werden wir in (8.4.1) und (8.4.2) noch einen zweiten Typ einer Schätzung antreffen, nämlich die *Intervallschätzung*. Hier geht es nicht mehr darum, den Parameter durch eine einzelne Zahl zu schätzen, vielmehr hätte man gerne ein Intervall, von dem man mit einer bestimmten Sicherheit sagen kann, dass es den fraglichen Parameter (z.B. μ) enthält. Man gibt also sozusagen eine vernünftige “Bandbreite” für den Parameter an.

In den nächsten Abschnitten gehen wir zu den Einzelheiten über.

(8.2.3) Schätzung des Erwartungswerts

Wir nehmen an, es sei eine Grundgesamtheit gegeben, welche im Sinn von (8.1.3) durch eine Zufallsgrösse X beschrieben wird. Ferner sei eine Stichprobe

$$x_1, x_2, \dots, x_n$$

vom Umfang n ermittelt worden.

Wir möchten nun den Erwartungswert $\mu = E(X)$ schätzen. Da dieser anschaulich dem Durchschnitt entspricht, liegt die folgende Festsetzung nahe:

Als Schätzung für den Erwartungswert $\mu = E(X)$ der Grundgesamtheit verwendet man den Durchschnitt \bar{x} der Stichprobe, also die Zahl

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i .$$

Diese Festsetzung ist zugegebenermassen völlig unspektakulär, denn sie ist wirklich einleuchtend. Das darf Sie aber nicht stören. Immerhin werden wir in (8.2.6) noch eine mathematische Rechtfertigung finden.

Auch die folgenden Zahlenbeispiele sind entsprechend banal. Da wir die Zahlwerte später noch brauchen werden, geben wir sie trotzdem an:

- 1) Zu Beispiel 8.1.2.A: In (2.2.3.3.b) haben wir das Durchschnittsgewicht der 50 Küken aus (2.2.2.3) angegeben: Die Zahl $\bar{x} = 102.96$ wird daher als Schätzung für das unbekannte μ verwendet.

Vielleicht schadet es aber nichts, nochmals den Unterschied zwischen \bar{x} und μ zu betonen:

- μ ist das Durchschnittsgewicht der Grundgesamtheit, also *aller* zweiwöchigen Küken. Diese Zahl ist unbekannt.
- \bar{x} ist der Mittelwert einer einzelnen Stichprobe von 50 Tieren. Eine andere Stichprobe (von gleichem oder verschiedenem Umfang) wird ein anderes \bar{x} ergeben (vgl. (8.2.5)). Man erwartet aber, dass alle so erhaltenen \bar{x} in der Nähe des unbekanntes Parameters μ liegen und deshalb als Schätzung verwendet werden können.

- 2) Zu Beispiel 8.1.2.D (vgl. die dortige Tabelle): 326 Minuten ergeben 2608 Intervalle zu 7.5 Sekunden (dies ist natürlich auch das Total der zweiten Zeile). Die gesamte Zahl der Zerfälle ist

$$57 \cdot 0 + 203 \cdot 1 + 383 \cdot 2 + \dots + 1 \cdot 14 = 10097 .$$

Daraus ergibt sich das arithmetische Mittel (auf 3 Stellen gerundet)

$$\bar{x} = \frac{10097}{2608} = 3.872 .$$

Dies ist also ein Schätzwert für die mittlere Anzahl der Zerfälle pro Zeitintervall von Polonium, und zwar nicht für den untersuchten Zeitraum von 326 Minuten, sondern generell.

Schon in (2.2.3.3) wurde darauf hingewiesen, dass der Durchschnitt \bar{x} mit den meisten Taschenrechnern direkt berechnet werden kann.

(8.2.4) Schätzung von Varianz und Standardabweichung

Wir gehen von derselben Situation wie in (8.2.3) aus und wollen nun die (wahrscheinlichkeitstheoretische) Varianz schätzen. Hierzu verwendet man — was niemanden überraschen wird — die in (2.2.3.7) definierte statistische (oder empirische) Varianz:

Als Schätzung für die Varianz $\sigma^2 = V(X)$ der Grundgesamtheit (präziser wahrscheinlichkeitstheoretische Varianz genannt) verwendet man die Varianz s^2 der Stichprobe (die statistische oder empirische Varianz), also die Zahl

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 .$$

Entsprechend wird die wahrscheinlichkeitstheoretische Standardabweichung σ durch die statistische (oder empirische) Standardabweichung

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

geschätzt.

Die in (8.2.3) gemachten Bemerkungen gelten mutatis mutandis auch hier. Für den späteren Gebrauch zitieren wir ein früheres Zahlenbeispiel (2.2.3.7.c): Bei den Küken aus (2.2.2.3) ist $s^2 = 40.5289$ und $s = 6.3662$.

Bei der Verwendung eines Taschenrechners ist, wie schon in (2.2.3.7.e) erwähnt, darauf zu achten, ob er die Standardabweichung mit dem Nenner $n - 1$ (wie er hier verwendet wird) oder mit dem Nenner n liefert.

(8.2.5) Wir fassen den Durchschnitt als Zufallsgrösse auf

Wir betrachten wie schon in den vorhergehenden Abschnitten eine Stichprobe

$$x_1, x_2, \dots, x_n$$

aus einer durch die Zufallsgrösse X beschriebenen Grundgesamtheit und berechnen das arithmetische Mittel

$$\bar{x} = \frac{1}{n} (x_1 + \dots + x_n) .$$

Dies kann für jede Stichprobe vom (als fest angenommenem) Umfang n durchgeführt werden. Natürlich wird \bar{x} — wie schon im Beispiel 1) von (8.2.3) erwähnt — für jede Stichprobe ein wenig anders ausfallen. Die Zahl \bar{x} kann daher als Realisierung einer neuen Zufallsgrösse \bar{X} aufgefasst werden.

Der obigen Formel für \bar{x} (Realisierung) entspricht die folgende Formel für \bar{X} (Zufallsgrösse); vgl. (8.1.5) für eine theoretische Erläuterung:

$$\bar{X} = \frac{1}{n} (X_1 + \dots + X_n) ,$$

wobei die Zufallsgrösse X_i wie in (8.1.4) als i -te Messung zu verstehen ist. (Der Umfang n der Stichprobe ist allerdings aus dem Symbol \bar{X} nicht direkt ersichtlich.)

Ein Hinweis zur Bezeichnung: Wenn immer möglich, werden wir die Zufallsgrössen mit grossen, die zugehörigen Realisierungen mit den entsprechenden kleinen Buchstaben bezeichnen. (In vielen Büchern wird hier allerdings nicht so genau unterschieden.)

Die soeben eingeführte Zufallsgrösse \bar{X} , also der Durchschnitt einer Stichprobe vom (festen) Umfang n , besitzt einen Erwartungswert $E(\bar{X})$ und eine Varianz $V(\bar{X})$, die man ausrechnen kann, wenn man den Erwartungswert $E(X) = \mu$ und die Varianz $V(X) = \sigma^2$ der durch X beschriebenen Grundgesamtheit kennt. Es gelten die folgenden Formeln, die wir ohne Beweis angeben:

$$(1) \quad E(\bar{X}) = E(X) = \mu, \quad (2) \quad V(\bar{X}) = \frac{V(X)}{n} = \frac{\sigma^2}{n}.$$

An dieser Stelle führen wir noch einen wichtigen Begriff ein, der auch in der Fehlerrechnung vorkommt, nämlich den *Standardfehler*. Dies ist einfach die Standardabweichung von \bar{X} , also die Wurzel aus der Varianz $V(\bar{X})$. Als Bezeichnung verwendet man das Symbol $\sigma_{\bar{x}}$. In Formeln also

$$(3) \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

Beachten Sie, dass der Standardfehler vom Umfang n der Stichprobe abhängt, obwohl dies im Zeichen $\sigma_{\bar{x}}$ nicht zum Ausdruck kommt.

Bei einer gegebenen Stichprobe wird gemäss (8.2.4) σ durch s geschätzt. Man kommt so auf folgendes:

Als Schätzung für den Standardfehler $\sigma_{\bar{x}}$ verwendet man die Zahl

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}}.$$

Beispiel

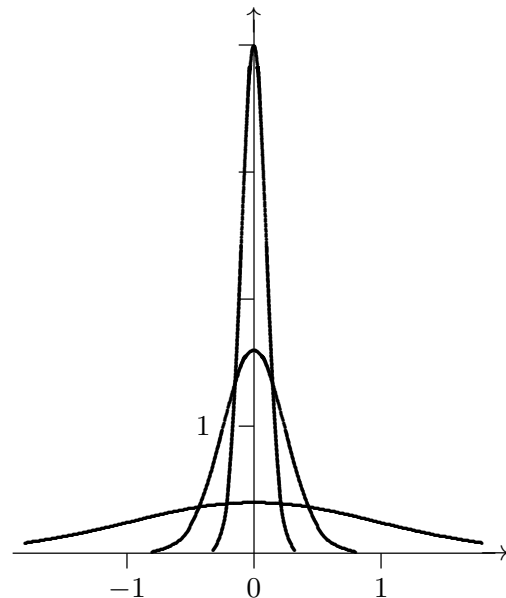
Im Fall der Küken ist die Standardabweichung $s = 6.36623$, ferner ist $n = 50$. Somit wird der Standardfehler geschätzt durch

$$s_{\bar{x}} = \frac{6.36623}{\sqrt{50}} = 0.9003. \quad \boxtimes$$

Wir besprechen nun die anschauliche Bedeutung der Formeln (1), (2) und (3). Für die Formel (1) betrachten wir wieder unsere Küken. Der Erwartungswert $\mu = E(X)$ (also anschaulich das Durchschnittsgewicht *aller* denkbaren zweiwöchigen Küken) ist nicht bekannt. Aufgrund unserer Stichprobe haben wir aber μ durch $\bar{x} = 102.96$ geschätzt. Diese Zahl ist eine Realisierung der Zufallsvariablen \bar{X} . Eine andere Stichprobe (aber vom *gleichen* Umfang 50) hätte einen etwas anderen Durchschnitt \bar{x} ergeben (z.B. $\bar{x} = 103.33$). Der Erwartungswert $E(\bar{X})$ der Zufallsgrösse \bar{X} ist nun anschaulich der “Durchschnitt aller dieser Durchschnitte”. So gesehen, überrascht es kaum, dass $E(\bar{X}) = E(X) = \mu$ herauskommt.

Statt Formel (2) betrachten wir die dazu gleichwertige Formel (3) für die Standardabweichung $\sigma_{\bar{x}}$ von \bar{X} . Die Standardabweichung von X ist gleich σ ; jene von \bar{X} ist gleich $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ und wird umso kleiner, je grösser n ist. Dieser Sachverhalt hat ebenfalls seinen guten Sinn: Die verschiedenen (jeweils aus einer Stichprobe vom Umfang n berechneten) Durchschnitte streuen um ihren Erwartungswert, der durch $E(\bar{X}) = E(X) = \mu$ gegeben ist. Die Standardabweichung $\sigma_{\bar{x}}$ von \bar{X} ist nun gerade ein Mass für die Grösse der Streuung. Nun leuchtet es doch ein, dass ein aus vielen Werten berechneter Durchschnitt im Allgemeinen dem wahren Wert näher kommen wird, als ein aus wenigen Werten berechnetes Mittel. Dies heisst übersetzt, dass die Standardabweichung von \bar{X} (also $\sigma_{\bar{x}}$) umso kleiner wird, je grösser n ist. Die Formel (3) gibt nun eine exakte, quantitative Beschreibung dieses Sachverhalts. Beachten Sie, dass im Nenner von (3) die Wurzel aus n steht. Will man also die “Genauigkeit” verdoppeln — präziser gesagt den Standardfehler $\sigma_{\bar{x}}$ halbieren — so muss die Anzahl n der Beobachtungswerte vervierfacht werden.

Wir sehen uns das Ganze noch etwas genauer an für den Fall, wo X normal verteilt ist, mit den Parametern μ, σ . Man kann zeigen (der nicht einfache Beweis sei hier unterschlagen), dass in diesem Fall auch \bar{X} normal verteilt ist. Gemäss den obigen Formeln (1) und (3) hat dann die normal verteilte Zufallsgrösse \bar{X} die Parameter μ und σ/\sqrt{n} . Die nebenstehende Figur zeigt sehr deutlich, wie die Dichtefunktion von \bar{X} mit wachsendem n immer schmaler wird. Dies belegt noch einmal, dass \bar{x} umso weniger um μ streut, je grösser n ist.



Gezeichnet sind die Verteilung von X mit $\sigma = 1$ (flachste Kurve) sowie die Verteilungen von \bar{X} mit $n = 16$ und $\sigma_{\bar{x}} = 0.25$ bzw. mit $n = 100$ und $\sigma_{\bar{x}} = 0.1$ (steilste

Kurve).

Zu dieser Figur geben wir noch ein

Zahlenbeispiel

Wir nehmen an, die Zufallsgrösse X sei gemäss $N(0; 1)$ normal verteilt (Standard-Normalverteilung). Die Wahrscheinlichkeit dafür, dass ein Wert von X (also ein Wert einer einzelnen Messung) im Intervall $[-0.1, 0.1]$ liegt, ist dann gemäss (5.10.4) gegeben durch

$$\Phi(0.1) - \Phi(-0.1) = 0.5398 - 0.4602 = 0.0796 .$$

Führen wir nun aber 100 Messungen durch ($n = 100$), so ist $E(\bar{X}) = \mu = 0$, $\sigma_{\bar{x}} = \sigma/\sqrt{100} = \sigma/10 = 0.1$. Die Wahrscheinlichkeit dafür, dass der Durchschnitt \bar{x} aus 100 Messungen im erwähnten Intervall $[-0.1, 0.1]$ liegt, ist also gleich

$$\begin{aligned} \Phi_{0,0.1}(0.1) - \Phi_{0,0.1}(-0.1) &= \Phi\left(\frac{0.1 - 0}{0.1}\right) - \Phi\left(\frac{-0.1 - 0}{0.1}\right) \\ &= \Phi(1) - \Phi(-1) = 0.8413 - 0.1587 = 0.6826 \end{aligned}$$

und ist somit wesentlich grösser als im Fall der Einzelmessung. ☒

(8.2.6) Erwartungstreue Schätzungen

In diesem Abschnitt, der für das weitere Verständnis nicht notwendig ist, wird der im Titel erwähnte Begriff erklärt, und es wird erläutert, warum in der Formel für die statistische (oder empirische) Varianz

$$(A) \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

im Nenner $n-1$ und nicht n steht. (Es gibt allerdings auch Bücher, wo der Nenner n gewählt wird.)

Die Formel (1) von (8.2.5),

$$E(\bar{X}) = \mu ,$$

liefert ein mathematisches Argument für die Wahl von \bar{x} als Schätzung für μ . Die Formel besagt ja einfach, dass der Erwartungswert der Schätzung gleich dem gesuchten Parameter ist. Führt man also sehr viele derartige Schätzungen aus, so wird man im Mittel gerade μ erhalten. Eine Schätzung mit dieser Eigenschaft heisst *erwartungstreu*. Also: \bar{x} ist eine erwartungstreue Schätzung für μ .

Wie steht es nun mit der Schätzung s^2 für die Varianz σ^2 ? Der Zufallsgrösse \bar{X} , die soeben besprochen wurde, entspricht hier nun eine neue Zufallsgrösse S^2 , die gegeben ist durch die Formel

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

Die durch (A) gegebene Varianz s^2 ist eine Realisierung von S^2 . (Beachten Sie, dass sich die beiden Formeln völlig entsprechen: In (A) stehen Zahlen, in (B) die zugehörigen Zufallsgrössen.)

Man kann nun zeigen, dass die Schätzung von σ^2 durch s^2 ebenfalls erwartungstreu ist. Es gilt nämlich

$$E(S^2) = \sigma^2 ,$$

wie ohne Beweis erwähnt sei.

Zum Schluss dieses Abschnitts wollen wir wie angekündigt noch begründen, warum in der Definition von s^2 im Nenner die Zahl $n - 1$ steht (vgl. (2.2.3.7)). Nimmt man nämlich statt dessen n und setzt

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{n-1}{n} s^2$$

(die neue Schriftart für “ s ” [s für die Realisierung bzw. S für die Zufallsgrösse] wird für die Formeln mit dem Nenner n verwendet), so erhält man (unter Verwendung von (5.1), Bemerkung 5)) für den Erwartungswert dieser “neuen Varianz”

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

Die mit n statt $n - 1$ definierte Varianz ist also *nicht* erwartungstreu. Dies rechtfertigt die Wahl des Nenners $n - 1$. Allerdings ist der Unterschied bei grösseren Werten von n nurmehr gering.

Um keine falschen Vorstellungen aufkommen zu lassen, sei noch ohne Beweis erwähnt, dass die Schätzung der Standardabweichung σ durch s weder mit dem Nenner $n - 1$ noch mit dem Nenner n erwartungstreu ist.

8.3. DIE T -VERTEILUNG

Das wichtigste Beispiel einer stetigen Verteilung, welches wir bisher kennen gelernt haben, ist die Normalverteilung (5.10.2). Nun besprechen wir eine weitere Verteilung, deren praktische Bedeutung sich in den nächsten Abschnitten und Kapiteln erweisen wird. Es handelt sich dabei um die so genannte *t-Verteilung*, welche auch manchmal *Student-Verteilung* genannt wird, denn ihr Entdecker, der Engländer W.S. GOSSET (1876–1937), veröffentlichte seine Arbeit unter dem Pseudonym “Student”.

Um die *t-Verteilung* zu erhalten, gehen wir von einer Zufallsgrösse X aus, welche der Normalverteilung $N(\mu; \sigma^2)$ folgt. Der Bezug zur Praxis wird dadurch hergestellt, dass dort in vielen Fällen angenommen werden darf, die Grundgesamtheit sei (zumindest annähernd, vgl. dazu (7.2)) normal verteilt. Dieser Grundgesamtheit entnehmen wir eine Stichprobe vom Umfang n :

$$x_1, x_2, \dots, x_n .$$

Ihr Durchschnitt \bar{x} ist eine Realisierung der Zufallsgrösse \bar{X} (vgl. (8.2.5)).

Wir haben gegen das Ende von (8.2.5) ohne Beweis erwähnt, dass mit X auch die Zufallsgrösse \bar{X} normal verteilt ist. Diese Normalverteilung hat als Parameter den Erwartungswert $E(\bar{X}) = \mu$ und die Standardabweichung $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, wie wir in (8.2.5) (Formeln (1) und (3)) gesehen haben.

Wie in (5.10.3) standardisieren wir die normal verteilte Zufallsgrösse \bar{X} , indem wir

$$Y = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}}$$

setzen. Die Zufallsgrösse Y ist Dank der Z-Transformation (5.10.3) wieder normal verteilt. Beachten Sie, dass im Nenner dieses Ausdrucks eine *Zahl*, nämlich $\sigma_{\bar{x}}$ steht. Allerdings ist die Grösse $\sigma_{\bar{x}}$ sehr oft nicht genau bekannt, sondern muss durch

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}}$$

geschätzt werden. Dabei ist $s_{\bar{x}}$ die Realisierung der neuen Zufallsgrösse

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n(n-1)}}.$$

Diese ordnet jeder Stichprobe vom Umfang n ihren Standardfehler zu, der natürlich genauso wie der Durchschnitt von der Stichprobe abhängt und deshalb eben eine Zufallsgrösse ist. Ersetzt man nun in der Formel für Y die *Zahl* $\sigma_{\bar{x}}$ durch die *Zufallsgrösse* $S_{\bar{x}}$, so erhält man die neue Zufallsvariable

$$T = \frac{\bar{X} - \mu}{S_{\bar{x}}}.$$

Ihre Werte (d.h., die Realisierungen) erhält man durch Einsetzen der aus der Stichprobe

$$x_1, x_2, \dots, x_n$$

berechneten Grössen \bar{x} und $s_{\bar{x}}$ mit der Formel

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}.$$

Dabei wird die *Zahl* μ in diesem Zusammenhang jeweils als gegeben angenommen, was in den Anwendungen noch klarer zum Ausdruck kommen wird. (Im Gegensatz zu $\sigma_{\bar{x}}$ wird also μ nicht durch eine Zufallsgrösse ersetzt.)

Beachten Sie auch, dass (via $S_{\bar{x}}$ und \bar{x} selber) der Umfang n der Stichprobe in der Formel für T steckt. Trotzdem schreibt man nicht T_n oder so etwas, sondern gibt die Zahl n falls nötig separat an.

Diese Zufallsgrösse T ist nun nicht mehr normal verteilt. Dies kommt daher, dass, wie bereits erwähnt, im Gegensatz zu Y im Nenner nicht mehr eine Zahl, sondern eine Zufallsgrösse steht.

Es stellt sich daher die Aufgabe, die Verteilung von T zu bestimmen. Dies kann man z.B. dadurch erledigen, dass man versucht, die Dichtefunktion zu finden. Gemäss (4.3.3) ist also eine Funktion f mit den dort angegebenen Eigenschaften 1), 2) und 3) gesucht, so dass für alle $t \in \mathbb{R}$

$$P(T \leq t) = \int_{-\infty}^t f(x) dx ,$$

gilt, wobei — wie eingangs erwähnt wurde — die dem Ganzen zugrunde liegende Zufallsgrösse X normal verteilt sein muss. Diese Aufgabe ist alles andere als einfach; wir wollen uns daher darauf beschränken, das Ergebnis anzuführen, das wie folgt lautet:

Die oben definierte Zufallsgrösse T hat die Dichtefunktion

$$(*) \quad f(x) = c_n \left(1 + \frac{x^2}{n-1} \right)^{-\frac{n}{2}} .$$

In dieser Formel kommt die Zahl n vor, die bei unserer Herleitung dem Umfang der Stichprobe entsprach. Wenn man diese Formel mit der Dichte in (6.2.6) vergleicht (ignorieren Sie dazu besser die Konstante c_n), stellt man fest, dass obige Dichte offenbar zu einer t -Verteilung mit Freiheitsgrad $n-1$ (!) gehört. Das ist zwar irritierend, wird aber im Folgenden zu keinen Problemen führen, denn es gilt: wann immer wir eine *theoretische Bemerkung zu einer t -Verteilung mit Freiheitsgrad n* machen, dann orientiert man sich an der t -Verteilung aus Kapitel 6. In den Anwendungen wird man aber in Kapitel 8, 9 und 10 Parameter schätzen, zum Beispiel σ^2, σ , welche in die zugrundeliegende Teststatistik einfließen. Man kann zeigen, dass damit ein Freiheitsgrad “verloren” geht. In Kapitel 8 und 9 ist das jeweils 1 Freiheitsgrad, in Kapitel 10 sogar deren 2 (bei der späteren χ^2 -Verteilung haben wir ganz analoge Umstände, **siehe (9.x.x)**). Wir werden dies aber immer thematisieren.

Damit können wir die Verteilung der Zufallsgrösse T genau angeben:

Die Zufallsgrösse T folgt der t -Verteilung mit dem Freiheitsgrad $n-1$, wenn T die in der Formel (*) angegebene Dichtefunktion hat.

Bemerkungen

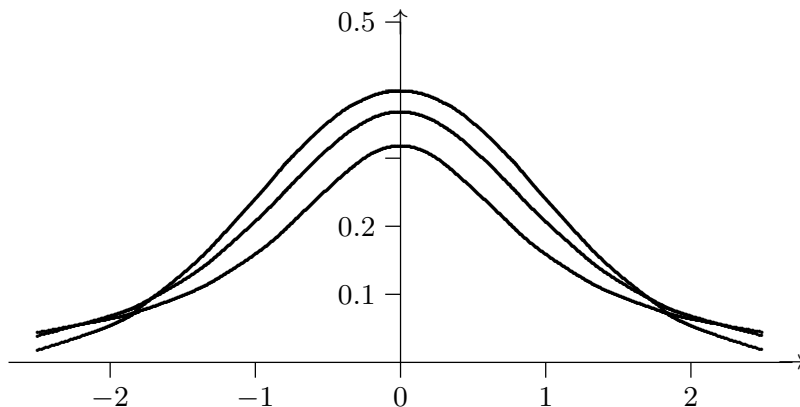
1. Der genaue Wert der Konstanten c_n soll hier nicht angegeben werden (siehe aber auch (6.2.6)). Er wird so festgelegt, dass

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

ist (Bedingung 3) von (4.3.3)).

2. Der Graph der Dichtefunktion f ähnelt jenem der Standard-Normalverteilung. Er ist symmetrisch zur y -Achse, aber etwas niedriger und breiter als jener der Normalverteilung. Je grösser der Freiheitsgrad n ist, desto mehr nähert sich die t -Verteilung der Standard-Normalverteilung. Die oberste Kurve gehört zur Standard-Normalverteilung, die mittlere zur t -Verteilung mit dem Freiheitsgrad $n=3$ und

die unterste zur t -Verteilung mit $n = 1$ (Freiheitsgrad wie in Kapitel 6 angegeben).



Zur Tabellierung der t -Verteilung

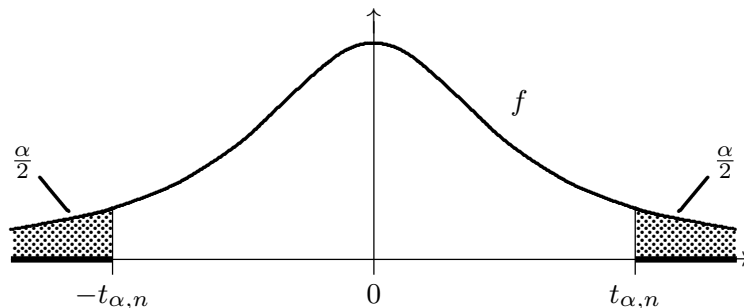
Genau wie für die Standard-Normalverteilung könnte man auch die Werte der Verteilungsfunktion der t -Verteilung tabellieren, wobei aber für jeden Freiheitsgrad n eine eigene Tabelle erforderlich wäre. Nun ist es aber so, dass man sich in den statistischen Anwendungen nicht so sehr für die Werte der Verteilungsfunktion als für die *kritischen Werte* $t_{\alpha,n}$ interessiert*. Wir haben diesen Begriff bereits in (5.10.5) für die Normalverteilung kennen gelernt. Die Situation ist hier ganz analog, soll aber trotzdem nochmals erläutert werden.

Gegeben ist also eine Zufallsgrösse T , welche eine t -Verteilung mit dem Freiheitsgrad n hat. Weiter sei eine Zahl α mit $0 < \alpha < 1$ gegeben. Unter dem zugehörigen *kritischen Wert* $t_{\alpha,n}$ versteht man die Zahl, für welche

$$P(|T| \geq t_{\alpha,n}) = \alpha$$

gilt. Oft schreibt man auch bloss t_α , da für eine gegebene Aufgabe der Freiheitsgrad n im vornherein feststeht.

Geometrisch sieht die Sache so aus (vgl. auch (5.10.5)): Die Werte, welche dem Betrage nach grösser als oder gleich $t_{\alpha,n}$ sind, entsprechen dem dick ausgezogenen Teil der horizontalen Achse:



* Ähnlich wie bei den kritischen Werten der Normalverteilung sind auch hier noch andere Bezeichnungen wie etwa $t_{1-\frac{\alpha}{2}}$ oder $t_{1-\frac{\alpha}{2}}(n)$ gebräuchlich.

Der markierte Bereich entspricht der gegebenen Wahrscheinlichkeit α ; aus Symmetriegründen hat dann jedes der beiden Flächenstücke den Inhalt $\alpha/2$. In Formeln ausgedrückt:

$$\int_{-\infty}^{-t_{\alpha,n}} f(x) dx = \frac{\alpha}{2}, \quad \int_{t_{\alpha,n}}^{\infty} f(x) dx = \frac{\alpha}{2}.$$

Diese kritischen Werte $t_{\alpha,n}$ findet man nun in Tabellen aufgeführt und zwar in Abhängigkeit von n und α , vgl. Tabelle (6.2.6).

Ein Zahlenbeispiel

Es liege eine t -Verteilung mit Freiheitsgrad 7 vor. Ferner sei $\alpha = 0.1$ gewählt. Der Tabelle (6.2.6) entnimmt man für $n = 7$ und $\alpha = 0.1$, dass

$$t_{\alpha,n} = 1.895$$

ist. Dies bedeutet, dass $P(|T| \geq 1.895) = 0.1$ ist. “Einseitig” gesehen ist entsprechend $P(T \leq -1.895) = 0.05$ und $P(T \geq 1.895) = 0.05$. \square

Im nächsten Abschnitt kommen wir nun zu einer ersten Anwendung der t -Verteilung. Diese Verteilung wird auch in den Kapiteln 9 und 10 von Bedeutung sein.

8.4. KONFIDENZINTERVALLE

(8.4.1) Das Konfidenzintervall für den Erwartungswert

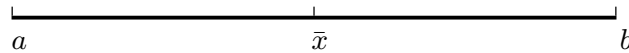
In (8.2.3) haben wir gesehen, dass der Erwartungswert μ der Grundgesamtheit durch den Durchschnitt \bar{x} der Stichprobe geschätzt wird. Es ist nützlich, an dieser Stelle nochmals den Unterschied zwischen μ und \bar{x} zu erläutern. Im Fall der Küken entspricht μ dem naturgemäss unbekanntem, aber im Sinne einer Idealisierung denkbaren Durchschnittsgewicht aller zweiwöchigen Küken. Die Zahl \bar{x} dagegen ist aufgrund einer Stichprobe konkret ermittelt worden. Da die Stichprobe nicht alle Küken umfasst, ist die Schätzung von μ durch \bar{x} nicht exakt, sondern eben nur ein approximativer Wert. Zudem wird \bar{x} mit jeder Stichprobe etwas anders ausfallen.

Trotzdem möchte man natürlich gerne wissen, wie zuverlässig diese Schätzung ist. Dies kann man dadurch erreichen, dass man eine Art “Bandbreite” angibt, in welcher der unbekannte Wert μ mit einer gewissen vorher festgelegten Wahrscheinlichkeit liegt*. Diese Wahrscheinlichkeit nennt man *Vertrauenswahrscheinlichkeit* oder auch *Konfidenzwahrscheinlichkeit* und bezeichnet sie meist mit Q . Eine mögliche (und übliche) Wahl ist $Q = 0.95$ (oder 95%).

* Die genaue Bedeutung dieser Formulierung wird weiter unten diskutiert.

Die oben erwähnte “Bandbreite” wird mathematisch durch ein abgeschlossenes Intervall $[a, b]$ gegeben. Da wir ja nicht wissen, ob unser Stichprobendurchschnitt \bar{x} grösser oder kleiner als μ ist, ist es sinnvoll, das Intervall so festzulegen, dass \bar{x} sein Mittelpunkt ist. Damit können wir unser Problem zusammenfassen:

Gegeben ist \bar{x} sowie die Vertrauenswahrscheinlichkeit Q . Gesucht ist ein Intervall $I = [a, b]$ mit Mittelpunkt \bar{x} , so dass gilt: Die Wahrscheinlichkeit dafür, dass μ in I liegt*, ist gleich Q .



Es ist von vornherein schon ohne Rechnung klar, dass bei einer gegebenen Stichprobe das Intervall umso grösser sein wird, je grösser Q gewählt wird.

Wir wollen nun eine Formel für das Intervall $I = [a, b]$ herleiten. Dazu setzen wir voraus, dass die zur Grundgesamtheit gehörende Zufallsgrösse X normal verteilt ist. Kleine Abweichungen von der Normalität, wie sie in der Praxis immer wieder vorkommen, bewirken aber keinen wesentlichen Fehler.

Ausgehend von der Stichprobe x_1, \dots, x_n schätzen wir die unbekannt Parameter μ bzw. $\sigma_{\bar{x}}$ wie gewohnt (vgl. (8.2.3), (8.2.5)) durch \bar{x} und $s_{\bar{x}}$. Nun haben wir in (8.3) gelernt, dass die Zahl

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$$

eine Realisierung der Zufallsgrösse

$$T = \frac{\bar{X} - \mu}{S_{\bar{x}}}$$

ist, welche einer t -Verteilung mit Freiheitsgrad $n - 1$ folgt, wobei n der Umfang der Stichprobe ist. Gegeben ist ferner die Vertrauenswahrscheinlichkeit Q ; zur Vereinfachung setzen wir noch $\alpha = 1 - Q$.

Zu diesen Werten von α und $n - 1$ können wir in der Tabelle den kritischen Wert $t_{\alpha, n-1}$ nachschlagen, den wir hier kurz mit t_{α} bezeichnen wollen. Nach Definition des kritischen Werts ist

$$P(|T| \geq t_{\alpha}) = \alpha .$$

Das Gegenereignis zu $|T| \geq t_{\alpha}$ ist $|T| < t_{\alpha}$ oder auch $-t_{\alpha} < T < t_{\alpha}$. Deshalb ist

$$P(-t_{\alpha} < T < t_{\alpha}) = 1 - \alpha = Q ,$$

aber auch (vgl. (4.3.4.c))

$$P(-t_{\alpha} \leq T \leq t_{\alpha}) = 1 - \alpha = Q .$$

Nun formen wir etwas um. Die Ungleichung $-t_{\alpha} \leq T \leq t_{\alpha}$ ist natürlich gleichbedeutend mit $-t_{\alpha} \leq -T \leq t_{\alpha}$. Setzen wir für $-T$ die oben stehende Definition ein, so folgt der Reihe nach

$$\begin{aligned} -t_{\alpha} &\leq \frac{\mu - \bar{X}}{S_{\bar{x}}} \leq t_{\alpha} \\ -t_{\alpha} S_{\bar{x}} &\leq \mu - \bar{X} \leq t_{\alpha} S_{\bar{x}} && \text{(Multiplikation mit } S_{\bar{x}}) \\ \bar{X} - t_{\alpha} S_{\bar{x}} &\leq \mu \leq \bar{X} + t_{\alpha} S_{\bar{x}} && \text{(Addition von } \bar{X}) . \end{aligned}$$

Da die obigen Schritte auch in umgekehrter Reihenfolge durchlaufen werden können, ist das Ereignis

$$\bar{X} - t_\alpha S_{\bar{x}} \leq \mu \leq \bar{X} + t_\alpha S_{\bar{x}}$$

gleich dem Ereignis

$$-t_\alpha \leq T \leq t_\alpha$$

und hat somit ebenfalls die Wahrscheinlichkeit Q . Mit andern Worten: Es ist

$$(*) \quad P(\bar{X} - t_\alpha S_{\bar{x}} \leq \mu \leq \bar{X} + t_\alpha S_{\bar{x}}) = Q.$$

Damit haben wir ein Intervall gefunden, nämlich

$$[\bar{X} - t_\alpha S_{\bar{x}}, \bar{X} + t_\alpha S_{\bar{x}}],$$

in welchem μ mit der Wahrscheinlichkeit Q liegt. Die Grenzen dieses Intervalls, in denen \bar{X} und $S_{\bar{x}}$ vorkommt, sind selbst wieder Zufallsgrößen.

Für jede effektiv gegebene Stichprobe

$$x_1, x_2, \dots, x_n$$

können wir aber die Werte

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{und} \quad s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}}$$

als Realisierungen der Zufallsgrößen \bar{X} und $S_{\bar{x}}$ berechnen und erhalten so das Intervall

$$[\bar{x} - t_\alpha s_{\bar{x}}, \bar{x} + t_\alpha s_{\bar{x}}].$$

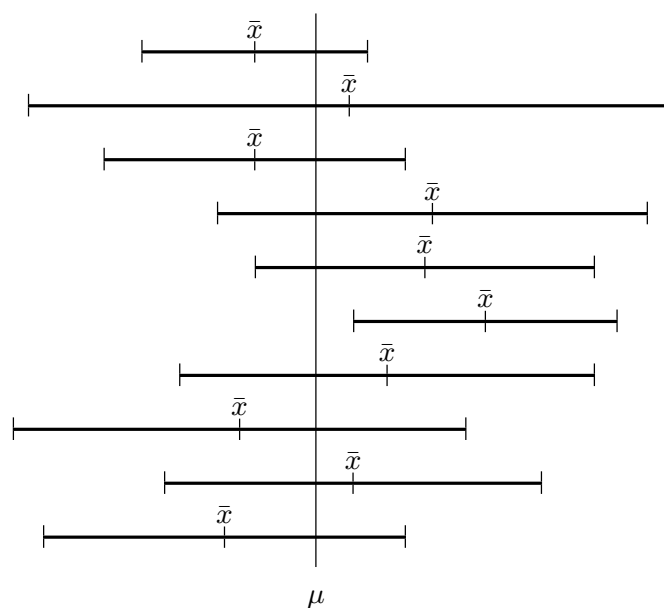
Es heisst *Vertrauensintervall* oder *Konfidenzintervall*.

Man ist nun aufgrund der Herleitung versucht zu sagen, der unbekannte Wert von μ liege mit der Wahrscheinlichkeit Q im Vertrauensintervall. Diese Aussage geht zwar in die richtige Richtung, ist aber, wörtlich genommen, falsch.

Dies lässt sich so begründen: Die Grenzen des aufgrund einer Stichprobe bestimmten Intervalls $[\bar{x} - t_\alpha s_{\bar{x}}, \bar{x} + t_\alpha s_{\bar{x}}]$ sind Zahlen, im Beispiel 8.4.1.A, a) etwa erhalten wir das Intervall $I = [101.15, 104.77]$. Nun gibt es aber für das zahlenmässig festgelegte Intervall I nur zwei Möglichkeiten. Entweder liegt der unbekannte Wert von μ in I oder aber dies ist nicht der Fall; wir wissen allerdings nicht, was zutrifft. Im ersten Fall ist die Wahrscheinlichkeit dafür, dass μ in I liegt gleich 1, im zweiten ist sie gleich 0; jedenfalls ist sie nicht gleich Q . Insofern ist die Aussage “ μ liegt mit der Wahrscheinlichkeit Q in I ” falsch.

Die folgenden Überlegungen sollen aufzeigen, was das Konfidenzintervall wirklich bedeutet. Wenn wir viele Stichproben vom Umfang n bilden, dann werden die Grössen \bar{x} und $s_{\bar{x}}$ jedes Mal etwas anders herauskommen und damit aber auch die Grenzen $\bar{x} - t_{\alpha} s_{\bar{x}}$ und $\bar{x} + t_{\alpha} s_{\bar{x}}$ des Konfidenzintervalls I (es handelt sich ja um Realisierungen von Zufallsgrössen). Auch die Länge des Intervalls (sie beträgt $2t_{\alpha} s_{\bar{x}}$) hängt von der Stichprobe ab. Je nach Lage und Grösse des Intervalls wird nun der Erwartungswert μ in I liegen oder nicht. Wegen der in (*) oben angegebenen Wahrscheinlichkeit wird man aber in einer langen Stichprobenreihe erwarten, dass der Anteil der Fälle, wo μ in I liegt, näherungsweise gleich der Vertrauenswahrscheinlichkeit Q ist.

Wenn also beispielsweise 10 Stichproben vorliegen, und wenn wir die Vertrauenswahrscheinlichkeit $Q = 90\%$ gewählt haben, dann werden wir 10 verschiedene Vertrauensintervalle erhalten, und wir rechnen damit, dass der unbekannte Erwartungswert μ in 9 von diesen Intervallen liegen wird. Die unten stehende schematische Darstellung zeigt dies, wobei wir hier so tun, als ob μ bekannt sei (sonst könnten wir diese Grösse ja gar nicht eintragen).



Wir stellen abschliessend fest, dass zur Bestimmung des Vertrauensintervalls I ein Verfahren verwendet wurde, für das in (beispielsweise) 90% aller Fälle die Behauptung, dass μ in I liegt, richtig ist. Wir haben demnach in Form der Wahrscheinlichkeit $Q = 90\%$ sozusagen ein Mass für die Sicherheit dieser Behauptung oder, noch anders formuliert, wir vertrauen zu 90% darauf, dass μ in I liegt. Die kurze Aussage “ μ liegt mit einer Wahrscheinlichkeit von 90% in I ” ist also in diesem Sinn zu verstehen.

Das folgende anschauliche Beispiel soll, im Sinne einer Analogie, den Sachverhalt noch etwas näher erläutern. Ein Produzent verwendet ein Prüfverfahren, das bewirkt, dass 99% der auf den Markt gelangenden Artikel in Ordnung sind. Wenn ich nun ein solches Produkt gekauft habe, dann bin ich versucht zu sagen: “Dieser Artikel ist mit einer Wahrscheinlichkeit von 99% intakt.” Nun ist aber der Artikel entweder defekt oder in Ordnung; allerdings weiss ich (a priori) nicht, welche der beiden Möglichkeiten zutrifft. Meine Aussage ist also, wörtlich genommen, falsch. Die Wahrscheinlichkeit 99% betrifft nämlich nicht den Artikel als solchen, sondern mein Urteil (bzw. das Prüfverfahren), sie ist ein Mass für mein Vertrauen: Ich rechne damit, dass ich bei 100 Käufen (die ich in der Praxis natürlich nicht durchführe) 99 gute Artikel erhalten würde.

Beispiel 8.4.1.A

Wir belästigen einmal mehr unsere Küken (2.2.2.3). Hier ist $n = 50$, $\bar{x} = 102.96$ und $s_{\bar{x}} = 0.9003$, vgl. (8.2.3) und (8.2.5). Man darf annehmen, dass die Grundgesamtheit ungefähr normal verteilt ist; die Voraussetzung für die obigen Überlegungen sind also erfüllt. Der Freiheitsgrad beträgt $n - 1 = 49$. Dieser Wert ist in unserer Tabelle (6.2.6) nicht enthalten. In a) und b) ermitteln wir ihn deshalb durch Interpolation. (Der Unterschied der so gefundenen Werte für $n - 1 = 49$ zu jenen für $n - 1 = 50$ ist allerdings nur gering.)

- a) Wir wählen $Q = 95\%$, somit wird $\alpha = 0.05$. Dann ist gemäss Tabelle $t_{\alpha,45} = 2.014$, $t_{\alpha,50} = 2.009$. Es folgt $t_{\alpha,49} = t_{\alpha} = 2.010$. Man berechnet sofort

$$\bar{x} - t_{\alpha}s_{\bar{x}} = 101.15, \quad \bar{x} + t_{\alpha}s_{\bar{x}} = 104.77.$$

Das Vertrauensintervall ist also gegeben durch

$$[101.15, 104.77].$$

Die Annahme, dass das Durchschnittsgewicht aller zweiwöchigen Küken in diesem Intervall liegt, wird in 95% aller Fälle richtig sein.

- b) Wählen wir $Q = 99\%$, also $\alpha = 0.01$, so wird $t_{\alpha} = 2.680$ (wie oben interpoliert). Als Vertrauensintervall findet man nun

$$[100.55, 105.37].$$

Es ist also grösser als im Fall a), was einleuchtet, denn grössere Sicherheit muss mit einem grösseren Intervall erkaufte werden, was wir übrigens weiter oben schon einmal bemerkt haben. Umgekehrt ist die Situation im folgenden Fall c):

- c) Es sei $Q = 80\%$, also $\alpha = 0.20$. Hier unterscheidet sich $t_{\alpha} = 1.299$ in einer Darstellung mit 3 Stellen nach dem Komma nicht, ob der Freiheitsgrad 50 (siehe Tabelle) oder 49 ist (Zur Kontrolle, mit R erhalten wir bei Freiheitsgrad 49: $qt(0.9, 49) = 1.299069$). Man erhält das Konfidenzintervall

$$[101.79, 104.13].$$

Zwar ist das Intervall jetzt relativ klein geworden, aber wir können nur noch in 80% aller Fälle erwarten, dass α in dieses Intervall fällt. \square

Zusammenfassung

Gegeben sei eine Stichprobe x_1, \dots, x_n aus einer (annähernd) normal verteilten Grundgesamtheit mit unbekanntem Erwartungswert μ und unbekannter Standardabweichung σ . Ferner sei eine Vertrauenswahrscheinlichkeit Q gewählt worden.

Dann ist das Vertrauensintervall für μ zur Vertrauenswahrscheinlichkeit Q gegeben durch

$$[\bar{x} - t_{\alpha}s_{\bar{x}}, \bar{x} + t_{\alpha}s_{\bar{x}}].$$

Dabei ist

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{und} \quad s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}},$$

und $t_{\alpha} = t_{\alpha, n-1}$ ist der kritische Wert der t -Verteilung (Tabelle 6.2.6) mit Freiheitsgrad $n-1$ und "Signifikanzschwelle" $\alpha = 1 - Q$.

Die Behauptung, dass μ im Vertrauensintervall liegt, wird aufgrund eines Verfahrens aufgestellt, das mit der Wahrscheinlichkeit Q richtige Resultate liefert.

Das Vertrauensintervall lässt sich auch als eine Art Schätzung betrachten, wobei aber im Gegensatz zu (8.2.3), wo eine so genannte Punktschätzung vorliegt, eben ein Intervall betrachtet wird, weshalb man auch von einer Intervallschätzung spricht, vgl. (8.2.2).

Schliesslich sei noch darauf hingewiesen, dass es nicht nur für den Erwartungswert, sondern auch für andere Masszahlen, wie etwa die Varianz, Vertrauensintervalle gibt. Für die entsprechenden Formeln sei aber auf die Literatur verwiesen.

(8.4.2) Das Konfidenzintervall für eine unbekannte Wahrscheinlichkeit

Vertrauensintervalle treten auch auf, wenn es darum geht, eine unbekannte Wahrscheinlichkeit (oder einen prozentualen Anteil) zu schätzen. Zur Illustration betrachten wir die folgende Situation.

Bei einer Umfrage wurden 500 Personen gefragt, ob sie ein bestimmtes Projekt befürworten. Dabei antworteten 300 mit “ja”. Nun interessiert man sich natürlich vor allem für den Anteil der Befürwortenden innerhalb der ganzen Bevölkerung und nicht nur innerhalb der Stichprobe. Wenn wir die Wahrscheinlichkeit dafür, dass eine zufällig ausgewählte Person aus der Gesamtpopulation für das Projekt ist, mit p bezeichnen, dann werden wir aufgrund der Stichprobe p durch den Wert 0.6 schätzen. Wie schon in (8.4.1) möchte man nun eine Aussage über die Zuverlässigkeit der Schätzung haben; dazu konstruiert man ein Vertrauensintervall für p . Dabei geht man wie folgt vor:

Die Zufallsgrösse X bezeichne die Anzahl der befürwortenden Personen aus der Stichprobe. Da die Antwort auf die Befragung nur “ja” oder etwas anderes (“nein” bzw. “keine Meinung”) sein kann und da — so nehmen wir wenigstens an — die Personen unabhängig voneinander antworten, ist X binomial verteilt. Dabei ist der Parameter n gleich dem Umfang der Stichprobe, der Parameter p aber ist unbekannt. (In unserem Fall ist $n = 500$, p ist aber nicht etwa gleich 0.6; vielmehr ist dies bloss eine Schätzung für das unbekannte p .)

Um die Betrachtungen allgemein durchzuführen, gehen wir einfach von einer binomial verteilten Zufallsgrösse X aus, ohne ihr eine konkrete Bedeutung zu geben.

Wie wir aus (5.5) wissen, ist $E(X) = np$ und $V(X) = np(1 - p)$. In (7.2) haben wir gesehen, dass die Zufallsgrösse

$$\frac{X - np}{\sqrt{np(1 - p)}}$$

für grosse n annähernd der Standard-Normalverteilung folgt. Wenn jetzt eine Vertrauenswahrscheinlichkeit Q gegeben ist, dann setzen wir wie in (8.4.1) $\alpha = 1 - Q$. Ist nun z_α der zugehörige kritische Wert der Standard-Normalverteilung, so gilt (nach Definition des kritischen Werts, (5.10.5))

$$P\left(\left|\frac{X - np}{\sqrt{np(1 - p)}}\right| \leq z_\alpha\right) \doteq 1 - \alpha = Q.$$

Wir betrachten das Ereignis

$$(1) \quad \left| \frac{X - np}{\sqrt{np(1-p)}} \right| \leq z_\alpha .$$

Mit einer längeren Rechnung, auf die hier verzichtet sei (man quadriert, bildet eine “quadratische Ungleichung” für p und löst diese), stellt man fest, dass das Ereignis (1) gleichwertig zum folgenden Ereignis ist:

$$(2) \quad \frac{1}{n + z_\alpha^2} \left(X + \frac{z_\alpha^2}{2} - z_\alpha \sqrt{\frac{X(n-X)}{n} + \frac{z_\alpha^2}{4}} \right) \leq p \leq \frac{1}{n + z_\alpha^2} \left(X + \frac{z_\alpha^2}{2} + z_\alpha \sqrt{\frac{X(n-X)}{n} + \frac{z_\alpha^2}{4}} \right) .$$

Dieses hat somit die Wahrscheinlichkeit Q .

Damit haben wir nun ein — allerdings ziemlich kompliziert aussehendes — Konfidenzintervall für p gefunden, dessen Grenzen durch die Ausdrücke in (2) gegeben sind.

Wenn man das zu X gehörende Experiment durchführt, dann erhält man eine Realisierung k der Zufallsgrösse X . (In unserm Beispiel ist $k = 300$.) Damit erhält man als Realisierung des durch (2) beschriebenen “Zufallsintervalls” das Intervall mit den Grenzen

$$(3) \quad \frac{1}{n + z_\alpha^2} \left(k + \frac{z_\alpha^2}{2} - z_\alpha \sqrt{\frac{k(n-k)}{n} + \frac{z_\alpha^2}{4}} \right) \quad \text{und} \quad \frac{1}{n + z_\alpha^2} \left(k + \frac{z_\alpha^2}{2} + z_\alpha \sqrt{\frac{k(n-k)}{n} + \frac{z_\alpha^2}{4}} \right) .$$

Mit den Angaben aus unserm Beispiel können wir nun das Intervall berechnen. Wir wählen dazu $Q = 0.95$, also $\alpha = 0.05$. Gemäss (5.10.5) ist dann $z_\alpha = 1.96$. Einsetzen von $n = 500$ und $k = 300$ ergibt nach einiger Rechnung das Vertrauensintervall

$$[0.556, 0.642] .$$

Wir können also sagen, dass der Anteil der Befürwortenden zwischen 55.6% und 64.2% liegt und dass diese Behauptung mit einem Verfahren zustande gekommen ist, das in 95% aller Fälle eine richtige Antwort liefert.

Die obige Formel (3) ist etwas kompliziert. Sind n , k und $n - k$ gross, so kann die Zahl z_α^2 als Summand gegenüber n , k und $k(n - k)$ vernachlässigt werden. Das Vertrauensintervall hat dann die Grenzen

$$(4) \quad \frac{k}{n} - \frac{z_\alpha}{n} \sqrt{\frac{k(n-k)}{n}} \quad \text{und} \quad \frac{k}{n} + \frac{z_\alpha}{n} \sqrt{\frac{k(n-k)}{n}} .$$

In unserem Beispiel liefert die Formel (4) praktisch dasselbe Intervall, nämlich

$$[0.557, 0.643] .$$

Wir führen unser Beispiel noch etwas weiter. Die “Bandbreite” (Länge des Vertrauensintervalls) von 0.086 (bzw. 8.6%) lässt sich — was unmittelbar einleuchtet — verkleinern, wenn wir den Umfang der Stichprobe erhöhen. Wenn wir uns nun bei einer fest gewählten Vertrauenswahrscheinlichkeit Q die Bandbreite b vorgeben, dann stellt sich die Frage, wie gross denn der Umfang der Stichprobe sein muss, damit das Vertrauensintervall die Breite b hat.

Zur Beantwortung dieser Frage arbeiten wir einfachheitshalber mit der Näherungsformel (4).

Wir müssen n so bestimmen, dass

$$2 \frac{z_\alpha}{n} \sqrt{\frac{k(n-k)}{n}} \leq b$$

ist (links steht die Breite des Konfidenzintervalls gemäss (4)). Quadrieren liefert

$$4 \frac{z_\alpha^2}{n^2} \frac{k(n-k)}{n} \leq b^2.$$

Eine einfache Umformung ergibt

$$4z_\alpha^2 \frac{k}{n} \left(1 - \frac{k}{n}\right) \leq b^2 n \quad \text{oder} \quad n \geq \frac{4z_\alpha^2}{b^2} h(1-h),$$

wobei noch die Abkürzung $h = k/n$ (relative Häufigkeit) verwendet wurde. Nun ist aber $h(1-h)$ stets $\leq 1/4$, denn die Funktion $f(h) = h(1-h)$ nimmt ihr absolutes Maximum an der Stelle $h = 1/2$ an.

Es genügt also,

$$n \geq \frac{4z_\alpha^2}{b^2} \cdot \frac{1}{4} = \frac{z_\alpha^2}{b^2}$$

zu wählen. Arbeiten wir mit der Vertrauenswahrscheinlichkeit $Q = 0.95$, so ist $z_\alpha = 1.96$. Soll die “Bandbreite” z.B. 4% betragen, so ist $b = 0.04$ zu setzen, und wir erhalten $n \geq (1.96/0.04)^2 = 2401$. Es sind also rund 2400 Personen zu befragen, damit die gewünschte Sicherheit zustande kommt.

(8.∞) Aufgaben

8–1 Aus einer grossen Zahl gleichartiger Samenkörner wurden 10 Körner zufällig herausgegriffen und gewogen. Die Gewichte (in mg) betragen

245, 233, 249, 255, 238, 251, 245, 250, 236, 238.

a) Schätzen Sie Erwartungswert und Standardabweichung für die Grundgesamtheit sowie den Standardfehler. b) Wir nehmen an, die Grundgesamtheit sei normal verteilt. Wieviel Prozent der Samenkörner wiegen dann über 250 mg?

8–2 Für die Körpergrösse von 12 zehnjährigen Knaben erhielt man folgende Werte (in cm):

141, 142, 140, 145, 135, 138, 144, 143, 150, 148, 134, 144.

a) Schätzen Sie den Erwartungswert und die Standardabweichung für die Grundgesamtheit sowie den Standardfehler. b) Wir nehmen an, die Grundgesamtheit sei normal verteilt. Wieviel Prozent aller zehnjährigen Knaben sind dann grösser als 145 cm?

8–3 Eine Untersuchung über das Gewicht einer Singvogelart ergab folgende Daten:

Gewicht in g	22	23	24	25	26
abs. Häufigkeit	10	14	20	12	4

Schätzen Sie Erwartungswert und Varianz der zugehörigen Grundgesamtheit.

8–4 Bei 100 Blättern eines Apfelbaumes wurde die Blattfläche bestimmt. Man erhielt als Mittelwert 18 cm^2 mit einem Standardfehler von 0.3 cm^2 . Wieviel Blätter dieses Baumes müsste man ausmessen, um den Standardfehler auf 0.1 cm^2 zu verkleinern?

8–5 Der Intelligenzquotient (IQ) ist so normiert, dass er $N(100; 225)$ -verteilt ist (vgl. Beispiel 5.10.5.A). Es werden nun Gruppen von 25 bzw. 400 Personen gebildet; in jeder Gruppe wird der durchschnittliche IQ berechnet. a) Wie gross sind Erwartungswert und Standardabweichung dieses Durchschnitts? b) Mit welcher Wahrscheinlichkeit liegt der IQ einer einzelnen Person zwischen 97 und 103? c) Mit welcher Wahrscheinlichkeit liegt der durchschnittliche IQ einer Gruppe von 25 Personen zwischen 97 und 103?

8–6 Die Zufallsgrösse T folge der t -Verteilung mit Freiheitsgrad 20. Bestimmen Sie unter Verwendung der Tabelle (51.4) die Zahl t so, dass gilt

a) $P(T \geq t) = 0.01$,

b) $P(t \leq T \leq 0) = 0.45$.

Illustrieren Sie beides mit einer Skizze.

8–7 **Für Freiheitsgrad** $n = 1$ hat die Dichtefunktion der t -Verteilung die einfache Form $f(x) = c_2(1 + x^2)^{-1}$ (vgl. Formel (*) aus (8.3)).

a) Bestimmen Sie den Wert der Konstanten c_2 , vgl. Beispiel 5.8.B.

b) Im Allgemeinen ist die Berechnung der kritischen Werte nicht einfach. In diesem Fall ist eine direkte Rechnung aber möglich. Bestimmen Sie den kritischen Wert $t_{\alpha,1}$ für $\alpha = 0.05$ und 0.01 . Vergleichen Sie Ihr Resultat mit der Tabelle (6.2.6).

8–8 Aus einer normal verteilten Grundgesamtheit mit unbekanntem Erwartungswert μ und unbekannter Standardabweichung σ wird folgende Stichprobe entnommen:

24, 34, 32, 36, 38, 32, 28.

a) Schätzen Sie μ und σ .

b) Bestimmen Sie das Konfidenzintervall für μ mit dem Koeffizienten $Q = 95\%$.

c) Dasselbe für $Q = 90\%$.

8–9 Aus der laufenden Produktion von Nägeln wurden acht Stück herausgegriffen und gemessen. Es ergaben sich folgende Längen (in mm):

50.12, 49.96, 50.35, 50.02, 49.80, 51.00, 50.12, 49.75.

Geben Sie das Konfidenzintervall für die mittlere Länge der gesamten Produktion an. a) $Q = 90\%$, b) $Q = 99.9\%$.

8–10 Eine Untersuchung von 100 Eiern von Stockenten lieferte das Durchschnittsgewicht 51.2 g mit einer Varianz von 16 g^2 . Berechnen Sie das Konfidenzintervall mit $Q = 99\%$ für das mittlere Ei-Gewicht der Stockente schlechthin.

8–11 Die Untersuchung der Körperlänge von zehnjährigen Knaben ergab die Werte $\bar{x} = 141 \text{ cm}$, $s = 6 \text{ cm}$. Berechnen Sie das 95% -Vertrauensintervall für die mittlere Körperlänge, wenn der Umfang der Stichprobe a) $= 30$, b) $= 300$ war.

- 8–12 Eine Kontrolle von 100 Zuckersäcken ergab das folgende Konfidenzintervall (mit $Q = 95\%$) für das mittlere Gewicht (in Gramm): [1996, 2008]. Welche der folgenden Aussagen sind richtig?
- A: 95% aller produzierten Zuckersäcke haben ein Gewicht zwischen 1996 g und 2008 g.
 - B: Wir vertrauen zu 95% darauf, dass das mittlere Gewicht der gesamten Produktion zwischen 1996 g und 2008 g liegt.
 - C: Das mittlere Gewicht der gesamten Produktion beträgt mit 95% Wahrscheinlichkeit 2002 g.
 - D: Das mittlere Gewicht der 100 kontrollierten Säcke beträgt 2002 g.
- 8–13 Eine Stichprobe (betreffend eine Längenmessung) vom Umfang $n = 10$ ergibt für $Q = 95\%$ ein Konfidenzintervall der Länge 10 cm. Wie gross muss der Umfang n mindestens gewählt werden, damit das Konfidenzintervall nur noch die Länge 5 cm hat? Dabei wird angenommen, dass die Standardabweichung s bei beiden Stichproben dieselbe ist. Bestimmen Sie n so gut, wie es anhand der Tabelle (6.2.6) möglich ist.
- 8–14 Im Jahr 1983 wurden in der Stadt Zürich 2994 Kinder geboren. Davon waren 1562 Knaben. Geben Sie das Konfidenzintervall für die Wahrscheinlichkeit einer Knabengeburt an a) für $Q = 95\%$, b) für $Q = 99\%$.
- 8–15 Von 1000 Werkstücken aus einer Sendung erwiesen sich 30 als defekt. Bestimmen Sie das Vertrauensintervall für den Anteil der defekten Stücke in der Gesamtproduktion. Arbeiten Sie mit $Q = 90\%$.