WEIHENSTEPHAN · TRIESDORF
University of Applied Sciences

Prof. Dr. Klaus Eckhardt

Optimaler Stichprobenumfang

veröffentlicht im Internet unter aufgabomat.de

Inhalt

1 Einleitung

Statistische Untersuchungen müssen immer mit begrenzten Ressourcen durchgeführt werden, ob personell, finanziell oder zeitlich. Im Rahmen der Versuchsplanung sollte daher geklärt werden, welcher Aufwand notwendig ist, um eine gewisse Aussagegenauigkeit zu erzielen, oder, umgekehrt, welche statistische Sicherheit in den Ergebnissen erreicht werden kann, wenn nur Untersuchungen eines vorgegebenen Umfangs durchgeführt werden können. Dies betrifft sehr wesentlich den Stichprobenumfang. Im Folgenden wird erörtert, wie sich der optimale Stichprobenumfang im Fall dreier statistischer Verfahren bestimmen lässt.

Die Berechnung der Stichprobenumfänge basiert unter anderem auf Annahmen über den noch zu gewinnenden Datensatz. Da diese Annahmen unsicher sind, sind es auch die ausgewiesenen Stichprobenumfänge. Daher wird jeweils eine Unsicherheitsanalyse angeschlossen.

Entstanden ist auf diese Weise ein Skript, das viel an mathematischen Hintergrundinformationen bietet. Wollen oder können Sie sich nicht so weit in die Materie einarbeiten, so können Sie sich aber auch auf die zentralen Ergebnisse der Ausführungen konzentrieren: Die Beziehungen 3, 10 und 18 und wie sie zur Bestimmung des optimalen Stichprobenumfangs eingesetzt werden, sowie die Gleichungen 5, 12 und 20, die eine Abschätzung erlauben, welche Unsicherheit die entsprechenden Resultate aufweisen.

2 Intervallschätzung für den Erwartungswert

2.1 Bestimmung des optimalen Stichprobenumfangs

Das Konfidenzintervall bzw. der Vertrauensbereich für den Erwartungswert E(X) einer normalverteilten Zufallsvariable X zur statistischen Sicherheit bzw. zum **Konfidenzniveau** $\gamma = 1 - \alpha$ wird angegeben als $\mu = \bar{x} \pm \Delta x$ mit

$$\Delta x = t_{1-\alpha/2} \frac{s}{\sqrt{N}} .^{1} \qquad (1)$$

μ ist das Symbol für einen der beiden Parameter der Normalverteilung, den Mittelwert. Speziell im Fall normalverteilter Variablen stimmt der Erwartungswert der Variable mit diesem Parameter überein, d. h. es ist $E(X) = \mu$. Δx entspricht der halben Breite des Konfidenzintervalls und steht damit für die Genauigkeit, mit der die Aussage über den Erwartungswert getroffen wird. Ferner sind

- N: Stichprobenumfang
- $\bar{x}$ und s: empirischer Mittelwert und empirische Standardabweichung der Stichprobenwerte
- $t_{1-\alpha/2}$: $(1 - \alpha/2)$-Quantil der t-Verteilung mit dem Freiheitsgrad $f = N - 1$.

Wird Gleichung 1 nach dem Stichprobenumfang N aufgelöst, so ergibt sich

$$N = \left(t_{1-\alpha/2} \frac{s}{\Delta x}\right)^2 . \qquad (2)$$

[1] Siehe beispielsweise Skript „Intervallschätzung für den Erwartungswert“ unter aufgabomat.de.

Dies ist allerdings keine Gleichung, mit der sich N berechnen ließe, denn der Stichprobenumfang bestimmt den Freiheitsgrad der t-Verteilung und damit den Quantilwert $t_{1-\alpha/2}$. Er steckt also implizit auch in der rechten Seite der Gleichung. Wie lässt sich dennoch eine Aussage über N treffen?

Gegeben seien
- die Wahrscheinlichkeit α
- eine untere Begrenzung s* für die empirische Standardabweichung s
- die gewünschte Genauigkeit Δx in der Aussage über den Erwartungswert.

Tabelle 1 zeigt Quantile t_p zu unterschiedlichen Unterschreitungswahrscheinlichkeiten p für t-Verteilungen unterschiedlichen Freiheitsgrades f.

f	p						f	p					
	0,841	0,900	0,950	0,975	0,990	0,995		0,841	0,900	0,950	0,975	0,990	0,995
2	1,32	1,89	2,92	4,30	6,96	9,92	19	1,03	1,33	1,73	2,09	2,54	2,86
3	1,20	1,64	2,35	3,18	4,54	5,84	20	1,03	1,33	1,72	2,09	2,53	2,85
4	1,14	1,53	2,13	2,78	3,75	4,60	21	1,02	1,32	1,72	2,08	2,52	2,83
5	1,11	1,48	2,02	2,57	3,36	4,03	22	1,02	1,32	1,72	2,07	2,51	2,82
6	1,09	1,44	1,94	2,45	3,14	3,71	23	1,02	1,32	1,71	2,07	2,50	2,81
7	1,08	1,41	1,89	2,36	3,00	3,50	24	1,02	1,32	1,71	2,06	2,49	2,80
8	1,07	1,40	1,86	2,31	2,90	3,36	25	1,02	1,32	1,71	2,06	2,49	2,79
9	1,06	1,38	1,83	2,26	2,82	3,25	26	1,02	1,31	1,71	2,06	2,48	2,78
10	1,05	1,37	1,81	2,23	2,76	3,17	27	1,02	1,31	1,70	2,05	2,47	2,77
11	1,05	1,36	1,80	2,20	2,72	3,11	28	1,02	1,31	1,70	2,05	2,47	2,76
12	1,04	1,36	1,78	2,18	2,68	3,05	29	1,02	1,31	1,70	2,05	2,46	2,76
13	1,04	1,35	1,77	2,16	2,65	3,01	30	1,02	1,31	1,70	2,04	2,46	2,75
14	1,04	1,35	1,76	2,14	2,62	2,98	40	1,01	1,30	1,68	2,02	2,42	2,70
15	1,03	1,34	1,75	2,13	2,60	2,95	50	1,01	1,30	1,68	2,01	2,40	2,68
16	1,03	1,34	1,75	2,12	2,58	2,92	70	1,01	1,29	1,67	1,99	2,38	2,65
17	1,03	1,33	1,74	2,11	2,57	2,90	100	1,01	1,29	1,66	1,98	2,36	2,63
18	1,03	1,33	1,73	2,10	2,55	2,88	∞	1,00	1,28	1,64	1,96	2,33	2,58

Tabelle 1: Quantile t_p der t-Verteilung.

Bei gegebener Unterschreitungswahrscheinlichkeit werden die Quantilwerte mit steigendem Freiheitsgrad und damit steigendem Stichprobenumfang N immer kleiner. Der betreffende Grenzwert für $N \to \infty$ sei mit t_p* bezeichnet. Wird p gemäß der Notation in den Gleichungen 1 und 2 durch 1 – α/2 ersetzt, so lässt sich schreiben:

$$\lim_{N\to\infty} t_{1-\alpha/2} = t_{1-\alpha/2}* .$$

In der Realität sind Stichprobenumfänge aber endlich und es gilt daher $t_{1-\alpha/2} > t_{1-\alpha/2}*$. Setzt man dies in Gleichung 2 ein, so ergibt sich

$$N \geq \left(t_{1-\alpha/2} * \frac{s*}{\Delta x} \right)^2 . \qquad (3)$$

Damit lässt sich ein Mindeststichprobenumfang berechnen. Der optimale Stichprobenumfang ist derjenige, für den in Beziehung 3 linke und rechte Seite mit angepasstem Wert des (1 – α/2)-Quantils gleich werden. Dies erfordert, wie nachfolgend an einem Beispiel erläutert, eine iterative Berechnung. Dabei werden prinzipiell gleiche Rechenschritte wiederholt ausgeführt, wobei im aktuellen Schritt jeweils das Ergebnis des vorangehenden Schrittes verwendet wird.

Beispiel: In einer Bevölkerungsgruppe soll die mittlere Körpergröße erwachsener Frauen mit einer statistischen Sicherheit von 95 % auf 2,0 cm genau bestimmt werden. Aus anderen Datenerhebungen lässt sich ableiten, dass mit einer Standardabweichung von 6,0 cm zu rechnen ist. Wie viele Messwerte müssen erfasst werden?

$1 - \alpha = 0{,}95 \Rightarrow 1 - \alpha/2 = 0{,}975$

$t_{1-\alpha/2}^* = t_{0,975}^*$
$= 1{,}96$ (Tabelle 1)

$$N \geq \left(1{,}96 \, \frac{6{,}0 \text{ cm}}{2{,}0 \text{ cm}}\right)^2$$
$$\geq 35$$

Bei der Angabe des Mindeststichprobenumfangs sollte aufgerundet werden.

Nun zur iterativen Berechnung des optimalen Stichprobenumfangs. Wäre N = 35, so müsste auf der rechten Seite der obigen Beziehung $t_{0,975}^* = 1{,}96$ ersetzt werden durch $t_{0,975} = 2{,}03$, denn dies ist nach Tabelle 1 näherungsweise der Wert für das 0,975-Quantil der t-Verteilung mit f = 35 – 1 = 34. Damit ergäbe sich auf der rechten Seite

$$\left(2{,}03 \, \frac{6{,}0 \text{ cm}}{2{,}0 \text{ cm}}\right)^2 = 37.$$

Linke und rechte Seite stimmen noch nicht überein, daher wird der vorige Rechenschritt mit angepasstem Wert für N wiederholt. Wäre N = 37, so müsste die Berechnung mit dem 0,975-Quantil der t-Verteilung mit f = 37 – 1 = 36 durchgeführt werden. Im Rahmen der Genauigkeit, mit der die Quantile in Tabelle 1 angegeben sind, ist dies ebenfalls der Wert 2,03, sodass sich an dem Ergebnis nichts mehr ändert. Mit N = 37 ist die gesuchte Gleichheit zwischen linker und rechter Seite erreicht. Der optimale Stichprobenumfang ist 37. In Kurzform ließe sich diese iterative Bestimmung von N wie folgt dokumentieren:

N	rechte Seite von Beziehung 3
35	$\left(2{,}03 \, \frac{6{,}0 \text{ cm}}{2{,}0 \text{ cm}}\right)^2 = 37$
37	$\left(2{,}03 \, \frac{6{,}0 \text{ cm}}{2{,}0 \text{ cm}}\right)^2 = 37$

N = 37

Dieser Wert weist allerdings eine erhebliche Unsicherheit auf, die im nachfolgenden Abschnitt 2.2 analysiert wird.

Es kann geschehen, dass die iterativ ermittelten Zwischenergebnisse nicht kontinuierlich auf einen Endwert zulaufen, sondern zu oszillieren beginnen. In diesem Fall wechseln die Resultate von Iterationsschritt zu Iterationsschritt endlos zwischen zwei Werten hin und her. Sicherheitshalber sollte dann der größere dieser beiden Werte als Endergebnis ausgewiesen werden. Ein Beispiel dafür ist am Ende des Abschnitts 3 zu sehen.

Weitere Übungsaufgaben finden sich beispielweise im Internet unter der Adresse aufgabomat.de in der Rubrik Statistik.

2.2 Unsicherheitsanalyse

Zwei Größen sind bei der Bestimmung des Stichprobenumfangs N nach Beziehung 3 unsicher, das (1 – α/2)-Quantil und die Standardabweichung. Die Unsicherheit dieser beiden Größen pflanzt sich in den berechneten Stichprobenumfang fort. In welchem Maße dies der Fall ist und von welcher Eingangsgröße der Berechnung der größte Beitrag zur Unsicherheit des Stichprobenumfangs kommt, wird in der folgenden Analyse geklärt. Die Grundlagen der Unsicherheitsanalyse lassen sich beispielsweise dem Skript „Messunsicherheit“ unter aufgabomat.de entnehmen.

Nach dem Unsicherheitsfortpflanzungsgesetz ergibt sich für die kombinierte Standardunsicherheit des Stichprobenumfangs

$$\begin{aligned} u_C(N) &= \sqrt{\left[\frac{\partial N}{\partial t_{1-\alpha/2}}\, u(t_{1-\alpha/2})\right]^2 + \left[\frac{\partial N}{\partial s^*}\, u(s^*)\right]^2} \\ &= \sqrt{\left[2\, t_{1-\alpha/2} \left(\frac{s^*}{\Delta x}\right)^2 u(t_{1-\alpha/2})\right]^2 + \left[2\, s^* \left(\frac{t_{1-\alpha/2}}{\Delta x}\right)^2 u(s^*)\right]^2}\,. \end{aligned} \tag{4}$$

Darin sind $u(t_{1-\alpha/2})$ und $u(s^*)$ die Standardunsicherheiten des (1 – α/2)-Quantils und der Standardabweichung.

Beispiel: Bestimmung der mittleren Körpergröße erwachsener Frauen (Abschnitt 2.1)

Für das (1–α/2)-Quantils kann bei Verwendung von Tabelle 1 nur ein auf zwei Nachkommastellen genauer Wert $t_{1-\alpha/2}$ ermittelt werden. Da zudem mit der Unsicherheit bezüglich des Stichprobenumfangs eine Unsicherheit bezüglich des Freiheitsgrades der t-Verteilung einhergeht, wird für die maximale Unsicherheit dieses Wertes $\Delta t_{1-\alpha/2} = 0{,}01$ angesetzt. Unter der Annahme, dass der tatsächliche Quantilwert mit gleichverteilter Wahrscheinlichkeit im Intervall $[t_{1-\alpha/2} - \Delta t_{1-\alpha/2};\ t_{1-\alpha/2} + \Delta t_{1-\alpha/2}]$ liegt, ergibt sich seine Standardunsicherheit als

$$\begin{aligned} u(t_{1-\alpha/2}) &= \frac{0{,}01}{\sqrt{3}} \\ &= 0{,}006. \end{aligned}$$

Die maximale Abweichung Δs* der Standardabweichung hängt vom betrachteten Anwendungsfall ab. Im Beispiel des Abschnitts 2.1 könnte Δs* = 1,0 cm angesetzt werden, sodass die Standardunsicherheit den Wert

$$\begin{aligned} u(s^*) &= \frac{1{,}0\ \text{cm}}{\sqrt{3}} \\ &= 0{,}6\ \text{cm} \end{aligned}$$

annimmt. Es folgt

$$u_C(N) = \sqrt{\left[2 \cdot 1{,}96 \left(\frac{6{,}0\ \text{cm}}{2{,}0\ \text{cm}}\right)^2 0{,}006\right]^2 + \left[2 \cdot 6{,}0\ \text{cm} \left(\frac{1{,}96}{2{,}0\ \text{cm}}\right)^2 0{,}6\ \text{cm}\right]^2}$$
$$= \sqrt{0{,}04 + 47{,}8}$$
$$= 7.$$

Der weitaus größere Beitrag zur Unsicherheit in N kommt in diesem Beispiel von der Unsicherheit bezüglich der Standardabweichung s*. Die Verwendung der auf zwei Nachkommastellen gerundeten Quantilwerte aus Tabelle 1 ist demgegenüber völlig ausreichend. Der Stichprobenumfang selbst sollte angegeben werden als

N = 37 mit einer kombinierten Standardunsicherheit von 7.

Berücksichtigt man außerdem, dass Quantilwerte in der Praxis häufig mit einem Tabellenkalkulations- oder Statistikprogramm bestimmt werden können, sodass sie mit noch geringerer Unsicherheit vorliegen als in Tabelle 1, kann Gleichung 4 auch vereinfacht werden zu

$$u(N) = 2\, s^* \left(\frac{t_{1-\alpha/2}}{\Delta x}\right)^2 u(s^*)\ . \tag{5}$$

3 Einstichproben-t-Test

3.1 Bestimmung des optimalen Stichprobenumfangs

Der Einstichproben-t-Test, auch **Parametertest für den Erwartungswert** genannt, dient der Überprüfung, ob der Erwartungswert E(X) einer normalverteilten Zufallsvariable X mit einem vorgegebenen Sollwert μ_0 übereinstimmt. In seiner zweiseitigen Variante wird er mit den folgenden Hypothesen durchgeführt:
H_0: $E(X) = \mu_0$
H_1: $E(X) \neq \mu_0$.

Der Wert der Teststatistik bzw. Prüfwert wird berechnet als

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{N}}\ . \tag{6}$$

H_0 wird beibehalten, falls t in das Annahmeintervall $[t_{\alpha/2};\ t_{1-\alpha/2}]$ fällt, d. h. falls $t_{\alpha/2} \leq t \leq t_{1-\alpha/2}$. Aufgrund der speziellen Symmetrie der Wahrscheinlichkeitsdichtefunktion der t-Verteilung ist $t_{\alpha/2} = -t_{1-\alpha/2}$ (Abbildung 1). Daher lässt sich auch schreiben: H_0 wird beibehalten, falls

$$|t| \leq t_{1-\alpha/2}$$

$$\frac{|\bar{x} - \mu_0|}{s / \sqrt{N}} \leq t_{1-\alpha/2}$$

$$|\bar{x} - \mu_0| \leq Q_{1-\alpha/2}. \tag{7}$$

Darin ist $t_{1-\alpha/2}$ das $(1-\alpha/2)$-Quantil der t-Verteilung mit dem Freiheitsgrad $f = N - 1$ und

$$Q_{1-\alpha/2} = \frac{s\, t_{1-\alpha/2}}{\sqrt{N}} \tag{8}$$

das $(1-\alpha/2)$-Quantil derjenigen Verteilung, welche die Zufallsvariable $\overline{X} - \mu_0$ beschreibt.

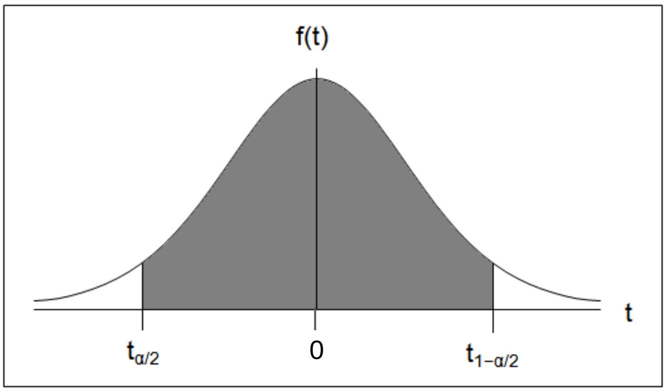

Abbildung 1: Wahrscheinlichkeitsdichtefunktion der t-Verteilung.

Gegeben seien

- die Wahrscheinlichkeit α
- eine untere Begrenzung s^* für die empirische Standardabweichung s
- die Mindestabweichung $|\,\overline{x} - \mu_0\,|^* = \Delta$ zwischen empirischem Mittelwert und Sollwert, die als Hinweis auf die Gültigkeit der Alternativhypothese zu werten sein soll.

Bei der letzten dieser drei Angaben geht es darum, mit welcher Auflösung der Test arbeiten soll.

Auch wenn die Nullhypothese gilt, liegen die Werte der Teststatistik mit der Wahrscheinlichkeit α außerhalb des Annahmeintervalls. Mit der Wahrscheinlichkeit α wird die Nullhypothese daher irrtümlich verworfen, oder, wie man auch sagt, der **Fehler erster Art** begangen. Die Wahrscheinlichkeit für diesen Fehler lässt sich durch den Wert von α direkt vorgeben.

Wird dagegen mit $|\,\overline{x} - \mu_0\,|^* \leq \Delta$ die Nullhypothese beibehalten, kann der **Fehler zweiter Art** begangen werden: Die Nullhypothese kann irrtümlich beibehalten, die Abweichung $|\,\overline{x} - \mu_0\,|$ irrtümlich als *nicht* signifikant ausgewiesen werden. Auch diese Wahrscheinlichkeit, die nachfolgend mit β bezeichnet wird, sollte möglichst kontrolliert werden. Dazu die folgenden Überlegungen:

Angenommen, die Nullhypothese H_0 gilt, d. h. es ist $E(X) = \mu_0$. Dann ist der Erwartungswert der Variable $\overline{X} - \mu_0$

$$\begin{aligned} E(\overline{X} - \mu_0) &= E(\overline{X}) - \mu_0 \\ &= E(X) - \mu_0 \\ &= 0 \end{aligned}$$

und die Variable wird durch die in Abbildung 2 dargestellte Wahrscheinlichkeitsdichtefunktion beschrieben. Gemäß Beziehung 7 wird H_0 beibehalten, falls $|\,\overline{x} - \mu_0\,| \leq Q_{1-\alpha/2}$. Da hier vorausgesetzt wurde, dass die Nullhypothese gilt, ist dies eine korrekte Entscheidung. Die Wahrscheinlichkeit für sie entspricht der in Abbildung 2 grau ausgefüllten Fläche unter der Dichtefunktion und beträgt $1 - \alpha$.

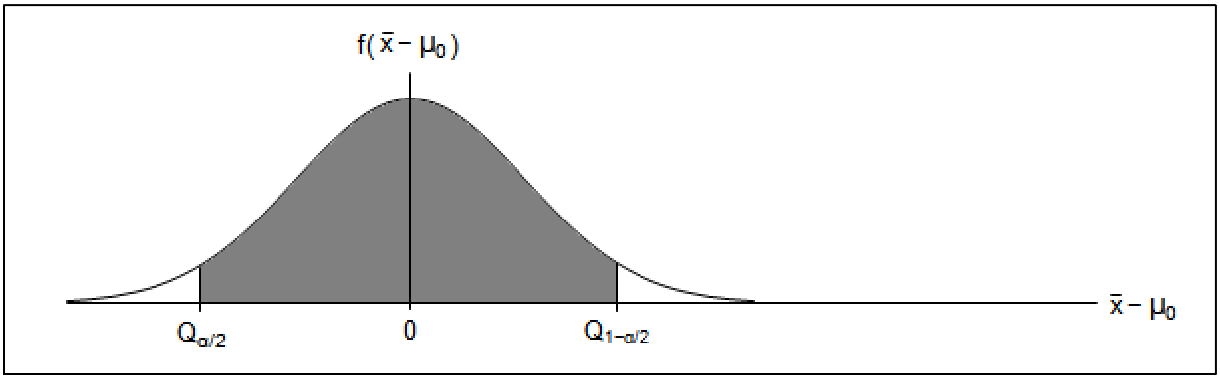

Abbildung 2: Wahrscheinlichkeit, H_0 korrekterweise beizubehalten, falls $E(X) - \mu_0 = 0$.

Angenommen, die Nullhypothese H_0 gilt nicht und der Erwartungswert $E(X)$ weicht um Δ vom Sollwert μ_0 ab: $E(X) = \mu_0 + \Delta$. In diesem Fall ist

$$\begin{aligned} E(\overline{X} - \mu_0) &= E(\overline{X}) - \mu_0 \\ &= E(X) - \mu_0 \\ &= \Delta \end{aligned}$$

und die Variable $\overline{X} - \mu_0$ wird durch eine Wahrscheinlichkeitsdichtefunktion beschrieben, die gegenüber der in Abbildung 2 gezeigten um Δ verschoben ist (Abbildung 3).

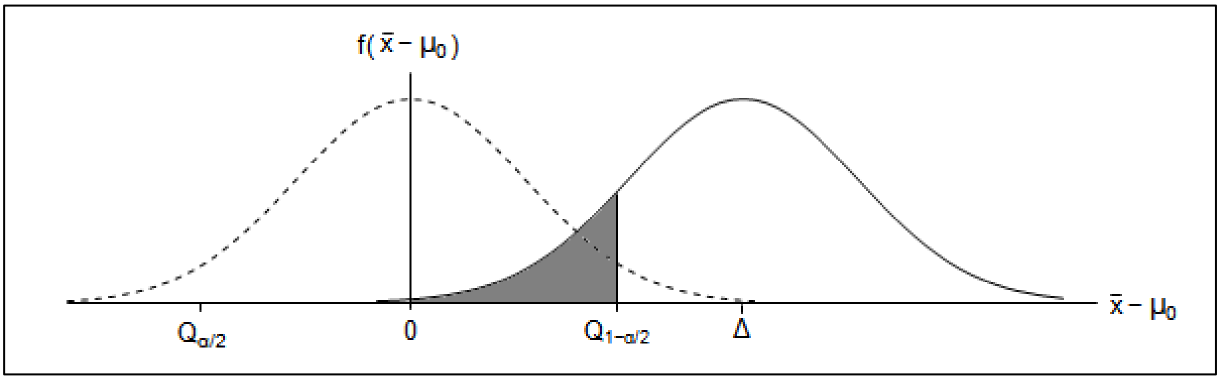

Abbildung 3: Wahrscheinlichkeit, H_0 irrtümlich beizubehalten, falls $E(X) - \mu_0 = \Delta$.

Diejenige Person, die den Test durchführt, weiß dies nicht und bleibt daher dabei, H_0 als gültig anzusehen, falls $|\bar{x} - \mu_0| \leq Q_{1-\alpha/2}$. Die Wahrscheinlichkeit, dass die Variable $\overline{X} - \mu_0$ einen Wert kleiner gleich $Q_{1-\alpha/2}$ annimmt, entspricht der in Abbildung 3 grau ausgefüllten Fläche. Mit dieser Wahrscheinlichkeit wird – da vorausgesetzt wurde, dass die Nullhypothese nicht gilt – eine falsche Entscheidung getroffen, nämlich die Nullhypothese irrtümlich beibehalten. Die grau ausgefüllte Fläche entspricht damit der bereits angesprochenen Wahrscheinlichkeit β, den Fehler zweiter Art zu begehen. $Q_{1-\alpha/2}$ kann auch als das β-Quantil der in Abbildung 3 rechts dargestellten Dichtefunktion bezeichnet werden:

$(1-\alpha/2)$-Quantil $Q_{1-\alpha/2}$ der linken Dichtefunktion = β-Quantil der rechten Dichtefunktion.

Die beiden Dichtefunktionen in Abbildung 3 sind um Δ gegeneinander verschoben, ansonsten aber deckungsgleich. Daher ergibt sich nicht nur der Erwartungswert der rechten Dichtefunktion durch Verschiebung des Erwartungswertes der linken Dichtefunktion um Δ. Auch alle Quantile der rechten Dichtefunktion ergeben sich durch Verschiebung der entsprechenden Quantile der linken Dichtefunktion um Δ. So auch das β-Quantil:

(1−α/2)-Quantil $Q_{1-\alpha/2}$ der linken Dichtefunktion = β-Quantil der rechten Dichtefunktion
= β-Quantil der linken Dichtefunktion + Δ.

Schließlich überlegt man sich noch, dass im Fall der linken Dichtefunktion aufgrund ihrer speziellen Symmetrie $Q_\beta = -Q_{1-\beta}$ ist. Dies entspricht der Situation in Abbildung 1, wenn dort α/2 durch β ersetzt wird.

(1−α/2)-Quantil $Q_{1-\alpha/2}$ der linken Dichtefunktion = β-Quantil der rechten Dichtefunktion
= β-Quantil der linken Dichtefunktion + Δ
= −(1− β)-Quantil der linken Dichtefunktion + Δ.

Mathematisch lässt sich das entsprechend Gleichung 8 auch wie folgt ausdrücken:

$$\frac{s\, t_{1-\alpha/2}}{\sqrt{N}} = -\frac{s\, t_{1-\beta}}{\sqrt{N}} + \Delta.$$

Durch Umformung ergibt sich

$$\Delta \sqrt{N} = s\,(t_{1-\alpha/2} + t_{1-\beta})$$

$$N = \frac{s^2\,(t_{1-\alpha/2} + t_{1-\beta})^2}{\Delta^2}. \qquad (9)$$

Dies ist noch keine Gleichung, mit der sich N berechnen ließe, denn der Stichprobenumfang bestimmt den Freiheitsgrad der t-Verteilung und damit die Quantilwerte $t_{1-\alpha/2}$ und $t_{1-\beta}$. Er steckt also implizit auch in der rechten Seite der Gleichung. Wie lässt sich dennoch eine Aussage über N treffen?

Gegeben seien

- die Wahrscheinlichkeiten α und β für die Fehler erster und zweiter Art
- eine untere Begrenzung s* für die empirische Standardabweichung s
- die Mindestabweichung Δ zwischen empirischem Mittelwert und Sollwert, die als signifikant gelten soll.

Es ist $\lim\limits_{N\to\infty} t_{1-\alpha/2} = t_{1-\alpha/2}{}^*$ und $\lim\limits_{N\to\infty} t_{1-\beta} = t_{1-\beta}{}^*$.

In der Realität sind Stichprobenumfänge aber endlich und es gilt daher $t_{1-\alpha/2} > t_{1-\alpha/2}{}^*$ und $t_{1-\beta} > t_{1-\beta}{}^*$. Setzt man dies in Gleichung 9 ein, so ergibt sich

$$N \geq \frac{s^{*2}\,(t_{1-\alpha/2}{}^* + t_{1-\beta}{}^*)^2}{\Delta^2}. \qquad (10)$$

β ist die Wahrscheinlichkeit, die Nullhypothese des Tests irrtümlich beizubehalten, 1 – β die Wahrscheinlichkeit, die Nullhypothese korrekterweise zu verwerfen. 1 – β steht somit für die Fähigkeit des Einstichproben-t-Tests, eine Abweichung des Mittelwerts vom Sollwert korrekt zu erkennen. Darum wird 1 – β auch als die **Trennschärfe des Tests** oder **Teststärke** bezeichnet.

Beispiel: Eine Maschine soll Schrauben eines möglichst exakt einzuhaltenden Durchmessers herstellen. Inwieweit sie die Vorgabe einhält, soll experimentell überprüft werden. Als Irrtumswahrscheinlichkeit bzw. Wahrscheinlichkeit für den Fehler erster Art wird $\alpha = 0{,}05$ gewählt. Kritischer ist es, den Fehler zweiter Art zu begehen, d. h. irrtümlich von der Annahme auszugehen, dass die Maschine ausreichend präzise arbeitet, sodass bei ihrem Praxiseinsatz ein unerwartet hoher Anteil an unbrauchbaren Schrauben produziert wird. Daher wird $\beta = 0{,}01$ bzw. $1 - \beta = 0{,}99$ gesetzt. Mit dieser Teststärke sollen Abweichungen Δ des mittleren Durchmessers vom Sollwert ab 0,20 mm identifiziert werden können. Wie viele Messwerte müssen erhoben werden, wenn für die Standardabweichung des Durchmessers $s^* = 0{,}10$ mm anzusetzen ist?

$$t_{1-\alpha/2}^* = t_{0,975}^* = 1{,}96 \qquad t_{1-\beta}^* = t_{0,99}^* = 2{,}33$$

Die Quantilwerte sind Tabelle 1 entnommen. Nach Beziehung 10 ist

$$N \geq \frac{(0{,}10\text{ mm})^2\,(1{,}96 + 2{,}33)^2}{(0{,}20\text{ mm})^2} \geq 5.$$

Die iterative Ermittlung des optimalen Stichprobenumfangs[2] liefert die folgenden Ergebnisse:

N	rechte Seite von Beziehung 10
5	$\frac{(0{,}10\text{ mm})^2\,(2{,}78 + 3{,}75)^2}{(0{,}20\text{ mm})^2} = 11$
11	$\frac{(0{,}10\text{ mm})^2\,(2{,}23 + 2{,}76)^2}{(0{,}20\text{ mm})^2} = 6$
6	$\frac{(0{,}10\text{ mm})^2\,(2{,}57 + 3{,}36)^2}{(0{,}20\text{ mm})^2} = 9$
9	$\frac{(0{,}10\text{ mm})^2\,(2{,}31 + 2{,}90)^2}{(0{,}20\text{ mm})^2} = 7$
7	$\frac{(0{,}10\text{ mm})^2\,(2{,}45 + 3{,}14)^2}{(0{,}20\text{ mm})^2} = 8$
8	$\frac{(0{,}10\text{ mm})^2\,(2{,}36 + 3{,}00)^2}{(0{,}20\text{ mm})^2} = 7$

Die Berechnungen können an dieser Stelle abgebrochen werden, da es zu einem permanenten Wechsel der Resultate zwischen 7 und 8 kommt. Als Ergebnis wird der höhere dieser beiden Werte ausgewiesen:

$N = 8.$

Dieser Wert weist allerdings eine erhebliche Unsicherheit auf, die im nachfolgenden Abschnitt 3.2 analysiert wird.

Weitere Übungsaufgaben finden sich beispielsweise im Internet unter der Adresse aufgabomat.de in der Rubrik Statistik.

[2] Nähere Erläuterungen zum Prinzip des Verfahrens im Beispiel des Abschnitts 2.

3.2 Unsicherheitsanalyse

Drei Größen sind bei der Bestimmung des Stichprobenumfangs N nach Beziehung 10 unsicher, das (1 – α/2)-Quantil, das (1 – β)-Quantil und die Standardabweichung. Die Unsicherheit dieser drei Größen pflanzt sich in den berechneten Stichprobenumfang fort. In welchem Maße dies der Fall ist und von welcher Eingangsgröße der Berechnung der größte Beitrag zur Unsicherheit des Stichprobenumfangs kommt, wird in der folgenden Analyse geklärt. Die Grundlagen der Unsicherheitsanalyse lassen sich beispielsweise dem Skript „Messunsicherheit" unter aufgabomat.de entnehmen.

Nach dem Unsicherheitsfortpflanzungsgesetz ergibt sich für die kombinierte Standardunsicherheit des Stichprobenumfangs

$$u_C(N) = \sqrt{\left[\frac{\partial N}{\partial t_{1-\alpha/2}}\,u(t_{1-\alpha/2})\right]^2 + \left[\frac{\partial N}{\partial t_{1-\beta}}\,u(t_{1-\beta})\right]^2 + \left[\frac{\partial N}{\partial s^*}\,u(s^*)\right]^2}$$

$$= \sqrt{\left[2\left(\frac{s^*}{\Delta}\right)^2 (t_{1-\alpha/2} + t_{1-\beta})\right]^2 \left\{[u(t_{1-\alpha/2})]^2 + [u(t_{1-\beta})]^2\right\} + \left[2\,s^*\left(\frac{t_{1-\alpha/2} + t_{1-\beta}}{\Delta}\right)^2 u(s^*)\right]^2}\ . \qquad (11)$$

Darin sind $u(t_{1-\alpha/2})$, $u(t_{1-\beta})$ und $u(s^*)$ die Standardunsicherheiten der Quantile und der Standardabweichung.

Beispiel: Vergleich des mittleren Schraubendurchmessers mit seinem Sollwert (Abschnitt 3.1)

Für das (1–α/2)-Quantil und das (1–β)-Quantil können bei Verwendung von Tabelle 1 nur auf zwei Nachkommastellen genaue Werte $t_{1-\alpha/2}$ und $t_{1-\beta}$ ermittelt werden. Da zudem mit der Unsicherheit bezüglich des Stichprobenumfangs eine Unsicherheit bezüglich des Freiheitsgrades der t-Verteilung einhergeht, wird für die maximale Unsicherheit dieser Werte $\Delta t_{1-\alpha/2} = \Delta t_{1-\beta} = 0{,}01$ angesetzt. Unter der Annahme, dass die tatsächlichen Quantilwerte mit gleichverteilter Wahrscheinlichkeit in den Intervallen $[t_{1-\alpha/2} - \Delta t_{1-\alpha/2};\ t_{1-\alpha/2} + \Delta t_{1-\alpha/2}]$ und $[t_{1-\beta} - \Delta t_{1-\beta};\ t_{1-\beta} + \Delta t_{1-\beta}]$ liegen, ergibt sich ihre Standardunsicherheit als

$$u(t_{1-\alpha/2}) = \frac{0{,}01}{\sqrt{3}} = 0{,}006 \qquad u(t_{1-\beta}) = \frac{0{,}01}{\sqrt{3}} = 0{,}006.$$

Die maximale Abweichung Δs^* der Standardabweichung hängt vom betrachteten Anwendungsfall ab. Im Beispiel des Abschnitts 3.1 könnte $\Delta s^* = 0{,}05$ mm angesetzt werden, sodass die Standardunsicherheit den Wert

$$u(s^*) = \frac{0{,}05\ \text{mm}}{\sqrt{3}} = 0{,}03\ \text{mm}$$

annimmt. Es folgt

$$u_C(N) = \sqrt{\left[2\left(\frac{0{,}10\ \text{mm}}{0{,}20\ \text{mm}}\right)^2 (1{,}96 + 2{,}33)\right]^2 2\cdot 0{,}006^2 + \left[2\cdot 0{,}10\ \text{mm}\left(\frac{1{,}96 + 2{,}33}{0{,}20\ \text{mm}}\right)^2 0{,}03\ \text{mm}\right]^2}$$

$$= \sqrt{3 \cdot 10^{-4} + 7{,}6}$$
$$= 3.$$

Der weitaus größere Beitrag zur Unsicherheit in N kommt in diesem Beispiel von der Unsicherheit bezüglich der Standardabweichung s*. Die Verwendung der auf zwei Nachkommastellen gerundeten Quantilwerte aus Tabelle 1 ist demgegenüber völlig ausreichend. Der Stichprobenumfang selbst sollte angegeben werden als

N = 8 mit einer kombinierten Standardunsicherheit von 3.

Berücksichtigt man außerdem, dass Quantilwerte in der Praxis häufig mit einem Tabellenkalkulations- oder Statistikprogramm bestimmt werden können, sodass sie mit noch geringerer Unsicherheit vorliegen als in Tabelle 1, kann Gleichung 11 auch vereinfacht werden zu

$$u(N) = 2\, s^* \left(\frac{t_{1-\alpha/2} + t_{1-\beta}}{\Delta} \right)^2 u(s^*)\,. \tag{12}$$

4 Zweistichproben-t-Test

4.1 Bestimmung des optimalen Stichprobenumfangs

Der Zweistichproben-t-Test dient der Überprüfung, ob die Erwartungswerte $E(X_1)$ und $E(X_2)$ zweier normalverteilter Zufallsvariablen X_1 und X_2 gleicher Varianz übereinstimmen. In seiner zweiseitigen Variante wird er mit den folgenden Hypothesen durchgeführt:
H_0: $E(X_1) = E(X_2)$
H_1: $E(X_1) \neq E(X_2)$.

Der Wert der Teststatistik bzw. Prüfwert wird berechnet als

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}\,. \tag{13}$$

Im Folgenden wird der Spezialfall zweier gleichgroßer Stichproben mit gleichgroßen empirischen Standardabweichungen betrachtet. Mit $N_1 = N_2 = N$ und $s_1 = s_2 = s$ ist

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{2\,s^2}{N}}}$$
$$= \sqrt{\frac{N}{2}}\, \frac{\bar{x}_1 - \bar{x}_2}{s}\,. \tag{14}$$

H_0 wird beibehalten, falls $t_{\alpha/2} \le t \le t_{1-\alpha/2}$, d. h.

$|t| \le t_{1-\alpha/2}$

$$\sqrt{\frac{n}{2}}\,\frac{|\,\bar{x}_1 - \bar{x}_2\,|}{s} \le t_{1-\alpha/2}$$

$$|\,\bar{x}_1 - \bar{x}_2\,| \le Q_{1-\alpha/2} \tag{15}$$

Darin ist $t_{1-\alpha/2}$ das (1−α/2)-Quantil der t-Verteilung mit dem Freiheitsgrad f = 2 N – 2 und

$$Q_{1-\alpha/2} = \sqrt{\frac{2}{N}}\; s\, t_{1-\alpha/2} \tag{16}$$

das (1−α/2)-Quantil derjenigen Verteilung, welche die Zufallsvariable $\bar{X}_1 - \bar{X}_2$ beschreibt.

Die weitere Argumentation entspricht derjenigen in Abschnitt 3, nur dass X durch X_1 und μ_0 durch $E(X_2)$ zu ersetzen ist.

Angenommen, die Nullhypothese H_0 gilt, d. h. es ist $E(X_1) = E(X_2)$. Dann ist der Erwartungswert der Variable $\bar{X}_1 - \bar{X}_2$

$$\begin{aligned} E(\bar{X}_1 - \bar{X}_2) &= E(\bar{X}_1) - E(\bar{X}_2) \\ &= E(X_1) - E(X_2) \\ &= 0 \end{aligned}$$

und die Variable wird durch die in Abbildung 4 gestrichelt gezeichnete linke Wahrscheinlichkeitsdichtefunktion beschrieben.

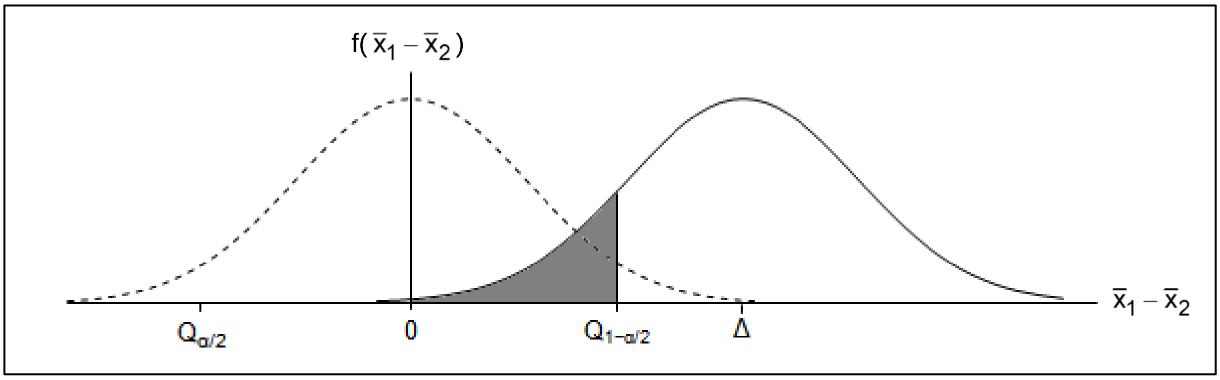

Abbildung 4: Wahrscheinlichkeit, H_0 irrtümlich beizubehalten, falls $E(X_1) - E(X_2) = \Delta$.

Angenommen, die Nullhypothese H_0 gilt nicht und die Erwartungswerte $E(X_1)$ und $E(X_2)$ weichen um Δ voneinander ab: $E(X_1) = E(X_2) + \Delta$. In diesem Fall ist

$$\begin{aligned} E(\bar{X}_1 - \bar{X}_2) &= E(\bar{X}_1) - E(\bar{X}_2) \\ &= E(X_1) - E(X_2) \\ &= \Delta \end{aligned}$$

und die Variable $\bar{X}_1 - \bar{X}_2$ wird durch die in Abbildung 4 durchgezogen gezeichnete rechte Wahrscheinlichkeitsdichtefunktion beschrieben. Es gilt

(1−α/2)-Quantil $Q_{1-\alpha/2}$ der linken Dichtefunktion = β-Quantil der rechten Dichtefunktion
= β-Quantil der linken Dichtefunktion + Δ

= –(1– β)-Quantil der linken Dichtefunktion + Δ.

Mathematisch lässt sich das entsprechend Gleichung 16 auch wie folgt ausdrücken:

$$\sqrt{\frac{2}{N}}\ s\ t_{1-\alpha/2} = -\sqrt{\frac{2}{N}}\ s\ t_{1-\beta} + \Delta$$

Durch Umformung ergibt sich

$$\Delta \sqrt{\frac{N}{2}} = s\ (t_{1-\alpha/2} + t_{1-\beta})$$

$$N = \frac{2\,s^2\,(t_{1-\alpha/2} + t_{1-\beta})^2}{\Delta^2}\,. \qquad (17)$$

Gegeben seien
- die Wahrscheinlichkeiten α und β für die Fehler erster und zweiter Art
- eine untere Begrenzung s* für die empirische Standardabweichung s
- die Mindestabweichung Δ zwischen den empirischen Mittelwerten, die als signifikant gelten soll.

Es ist $\lim\limits_{N\to\infty} t_{1-\alpha/2} = t_{1-\alpha/2}{*}$ und $\lim\limits_{N\to\infty} t_{1-\beta} = t_{1-\beta}{*}$.

In der Realität sind Stichprobenumfänge aber endlich und es gilt daher $t_{1-\alpha/2} > t_{1-\alpha/2}*$ und $t_{1-\beta} > t_{1-\beta}*$. Setzt man dies in Gleichung 17 ein, so ergibt sich

$$N \geq \frac{2\,s{*}^2\,(t_{1-\alpha/2}{*} + t_{1-\beta}{*})^2}{\Delta^2}\,. \qquad (18)$$

Beispiel: Es soll untersucht werden, ob sich das mittlere Geburtsgewicht von Kindern aus den Ländern A und B unterscheidet. Mit der Irrtumswahrscheinlichkeit α = 0,05 und der Teststärke 1 – β = 0,95 sollen Unterschiede Δ ≥ 100 g identifiziert werden können. Als Standardabweichung des Gewichts kann s* = 570 g angesetzt werden. Welchen Umfang müssen die beiden Stichproben jeweils haben?

$t_{1-\alpha/2}* = t_{0,975}* = 1{,}96$ $\qquad$ $t_{1-\beta}* = t_{0,95}* = 1{,}64$

Die Quantilwerte sind Tabelle 1 entnommen. Nach Beziehung 18 ist

$$N \geq \frac{2 \cdot (570\ \text{g})^2\,(1{,}96 + 1{,}64)^2}{(100\ \text{g})^2} \geq 843.$$

Die iterative Ermittlung des optimalen Stichprobenumfangs[3] liefert die folgenden Ergebnisse:

[3] Nähere Erläuterungen zum Prinzip des Verfahrens im Beispiel des Abschnitts 2.

N	rechte Seite von Beziehung 18
843	$\frac{2 \cdot (570\ g)^2\ (1{,}96 + 1{,}64)^2}{(100\ g)^2} = 847$
847	$\frac{2 \cdot (570\ g)^2\ (1{,}96 + 1{,}64)^2}{(100\ g)^2} = 847$

N = 847

Dieser Wert weist allerdings eine erhebliche Unsicherheit auf, die im nachfolgenden Abschnitt 4.2 analysiert wird.

Weitere Übungsaufgaben finden sich beispielweise im Internet unter der Adresse aufgabomat.de in der Rubrik Statistik.

4.2 Unsicherheitsanalyse

Drei Größen sind bei der Bestimmung des Stichprobenumfangs N nach Beziehung 18 unsicher, das (1 – α/2)-Quantil, das (1 – β)-Quantil und die Standardabweichung. Die Unsicherheit dieser drei Größen pflanzt sich in den berechneten Stichprobenumfang fort. In welchem Maße dies der Fall ist und von welcher Eingangsgröße der Berechnung der größte Beitrag zur Unsicherheit des Stichprobenumfangs kommt, wird in der folgenden Analyse geklärt. Die Grundlagen der Unsicherheitsanalyse lassen sich beispielsweise dem Skript „Messunsicherheit“ unter aufgabomat.de entnehmen.

Nach dem Unsicherheitsfortpflanzungsgesetz ergibt sich für die kombinierte Standardunsicherheit des Stichprobenumfangs

$$u_C(N) = \sqrt{\left[\frac{\partial N}{\partial t_{1-\alpha/2}}\, u(t_{1-\alpha/2})\right]^2 + \left[\frac{\partial N}{\partial t_{1-\beta}}\, u(t_{1-\beta})\right]^2 + \left[\frac{\partial N}{\partial s^*}\, u(s^*)\right]^2}$$

$$= \sqrt{\left[4\left(\frac{s^*}{\Delta}\right)^2 (t_{1-\alpha/2} + t_{1-\beta})\right]^2 \left\{[u(t_{1-\alpha/2})]^2 + [u(t_{1-\beta})]^2\right\} + \left[4\, s^* \left(\frac{t_{1-\alpha/2} + t_{1-\beta}}{\Delta}\right)^2 u(s^*)\right]^2}\ . \qquad (19)$$

Darin sind $u(t_{1-\alpha/2})$, $u(t_{1-\beta})$ und $u(s^*)$ die Standardunsicherheiten der Quantile und der Standardabweichung.

Beispiel: Vergleich des mittleren Geburtsgewichts (Abschnitt 4.1)

Für das (1–α/2)-Quantil und das (1–β)-Quantil können bei Verwendung von Tabelle 1 nur auf zwei Nachkommastellen genaue Werte $t_{1-\alpha/2}$ und $t_{1-\beta}$ ermittelt werden. Da zudem mit der Unsicherheit bezüglich des Stichprobenumfangs eine Unsicherheit bezüglich des Freiheitsgrades der t-Verteilung einhergeht, wird für die maximale Unsicherheit dieser Werte $\Delta t_{1-\alpha/2} = \Delta t_{1-\beta} = 0{,}01$ angesetzt. Unter der Annahme, dass die tatsächlichen Quantilwerte mit gleichverteilter Wahrscheinlichkeit in den Intervallen $[t_{1-\alpha/2} - \Delta t_{1-\alpha/2};\ t_{1-\alpha/2} + \Delta t_{1-\alpha/2}]$ und $[t_{1-\beta} - \Delta t_{1-\beta};\ t_{1-\beta} + \Delta t_{1-\beta}]$ liegen, ergibt sich ihre Standardunsicherheit als

$$u(t_{1-\alpha/2}) = \frac{0{,}01}{\sqrt{3}} = 0{,}006 \qquad u(t_{1-\beta}) = \frac{0{,}01}{\sqrt{3}} = 0{,}006.$$

Die maximale Abweichung Δs* der Standardabweichung hängt vom betrachteten Anwendungsfall ab. Im Beispiel des Abschnitts 4.1 könnte Δs* = 50 g angesetzt werden, sodass die Standardunsicherheit den Wert

$$
\begin{aligned}
u(s^*) &= \frac{50\text{ g}}{\sqrt{3}} \\
&= 29\text{ g}
\end{aligned}
$$

annimmt. Es folgt

$$
\begin{aligned}
u_C(N) &= \sqrt{\left[4\left(\frac{570\text{ g}}{100\text{ g}}\right)^2 (1{,}96+1{,}64)\right]^2 2\cdot 0{,}006^2 + \left[4\cdot 570\text{ g}\left(\frac{1{,}96+1{,}64}{100\text{ g}}\right)^2 29\text{ g}\right]^2} \\
&= \sqrt{16+7343} \\
&= 86.
\end{aligned}
$$

Der weitaus größere Beitrag zur Unsicherheit in N kommt in diesem Beispiel von der Unsicherheit bezüglich der Standardabweichung s*. Die Verwendung der auf zwei Nachkommastellen gerundeten Quantilwerte aus Tabelle 1 ist demgegenüber völlig ausreichend. Der Stichprobenumfang selbst sollte angegeben werden als

N = 847 mit einer kombinierten Standardunsicherheit von 86.

Berücksichtigt man außerdem, dass Quantilwerte in der Praxis häufig mit einem Tabellenkalkulations- oder Statistikprogramm bestimmt werden können, sodass sie mit noch geringerer Unsicherheit vorliegen als in Tabelle 1, kann Gleichung 19 auch vereinfacht werden zu

$$
u(N) = 4\, s^* \left(\frac{t_{1-\alpha/2} + t_{1-\beta}}{\Delta}\right)^2 u(s^*)\,. \tag{20}
$$

veröffentlicht im Internet unter aufgabomat.de