Statistische Intervalle

Andreas Handl

Inhaltsverzeichnis

1 Statistische Intervalle

Bei der Veröffentlichung der Ergebnisse statistischer Erhebungen werden vielfach Punktschätzer angegeben. So ist beim Umwelt- und Prognose-Institut am 16.01.2004 folgende Aussage zu finden:

> Die durchschnittliche Fahrleistung des Autofahrers liegt seit Jahren stabil bei 12000 Kilometern im Jahr.

Dieser Wert ist eine Schätzung. Dies merkt man schon daran, dass es sich um eine runde Zahl handelt. Es fehlt aber eine Genauigkeitsangabe in Form der Varianz. Oft wird ein Intervall angegeben. Dies trägt dem Umstand Rechnung, dass die Schätzung fehlerbehaftet ist. Wir wollen uns im Folgenden Intervalle anschauen.
Dabei gehen wir von einer Zufallsstichprobe $x_1, \ldots, x_n$ aus einer Grundgesamtheit mit stetiger Verteilungsfunktion $F_X(x)$ aus. Die Beobachtungen $x_1, \ldots, x_n$ sind also Realisationen der unabhängigen, identisch mit Verteilungsfunktion $F_X(x)$ verteilten Zufallsvariablen $X_1, \ldots, X_n$.
Wir betrachten im Folgenden das Intervall

$$\left[x_{(1)}, x_{(n)}\right] \tag{1}$$

Dabei ist $x_{(1)}$ die kleinste und $x_{(n)}$ die größte Beobachtung in der Stichprobe.

Beispiel 1
Ein Arbeitnehmer hat eine neue Stelle angenommen. Er will wissen, welche Charakteristika die Fahrzeit zur Arbeitsstelle besitzt. Deshalb notiert er die Fahrzeit an 9 aufeinander folgenden Tagen. Hier sind die Werte in Sekunden:

```
1670 1775 1600 1700 2000 1890 1740 1880 1945
```

Es gilt $x_{(1)} = 1600$ und $x_{(9)} = 2000$. Wir betrachten also das Intervall

$$[1600, 2000] \tag{2}$$

In Abhängigkeit von der Fragestellung kann man das Intervall in Gleichung (1) unterschiedlich interpretieren.
Ist man an einem Parameter wie dem Erwartungswert $E(X)$ oder dem Median $X_{0.5}$ interessiert, so wird man ein Konfidenzintervall aufstellen.

Definition 1.1
Seien $X_1, \ldots, X_n$ unabhängige, identisch verteilte Zufallsvariablen, deren Verteilungsfunktion $F_X(x)$ von einem Parameter θ abhängt. Außerdem seien

$T_1 = g_1(X_1, \ldots, X_n)$ und $T_2 = g_2(X_1, \ldots, X_n)$ zwei Stichprobenfunktionen. Dann heißt das Intervall

$$[T_1, T_2] \tag{3}$$

mit

$$P(T_1 \leq \theta \leq T_2) = 1 - \alpha \tag{4}$$

zweiseitiges Konfidenzintervall für θ zum **Konfidenzniveau** $1 - \alpha$.

Ist der Parameter θ gleich dem Median $X_{0.5}$, so können wir das Intervall in Gleichung (1) auf Seite 2 als Konfidenzintervall für den Median $X_{0.5}$ auffassen. Es ist einfach, das Konfidenzniveau des Intervalls $\left[x_{(1)}, x_{(n)}\right]$ bestimmen. Es gilt

$$P\left(X_{(1)} \leq X_{0.5} \leq X_{(n)}\right) = 1 - \left[P\left(X_{(1)} > X_{0.5}\right) + P\left(X_{(n)} < X_{0.5}\right)\right]$$

Nun gilt

$$\begin{aligned} P\left(X_{(1)} > X_{0.5}\right) &= P\left(\text{alle}\, X_i > X_{0.5}\right) = P\left(X_1 > X_{0.5}, \ldots, X_n > X_{0.5}\right) \\ &= P\left(X_1 > X_{0.5}\right) \cdots P\left(X_n > X_{0.5}\right) = 0.5^n \end{aligned}$$

Analog erhalten wir

$$P\left(X_{(n)} < X_{0.5}\right) = 0.5^n$$

Somit gilt

$$P\left(X_{(1)} \leq X_{0.5} \leq X_{(n)}\right) = 1 - 2 \cdot 0.5^n = 1 - 0.5^{n-1}$$

Beispiel 1 (fortgesetzt von Seite 2)
Der Arbeitnehmer ist an der mittleren Fahrzeit interessiert, die er mit dem Median misst. Das Konfidenzintervall für den Median ist in Gleichung (2) auf Seite 2 zu finden. Das Konfidenzniveau beträgt

$$1 - \alpha = 1 - 0.5^8 = 0.996$$

Wir wollen verdeutlichen, wie das Konfidenzniveau zu interpretieren ist. Für ein konkretes Intervall gibt es nur zwei Möglichkeiten. Der unbekannte Wert des Parameters liegt in dem Intervall oder er liegt nicht in dem Intervall. Wir wissen nicht, welche der beiden Möglichkeiten zutrifft. Wir können aber die Wahrscheinlichkeit bestimmen, dass wir ein Intervall gefunden haben, das den Wert des Parameters überdeckt. Hierzu führen wir eine Simulation durch.

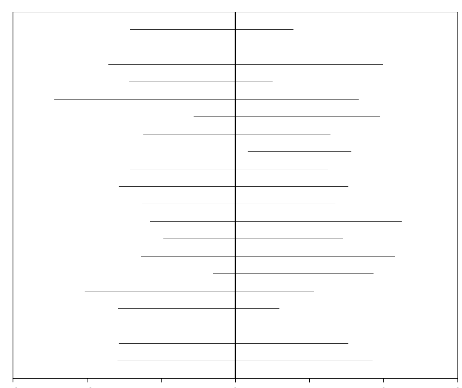
Wir unterstellen Standardnormalverteilung und ziehen 20 Stichproben vom Umfang $n = 9$. Für jede dieser Stichproben stellen wir das Konfidenzintervall aus Gleichung (1) auf Seite 2 für $X_{0.5}$ auf. Die Daten sind in Tabelle 1 auf 25 zu finden.

Die erste simulierte Stichprobe lautet:

```
1.05  1.07 -0.12  0.03 -1.59  0.35  1.85  0.40  0.36
```

Das Intervall ist $[-1.59, 1.85]$. Der Wert von $X_{0.5}$ ist uns bekannt. Er ist 0. Das Intervall enthält den Wert von $X_{0.5}$. Abbildung 1 verdeutlicht dies und zeigt die anderen 19 Konfidenzintervalle.

Abbildung 1: 20 Konfidenzintervalle



Wir sehen, dass 19 Konfidenzintervalle den Wert 0 enthalten.

Das Konfidenzniveau bezieht sich auf den Prozess und nicht das konkrete Intervall. Stellen wir sehr viele Konfidenzintervalle zum Konfidenzniveau $1-\alpha$ auf, so erwarten wir, dass $100 \cdot (1-\alpha)$ Prozent den wahren Wert des Parameters überdecken. Ist $1-\alpha$ nahe bei 1, so können wir uns bei einem konkreten Intervall also ziemlich sein, dass es den wahren Wert des Parameters enthält.

Im Beispiel haben wir den Stichprobenumfang vorgegeben. Wollen wir das Intervall $\left[x_{(1)}, x_{(n)}\right]$ als Konfidenzintervall zum vorgegebenen Konfidenzniveau $1-\alpha$ aufstellen, so bestimmen wir den Stichprobenumfang n folgendermaßen:

$$n \geq 1 + \frac{\ln(\alpha)}{\ln 0.5}$$

Dies sieht man folgendermaßen:

$$1 - 0.5^{n-1} \geq 1 - \alpha \iff 0.5^{n-1} \leq \alpha \iff (n-1)\ln 0.5 \leq \ln(\alpha)$$
$$\iff (n-1) \geq \frac{\ln(\alpha)}{\ln 0.5} \iff n \geq 1 + \frac{\ln(\alpha)}{\ln 0.5}$$

Beispiel 2

Wollen wir das Intervall $\left[x_{(1)}, x_{(n)}\right]$ als Konfidenzintervall zum Konfidenzniveau 0.95, so benötigen wir eine Stichprobe vom Umfang 6, denn

$$n \geq 1 + \frac{\ln(1 - 0.95)}{\ln 0.5} = 5.32$$

Wir können auch einseitige Konfidenzintervalle aufstellen.

Definition 1.2

Seien $X_1, \ldots, X_n$ unabhängige, identisch verteilte Zufallsvariablen, deren Verteilungsfunktion $F_X(x)$ von einem Parameter θ abhängt. Außerdem seien $T_1 = g_1(X_1, \ldots, X_n)$ und $T_2 = g_2(X_1, \ldots, X_n)$ zwei Stichprobenfunktionen. Dann heißen die Intervalle

$$(-\infty, T_2] \tag{5}$$

mit

$$P(\theta \leq T_2) = 1 - \alpha \tag{6}$$

und

$$[T_1, \infty) \tag{7}$$

mit

$$P(T_1 \leq \theta) = 1 - \alpha \tag{8}$$

einseitige Konfidenzintervalle für θ zum Konfidenzniveau $1 - \alpha$.

Einseitige Konfidenzintervalle werden verwendet, wenn man angeben will, welchen Wert ein Parameter mit einer vorgegebenen Wahrscheinlichkeit $1-\alpha$ mindestens oder höchstens annehmen kann.

Die Intervalle

$$\left[x_{(1)}, \infty\right) \tag{9}$$

und

$$\left(-\infty, x_{(n)}\right] \tag{10}$$

kann man als einseitige Konfidenzintervalle für den Median auffassen. Für beide Intervalle gilt

$$1 - \alpha = 1 - 0.5^n$$

Ein Konfidenzintervall ist ein Intervall für einen Parameter. Will man aber einen zukünftigen Wert x_{n+1} einer Zufallsvariablen auf Basis einer Stichprobe $x_1, \ldots, x_n$ vorhersagen, so spricht man von Prognose. Wie bei der Schätzung kann man auch bei der Prognose Intervalle für den zukünftigen Wert angeben.

Definition 1.3
Seien $X_1, \ldots, X_n, X_{n+1}$ unabhängige, identisch mit Verteilungsfunktion $F_X(x)$ verteilte Zufallsvariablen. Außerdem seien $T_1 = g_1(X_1, \ldots, X_n)$ und $T_2 = g_2(X_1, \ldots, X_n)$ zwei Stichprobenfunktionen.
Dann heißt das Intervall

$$[T_1, T_2] \tag{11}$$

mit

$$P\left(T_1 \leq X_{n+1} \leq T_2\right) = 1 - \alpha \tag{12}$$

Prognoseintervall für X_{n+1} zur **Sicherheit** $1 - \alpha$.

Wir können das Intervall $\left[x_{(1)}, x_{(n)}\right]$ als Prognoseintervall auffassen. Es gilt für jede stetige Verteilung

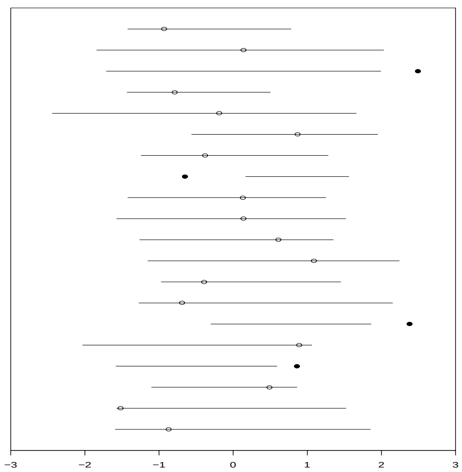
$$1 - \alpha = \frac{n-1}{n+1} \tag{13}$$

Bevor wir uns überlegen, warum $1 - \alpha$ diesen Wert annimmt, wollen wir zunächst $1 - \alpha$ interpretieren.
Wie auch bei Konfidenzintervallen gibt es bei Prognoseintervallen für ein konkretes Intervall $[t_1, t_2]$ zwei Möglichkeiten. Die Realisation x_{n+1} liegt im Intervall $[t_1, t_2]$ oder sie liegt nicht im Intervall $[t_1, t_2]$. Die Wahrscheinlichkeitsaussage bezieht sich auch hier nicht auf das Intervall sondern auf die Prozedur. Wenn wir sehr viele Stichproben $x_1, \ldots, x_n, x_{n+1}$ vom Umfang $n+1$ ziehen und für $x_1, \ldots, x_n$ das Prognoseintervall bestimmen, so erwarten wir, dass in $100 \cdot 1 - \alpha$ Prozent der Prognoseintervalle der Wert x_{n+1} liegt.
Wir wollen dies für das Intervall $\left[x_{(1)}, x_{(n)}\right]$ mit einer Simulation veranschaulichen. Wir unterstellen Standardnormalverteilung und ziehen 20 Stichproben vom Umfang $n = 9$. Für jede dieser Stichproben stellen wir das Prognoseintervall $\left[x_{(1)}, x_{(n)}\right]$ auf. Die Daten sind in Tabelle 1 auf 25 zu finden. Die erste simulierte Stichprobe lautet:

```
1.05  1.07 -0.12  0.03 -1.59  0.35  1.85  0.40  0.36
```

Das Intervall ist $[-1.59, 1.85]$. Der Wert von x_{10} ist -0.87. Das Intervall enthält den Wert von x_{10}. Abbildung 2 verdeutlicht dies. Außerdem sind noch die anderen 19 Prognoseintervalle zu finden.

Abbildung 2: 20 Prognoseintervalle



Wir sehen, dass 16 Prognoseintervalle den Wert x_{n+1} enthalten.

Beispiel 1 (fortgesetzt von Seite 3)
Will er nur wissen, wie lange er bei der nächsten Fahrt unterwegs ist, so wird er eine Prognose bestimmen. Ein Prognoseintervall sagt ihm, zwischen welchen Grenzen die nächste Fahrzeit liegen wird. Verwendet er das Intervall $[1600, 2000]$ als Prognoseintervall, so gilt

$$1 - \alpha = \frac{9-1}{9+1} = 0.8$$

Wenden wir uns jetzt der Frage zu, warum Gleichung (13) auf Seite 6 gilt. Wenn die Verteilungsfunktion $F_X(x)$ bekannt ist, gilt

$$P\left(X_{(1)} \leq X \leq X_{(n)}\right) = P\left(F_X\left(X_{(1)}\right) \leq F_X(X) \leq F_X\left(X_{(n)}\right)\right) \tag{14}$$

Besitzt X die Verteilungsfunktion $F_X(x)$, so ist die Zufallsvariable $F_X(X)$ gleichverteilt auf $(0, 1)$. Der Beweis ist bei Randles and Wolfe (1979) auf der Seite 6 zu finden. $F_X(X_{(i)})$ ist also die i-te Orderstatistik einer Zufallsstichprobe vom Umfang n aus einer Gleichverteilung auf $(0, 1)$. Diese bezeichnen wir mit $U_{(i)}$. Bickel and Doksum (2001) zeigen auf Seite 254:

$$P\left(X_{(1)} \leq X \leq X_{(n)}\right) = E\left(U_{(n)}\right) - E\left(U_{(1)}\right)$$

Es gilt

$$E\left(U_{(i)}\right)=\frac{i}{n+1}$$

(siehe dazu Randles and Wolfe (1979), S. 7). Hieraus folgt

$$E\left(U_{(n)}-U_{(1)}\right)=\frac{n}{n+1}-\frac{1}{n+1}=\frac{n-1}{n+1}$$

Im Beispiel haben wir den Stichprobenumfang vorgegeben. Wollen wir das Intervall in Gleichung (1) auf Seite 2 als Prognoseintervall zum vorgegebenen Prognoseniveau $1-\alpha$ aufstellen, so bestimmen wir den Stichprobenumfang n folgendermaßen:

$$n \geq \frac{2-\alpha}{\alpha}$$

Dies sieht man folgendermaßen:

$$\frac{n-1}{n+1} \geq 1-\alpha \iff n-1 \geq (n+1)(1-\alpha) \iff n-1 \geq n(1-\alpha)+1-\alpha$$

$$\iff n\,\alpha \geq 2-\alpha \iff n \geq \frac{2-\alpha}{\alpha}$$

Beispiel 3
Verwenden wir das Intervall $\left[x_{(1)}, x_{(n)}\right]$ als Prognoseintervall zur Sicherheit 0.95, so benötigen wir eine Stichprobe vom Umfang 39, denn

$$n \geq \frac{2-0.05}{0.05}=39$$

Wir können auch einseitige Prognoseintervalle aufstellen.

Definition 1.4
Seien $X_1, \ldots, X_n, X_{n+1}$ unabhängige, identisch mit Verteilungsfunktion $F_X(x)$ verteilte Zufallsvariablen. Außerdem seien $T_1 = g_1(X_1, \ldots, X_n)$ und $T_2 = g_2(X_1, \ldots, X_n)$ zwei Stichprobenfunktionen. Dann heißen die Intervalle

$$(-\infty, T_2] \tag{15}$$

mit

$$P\left(X_{n+1} \leq T_2\right)=1-\alpha \tag{16}$$

und

$$[T_1, \infty) \tag{17}$$

mit

$$P\left(T_1 \leq X_{n+1}\right) = 1 - \alpha \tag{18}$$

einseitige Prognoseintervalle für x_{n+1} zur Sicherheit $1 - \alpha$.

Einseitige Prognoseintervalle werden verwendet, wenn man angeben will, welchen Wert eine zukünftige Beobachtung mit einer vorgegebenen Wahrscheinlichkeit $1 - \alpha$ mindestens oder höchstens annehmen kann. Die Intervalle

$$\left[x_{(1)}, \infty\right) \tag{19}$$

und

$$\left(-\infty, x_{(n)}\right] \tag{20}$$

kann man als einseitige Prognoseintervalle auffassen. Für beide Intervalle gilt

$$1 - \alpha = \frac{n}{n+1}$$

Hahn (1970) nennt das Prognoseintervall auch das Astronauten-Intervall. Ein Astronaut interessiert sich nur für den nächsten Flug und will wissen, welche Werte der interessierenden Merkmale er erwarten kann. Der Hersteller eines Produktes ist aber nicht nur an einer Beobachtung interessiert, sondern an der gesamten zukünftigen Produktion. Er will ein Intervall angeben, in dem sich mindestens der Anteil p aller zukünftigen Beobachtungen befindet.

Definition 1.5
Seien $X_1, \ldots, X_n$ unabhängige, identisch mit Verteilungsfunktion $F_X(x)$ verteilte Zufallsvariablen.
Außerdem seien $T_1 = g_1(X_1, \ldots, X_n)$ und $T_2 = g_2(X_1, \ldots, X_n)$ zwei Stichprobenfunktionen.
Dann heißt das Intervall

$$[T_1, T_2] \tag{21}$$

mit

$$P\left(P(T_1 \leq X_{n+1} \leq T_2) \geq p\right) = 1 - \alpha \tag{22}$$

Toleranzintervall für den Mindestanteil p zur Sicherheit $1 - \alpha$.

Wie können wir diese Wahrscheinlichkeit interpretieren?
Das Toleranzintervall soll mindestens $100 \cdot p$ Prozent aller Realisationen der Zufallsvariablen enthalten. Für ein konkretes Intervall gibt es zwei Möglichkeiten. Es enthält mindestens die Wahrscheinlichkeitsmasse p oder es enthält sie nicht.
Wir wollen dies für das Intervall $\left[x_{(1)}, x_{(n)}\right]$ mit einer Simulation veranschaulichen. Wir unterstellen Standardnormalverteilung und ziehen 20 Stichproben vom Umfang $n = 9$. Für jede dieser Stichproben stellen wir das Konfidenzintervall aus Gleichung (1) auf Seite 2 für $X_{0.5}$ auf. Die Daten sind in Tabelle 1 auf 25 zu finden. Nehmen wir an, die erste simulierte Stichprobe lautet

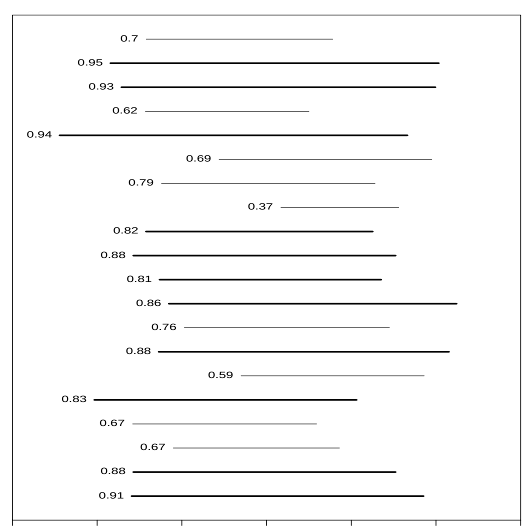
```
1.05  1.07 -0.12  0.03 -1.59  0.35  1.85  0.40  0.36
```

Das Intervall ist $[-1.59, 1.85]$. Sei $p = 0.9$. Ist X standardnormalverteilt, so gilt

$$P(-1.59 \leq X \leq 1.85) = \Phi(1.85) - \Phi(-1.59) = 0.9119.$$

Das Intervall $[-1.59, 1.85]$ enthält $100 \cdot 0.9119$ Prozent aller Realisationen einer standardnormalverteilten Zufallsvariablen. Abbildung 3 verdeutlicht dies und zeigt die 19 anderen Intervalle. Wir sehen, dass 11 Toleranzintervalle, also 55 Prozent mindestens die Wahrscheinlichkeitsmasse 0.8 enthalten.

Abbildung 3: 20 Toleranzzintervalle



Bei einem Toleranzintervall bezieht sich die innere Wahrscheinlichkeitsaussage in Gleichung (22) auf Seite 9 auf das Intervall, während sich die äußere Wahrscheinlichkeitsaussage auf die Prozedur bezieht. Wenn wir sehr viele Stichproben $x_1, \ldots, x_n$ und für jede das Toleranzintervall $[t_1, t_2]$ bestimmen, so erwarten wir, dass $100 \cdot (1 - \alpha)$ Prozent der Intervalle mindestens die Wahrscheinlichkeitsmasse p enthalten.

Fassen wir das Intervall $\left[x_{(1)}, x_{(n)}\right]$ als Toleranzintervall mit Mindestanteil p auf, so gilt für jede stetige Verteilung

$$1 - \alpha = 1 - p^n - n(1-p)p^{n-1} \tag{23}$$

Diese Beziehung wird in Mood et al. (1974) auf den Seiten 516-517 hergeleitet.

Beispiel 1 (fortgesetzt von Seite 7)

Der Arbeitnehmer ist an einem Toleranzintervall mit Mindestanteil 0.8 interessiert. Verwendet er das Intervall $[1600, 2000]$ als Toleranzintervall mit Mindestanteil 0.8, so beträgt die Sicherheit

$$1 - \alpha = 1 - 0.8^9 - 9 \cdot 0.2 \cdot 0.8^9 = 0.624$$

Im Beispiel haben wir den Stichprobenumfang vorgegeben. Wollen wir das Intervall $\left[x_{(1)}, x_{(n)}\right]$ als Toleranzintervall mit Mindestanteil p zur vorgegebenen Sicherheit $1 - \alpha$ aufstellen, so müssen wir die Gleichung (23) nach n auflösen. Dies ist nicht explizit möglich. Eine gute Approximation ist bei Conover (1999) auf Seite 151 zu finden. Sie lautet

$$n \geq \frac{1+p}{4 \cdot (1-p)} \cdot \chi^2_{1-\alpha,4} + 0.5 \tag{24}$$

Dabei ist $\chi^2_{1-\alpha,4}$ das $1 - \alpha$-Quantil der χ^2-Verteilung mit 4 Freiheitsgraden.

Beispiel 4

Wollen wir das Intervall $\left[x_{(1)}, x_{(n)}\right]$ als Toleranzintervall für den Mindesanteil 0.9 zur Sicherheit 0.95 verwenden, so benötigen wir eine Stichprobe vom Umfang 46, denn

$$n \geq \frac{1+0.9}{4 \cdot (1-0.9)} \cdot 9.49 + 0.5 = 45.6$$

Setzen wir $n = 46$ und $p = 0.9$ in Gleichung (23) ein, so erhalten wir

$$1 - \alpha = 1 - 0.9^{46} - 46 \cdot 0.1 \cdot 0.9^{45} = 0.952$$

Für $n = 45$ und $p = 0.9$ gilt

$$1 - \alpha = 1 - 0.9^{45} - 45 \cdot 0.1 \cdot 0.9^{44} = 0.948$$

Wir können auch einseitige Toleranzintervalle aufstellen.

Definition 1.6
Seien $X_1, \ldots, X_n$ unabhängige, identisch mit Verteilungsfunktion $F_X(x)$ verteilte Zufallsvariablen.
Außerdem seien $T_1 = g_1(X_1, \ldots, X_n)$ und $T_2 = g_2(X_1, \ldots, X_n)$ zwei Stichprobenfunktionen. Dann heißen die Intervalle

$$(-\infty, T_2] \tag{25}$$

mit

$$P(P\left(X \leq T_2\right) \geq p) = 1 - \alpha \tag{26}$$

und

$$[T_1, \infty) \tag{27}$$

mit

$$P(P\left(X \geq T_1\right) \geq p) = 1 - \alpha \tag{28}$$

einseitige Toleranzintervalle für den Mindestanteil p zur Sicherheit $1-\alpha$.

Die Intervalle

$$\left[x_{(1)}, \infty\right) \tag{29}$$

und

$$\left(-\infty, x_{(n)}\right] \tag{30}$$

kann man als einseitige Toleranzintervalle auffassen.
Für beide Intervalle gilt

$$1 - \alpha = 1 - p^n$$

2 Intervalle bei Normalverteilung

Im letzten Kapitel haben wir uns mit verteilungsfreien Intervallen beschäftigt. Ist die Verteilung der Grundgesamtheit bekannt, so kann man Intervalle aufstellen, die bei gleichem Niveau wesentlich schmaler sind. Wir wollen im Folgenden nur Normalverteilung unterstellen. Wir gehen also davon aus, dass die Zufallsvariablen $X_1, \ldots, X_n$ unabhängig und mit den Parametern μ und σ^2 normalverteilt sind.

2.1 Konfidenzintervalle

Die Dichtefunktion der Normalverteilung hängt von den Parametern μ und σ^2 ab. Wir wollen im Folgenden Konfidenzintervalle für jeden der beiden Parameter herleiten. Wie man ein simultanes Konfidenzintervall für beide Parameter erhält, wird von Mood et al. (1974) auf den Seiten 384-385 beschrieben.

2.1.1 Konfidenzintervall für μ

Ist die Varianz der Grundgesamtheit bekannt, so ist das Konfidenzintervall für μ bei Normalverteilung gegeben durch:

$$\left[\bar{X} - z_{1-\alpha/2}\,\frac{\sigma}{\sqrt{n}}, \bar{X} + \, z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \tag{31}$$

Dabei ist $z_{1-\alpha/2}$ das $1-\alpha/2$-Quantil der Standardnormalverteilung.
Es gilt nämlich

$$P\left(\bar{X} - z_{1-\alpha/2}\,\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right)$$

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

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

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

$$= \Phi(z_{1-\alpha/2}) - \Phi(-z_{1-\alpha/2})$$

$$= 1 - \alpha$$

Schauen wir uns das Konfidenzintervall in Gleichung (31) unter praxisrelevanten Gesichtspunkten an. Bei einer Datenanalyse wird σ^2 in der Regel unbekannt sein. Es liegt nahe, σ^2 durch

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

zu schätzen und diesen Schätzer in Gleichung (31) für σ^2 einzusetzen. Das Intervall sieht also folgendermaßen aus:

$$\left[\bar{X} - z_{1-\alpha/2}\,\frac{S}{\sqrt{n}}, \bar{X} + \, z_{1-\alpha/2}\frac{S}{\sqrt{n}}\right] . \tag{32}$$

Für kleine Stichprobenumfänge gilt aber

$$P\left(\bar{X} - z_{1-\alpha/2}\,\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2}\,\frac{S}{\sqrt{n}}\right) \neq 1-\alpha\,. \tag{33}$$

Eine kleine Simulation zeigt dies. Wir erzeugen 5000 Stichproben vom Umfang 4 aus der Standardnormalverteilung, stellen für jede das Konfidenzintervall in Gleichung (32) auf und zählen, wie viele der Konfidenzintervalle den Wert 0 überdecken. Das geschätze Konfidenzniveau beträgt 0.8757. Die Konfidenzintervalle sind also im Mittel zu schmal. Um zu sehen, woran dies liegt, formen wir den Ausdruck in der Klammer auf der linken Seite von Gleichung (33) um:

$$\begin{aligned}
&\bar{X} - z_{1-\alpha/2}\,\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{1-\alpha/2}\,\frac{S}{\sqrt{n}} \\
\Longleftrightarrow\ & -z_{1-\alpha/2}\,\frac{S}{\sqrt{n}} \leq \mu - \bar{X} \leq z_{1-\alpha/2}\,\frac{S}{\sqrt{n}} \\
\Longleftrightarrow\ & -z_{1-\alpha/2}\,\frac{S}{\sqrt{n}} \leq \bar{X} - \mu \leq z_{1-\alpha/2}\,\frac{S}{\sqrt{n}} \\
\Longleftrightarrow\ & -z_{1-\alpha/2} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq z_{1-\alpha/2}
\end{aligned}$$

Wir müssen also folgende Wahrscheinlichkeit bestimmen:

$$P\left(-z_{1-\alpha/2} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq z_{1-\alpha/2}\right)$$

Die Zufallsvariable

$$t = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

ist nicht standardnormalverteilt, wenn die $X_1, \ldots, X_n$ unabhängig und mit den Parametern μ und σ^2 normalverteilt sind. Die Schätzung von σ^2 führt dazu, dass t stärker streut als die Standardnormalverteilung. Von Gossett (1908) wurde gezeigt, dass t eine t-Verteilung mit $n-1$ Freiheitsgraden besitzt. Ist

$t_{n-1;1-\alpha/2}$ das $1-\alpha/2$-Quantil der t-Verteilung mit $n-1$ Freiheitsgraden, so gilt

$$P\left(-t_{n-1;1-\alpha/2} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{n-1;1-\alpha/2}\right) = 1-\alpha \tag{34}$$

Wir formen den Ausdruck in der Klammer auf der linken Seite von Gleichung (34) so um, dass zwischen den Ungleichheitszeichen nur noch μ steht. Es gilt

$$P(\bar{X} - t_{n-1;1-\alpha/2}\,\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1;1-\alpha/2}\frac{S}{\sqrt{n}}) = 1-\alpha$$

Durch diese Umformung haben wir ein Konfidenzintervall für μ bei Normalverteilung mit unbekannten σ^2 gefunden. Es lautet:

$$\left[\bar{X} - t_{n-1;1-\alpha/2}\,\frac{S}{\sqrt{n}}, \bar{X} + t_{n-1;1-\alpha/2}\frac{S}{\sqrt{n}}\right]. \tag{35}$$

Dabei ist $t_{n-1;1-\alpha/2}$ das $1-\alpha/2$-Quantil der t-Verteilung mit $n-1$ Freiheitsgraden.

Beispiel 1 (fortgesetzt von Seite 11)
Wir unterstellen Normalverteilung und stellen das Konfidenzintervall für μ zum Konfidenzniveau 0.95 auf.
Es gilt $\bar{x} = 1800$ und $s = 135.4$. Mit $n = 9$ gilt also $s/\sqrt{n} = 45.13$. Der Tabelle der t-Verteilung entnehmen wir $t_{8;0.975} = 2.306$. Mit

$$t_{n-1;1-\alpha/2}\,\frac{s}{\sqrt{n}} = 104.1$$

Wir erhalten also folgendes Konfidenzintervall

$$[1695.9, 1904.1]$$

Das verteilungsfreie Konfidenzintervall $\left[x_{(1)}, x_{(n)}\right]$ lautet $[1600, 2000]$. Das Konfidenzniveau ist 0.996. Zum Vergleich stellen wir noch das Konfidenzintervall für μ zum Konfidenzniveau 0.996 auf. Es gilt $t_{8;0.998} = 3.99$. Also erhalten wir folgendes Konfidenzintervall

$$[1619.9, 1980.1]$$

Dieses ist schmaler als das verteilungsfreie Konfidenzintervall.

Schauen wir die Länge L des Konfidenzintervalls an. Es gilt

$$L = 2\, t_{n-1;1-\alpha/2}\, \frac{s}{\sqrt{n}} \tag{36}$$

Dies sieht man folgendermaßen

$$\begin{aligned} L &= \bar{x} + t_{n-1;1-\alpha/2} \frac{s}{\sqrt{n}} - \left(\bar{x} - t_{n-1;1-\alpha/2}\, \frac{s}{\sqrt{n}} \right) \\ &= 2\, t_{n-1;1-\alpha/2}\, \frac{s}{\sqrt{n}} \end{aligned}$$

Das Intervall wird breiter, wenn wir das Konfidenzniveau $1-\alpha$ vergrößern. Es wird aber nicht notwendigerweise schmaler, wenn wir den Stichprobenumfang erhöhen. Für eine neue Stichprobe werden wir auch einen anderen Wert von s erhalten, sodass das Intervall größer werden kann. Es ist auch nicht möglich den Mindeststichprobenumfang zu bestimmen, um eine vorgegebene Länge des Intervalls nicht zu überschreiten, da der Stichprobenumfang von s abhängt. Aus $L \leq l$ folgt nämlich

$$n \geq \frac{4\, t^2_{n-1;1-\alpha/2}\, s^2}{l^2}\,.$$

Die einseitigen Konfidenzintervalle für μ bei Normalverteilung sind.

$$\left(-\infty, \bar{X} + t_{n-1;1-\alpha} \frac{S}{\sqrt{n}} \right] \,. \tag{37}$$

und

$$\left[\bar{X} - t_{n-1;1-\alpha}\, \frac{S}{\sqrt{n}}, \infty \right) \,. \tag{38}$$

Dabei ist $t_{n-1;1-\alpha}$ das $1-\alpha$-Quantil der t-Verteilung mit $n-1$ Freiheitsgraden.

Beispiel 1 (fortgesetzt von Seite 15)
Wir suchen den Wert, den μ mit Wahrscheinlichkeit 0.95 mindestens annimmt.
Es gilt $t_{0.95,8} = 1.86$. Mit $\bar{x} = 1800$, $s = 135.4$ und $s/\sqrt{n} = 45.13$ ist dieser Wert also gleich

$$1800 - 1.86 \cdot 45.13 = 1716.1$$

2.1.2 Konfidenzintervall für σ^2

Wir gehen wiederum davon aus, dass Normalverteilung vorliegt und wollen ein Konfidenzintervall für σ^2 aufstellen. Um zu verstehen, wie man dies erhält, schauen wir uns noch einmal das Konfidenzintervall für μ an. Bei diesem sind wir von

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

ausgegangen. Diese Zufallsvariable ist mit $n - 1$ Freiheitsgraden t-verteilt, wenn die Zufallsvariablen $X_1, \ldots, X_n$ unabhängig und mit den Parametern μ und σ^2 normalverteilt sind. Somit gilt

$$P\left(-t_{n-1,1-\alpha/2} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{n-1,1-\alpha/2}\right) = 1 - \alpha$$

Formen wir den Ausdruck in der Klammer so um, dass μ isoliert ist, so haben wir das Konfidenzintervall für μ gefunden.
Die Vorgehensweise zeigt, wie man ein Konfidenzintervall für einen Parameter finden kann. Man benötigt eine Stichprobenfunktion,

1. die von dem unbekannten Parameter abhängt

2. deren Verteilung bekannt ist

Um ein Konfidenzintervall für σ^2 aufzustellen, liegt es nahe, von

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

auszugehen, da S^2 eine erwartungstreue Schätzfunktion von σ^2 ist. Multipliziert man S^2 mit $n - 1$ und dividiert die so gewonnene Größe durch σ^2, so erhält man

$$\frac{(n-1)\,S^2}{\sigma^2} \tag{39}$$

Die Zufallsvariable in Gleichung (39) genügt den Anfoderungen 1 und 2. Sie hängt von σ^2 ab und ist χ^2-verteilt mit $n - 1$ Freiheitsgraden. Dies zeigen Mood et al. (1974) auf den Seiten 243-245. Somit gilt

$$P\left(\chi^2_{n-1,\alpha/2} \leq \frac{(n-1)\,S^2}{\sigma^2} \leq \chi^2_{n-1,1-\alpha/2}\right) = 1 - \alpha \tag{40}$$

Dabei sind $\chi^2_{n-1,\alpha/2}$ das $\alpha/2$-Quantil und $\chi^2_{n-1,1-\alpha/2}$ das $1 - \alpha/2$-Quantil der χ^2-Verteilung mit $n - 1$ Freiheitsgraden.

Wir formen den Ausdruck in der Klammer in Gleichung (40) so um, dass σ^2 isoliert ist.

$$\chi^2_{n-1,\alpha/2} \leq \frac{(n-1)\,S^2}{\sigma^2} \leq \chi^2_{n-1,1-\alpha/2}$$

$$\iff \frac{1}{\chi^2_{n-1,\alpha/2}} \geq \frac{\sigma^2}{(n-1)\,S^2} \geq \frac{1}{\chi^2_{n-1,1-\alpha/2}}$$

$$\iff \frac{(n-1)\,S^2}{\chi^2_{n-1,1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)\,S^2}{\chi^2_{n-1,\alpha/2}}$$

Das Konfidenzintervall für σ^2 bei Normalverteilung ist also:

$$\left[\frac{(n-1)\,S^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)\,S^2}{\chi^2_{n-1,\alpha/2}}\right] \tag{41}$$

Beispiel 1 (fortgesetzt von Seite 16)
Wir unterstellen Normalverteilung und stellen das Konfidenzintervall für σ^2 zum Konfidenzniveau 0.95 auf.
Es gilt $s^2 = 18331.25$. Der Tabelle der χ^2-Verteilung entnehmen wir $\chi^2_{8;0.025} = 2.18$ und $\chi^2_{8;0.975} = 17.53$. Wir erhalten also folgendes Konfidenzintervall

$$[8363.49, 67278.95]$$

2.2 Prognoseintervalle

Sind die Parameter μ und σ^2 bekannt, so ist das zentrale Schwankungsintervall

$$[\mu - z_{1-\alpha/2}\sigma, \mu + z_{1-\alpha/2}\sigma] \tag{42}$$

ein Prognoseintervall, denn es gilt

$$P(\mu - z_{1-\alpha/2}\sigma \leq X \leq \mu + z_{1-\alpha/2}\sigma) = 1 - \alpha$$

In der Regel sind μ und σ^2 aber unbekannt. Es liegt nahe, diese zu schätzen und die Schätzwerte in die Gleichung (42) einzusetzen. Dies ist aber nur für sehr große Stichprobenumfänge sinnvoll. Für kleine Stichprobenumfänge hingegen ist das exakte Prognoseintervall bei Normalverteilung gegeben durch

$$[\bar{x} - t_{n-1;1-\alpha/2}\, s\,\sqrt{1+1/n}, \bar{x} + t_{n-1;1-\alpha/2}\, s\,\sqrt{1+1/n}] \tag{43}$$

Dabei ist $t_{n-1;1-\alpha/2}$ das $1-\alpha/2$-Quantil der t-Verteilung mit $n-1$ Freiheitsgraden.
Schauen wir uns ein Beispiel an, bevor wir zeigen, wie man dieses Intervall gewinnt.

Beispiel 1 (fortgesetzt von Seite 18)
Der Arbeitnehmer sucht ein Prognoseintervall für die Fahrzeit des nächsten Tages zur sicherheit 0.95. Es gilt $n = 9$, $\bar{x} = 1800$ und $s = 135.4$. Mit $t_{8;0.975} = 2.306$ erhalten wir folgendes Prognoseintervall

$$[1470.9, 2129.1]$$

Um das Intervall herzuleiten, fragen wir uns zunächst, wie wir den Wert von X_{n+1} prognostizieren sollen. Da er möglichst nahe an allen Beobachtungen liegen sollte, bieten sich zwei Kriterien an. Wählt man als Maß für die Nähe, die euklidische Distanz, so erhält man folgendes Kriterium

$$\min \sum_{i=1}^{n} |x_i - x_{n+1}| \tag{44}$$

Die quadrierte euklidische Distanz liefert

$$\min \sum_{i=1}^{n} (x_i - x_{n+1})^2 \tag{45}$$

Im ersten Fall prognostiziert man x_{n+1} durch den Median, im zweiten Fall durch den Mittelwert der Beobachtungen. Da die Verteilung des Mittelwerts angegeben werden kann, verwenden wir diesen. Als Ausgangspunkt der Konstruktion des Prognoseintervalls wählen wir $X_{n+1} - \bar{X}$. Wir gehen im Folgenden davon aus, dass die Zufallsvariablen $X_1, \ldots, X_n, X_{n+1}$ unabhängig und identisch mit den Parametern μ und σ^2 normalverteilt sind. Unter diesen Annahmen gilt

$$E(X_{n+1} - \bar{X}) = E(X_{n+1}) - E(\bar{X}) = \mu - \mu = 0$$

und

$$Var(X_{n+1} - \bar{X}) = Var(X_{n+1}) + Var(\bar{X}) = \sigma^2 + \frac{\sigma^2}{n} = \sigma^2(1 + 1/n)$$

Außerdem ist $X_{n+1} - \bar{X}$ normalverteilt. Also ist

$$\frac{X_{n+1} - \bar{X}}{\sigma\sqrt{1 + 1/n}} \tag{46}$$

standardnormalverteilt. Schätzen wir σ durch S und setzen es in Gleichung (46) ein, so erhalten wir folgende mit $n-1$ Freiheitsgraden t-verteilte Zufallsvariable

$$\frac{X_{n+1}-\bar{X}}{S\sqrt{1+1/n}}$$

Es gilt also

$$P\left[-t_{n-1;1-\alpha/2} \leq \frac{X_{n+1}-\bar{X}}{S\sqrt{1+1/n}} \leq t_{n-1;1-\alpha/2}\right] = 1-\alpha \tag{47}$$

Formen wir diesen Ausdruck so um, dass zwischen den Ungleichungen X_{n+1} steht, so erhalten wir das Prognoseintervall in Gleichung (43):

$$-t_{n-1;1-\alpha/2} \leq \frac{X_{n+1}-\bar{X}}{S\sqrt{1+1/n}} \leq t_{n-1;1-\alpha/2}$$

$$\Longleftrightarrow t_{n-1;1-\alpha/2}\, S\sqrt{1+1/n} \leq X_{n+1}-\bar{X} \leq t_{n-1;1-\alpha/2}\, S\sqrt{1+1/n}$$

$$\Longleftrightarrow \bar{X} - t_{n-1;1-\alpha/2}\, S\sqrt{1+1/n} \leq X_{n+1} \leq \bar{X} + t_{n-1;1-\alpha/2}\, S\sqrt{1+1/n}$$

Da ein Konfidenzintervall ein Intervall für den Wert eines Parameters und ein Prognoseintervall ein Intervall für eine Realisation einer Zufallsvariablen ist, ist die Aussage beim Prognoseintervall mit größerer Unsicherheit behaftet. Dies zeigt sich in der größeren Länge des Prognoseintervalls. Die Länge des Konfidenzintervalls für μ ist nämlich

$$L = 2\, t_{n-1;1-\alpha/2}\, s\sqrt{1/n}$$

Die Länge des Prognoseintervalls beträgt

$$L = 2\, t_{n-1;1-\alpha/2}\, s\sqrt{1+1/n}$$

Da gilt

$$1+\frac{1}{n} > \frac{1}{n}$$

gilt auch

$$\sqrt{1+\frac{1}{n}} > \sqrt{\frac{1}{n}}$$

Die Länge des Konfidenzintervalls konvergiert gegen 0, während die Länge des Prognoseintervalls gegen die Länge des zentralen Schwankungsintervalls konvergiert.

Oft sucht man einseitige Prognoseintervalle. Man will also wissen, welchen Wert die nächste Beobachtung mindestens oder höchsten annimmt.
Bei Normalverteilung gibt es folgende einseitige Prognoseintervalle

$$[\bar{x} - t_{n-1;1-\alpha}\, s\, \sqrt{1+1/n}, \infty) \tag{48}$$

$$(-\infty, \bar{x} + t_{n-1;1-\alpha}\, s\, \sqrt{1+1/n}] \tag{49}$$

Dabei ist $t_{n-1;1-\alpha}$ das $1-\alpha$-Quantil der t-Verteilung mit $n-1$ Freiheitsgraden.

Beispiel 1 (fortgesetzt von Seite 19)
Der Arbeitnehmer will wissen, welchen Wert seine Fahrzeit am nächsten Tag nicht überschreiten wird. Er stellt ein einseitiges Prognoseintervall zur Sicherheit 0.95 auf. Es gilt $n = 9$, $\bar{x} = 1800$ und $s = 135.4$. Mit $t_{8;0.95} = 1.86$ erhalten wir folgendes Prognoseintervall

$$(-\infty, 2065.4]$$

Will er wissen, welchen Wert seine Fahrzeit am nächsten Tag überschreiten wird, so erhält er folgendes einseitige Prognoseintervall zur Sicherheit 0.95:

$$[1534.6, \infty)$$

2.3 Toleranzintervalle

Sind die Parameter μ und σ^2 unbekannt, so ist das zweiseitige Toleranzintervall bei Normalverteilung gegeben durch

$$[\bar{x} - q_{1-\alpha,p,n}\, s, \bar{x} + q_{1-\alpha,p,n}\, s] \tag{50}$$

Dabei gilt

$$q_{1-\alpha,p,n} = r\,\sqrt{\frac{n-1}{\chi^2_{\alpha,n-1}}} \tag{51}$$

mit

$$\Phi\left(\frac{1}{\sqrt{n}} + r\right) - \Phi\left(\frac{1}{\sqrt{n}} - r\right) = p \tag{52}$$

Dabei ist $\chi^2_{\alpha,n-1}$ das α-Quantil der χ^2-Verteilung mit $n-1$ Freiheitsgraden und $\Phi(z)$ der Wert der Verteilungsfunktion der Standardnormalverteilung an der Stelle z.

Die Gleichungen (51) und (52) werden von Kendall et al. (1991) auf den Seiten 774-776 hergeleitet.

Tabellen für $q_{1-\alpha,p,n}$ sind bei Hahn and Meeker (1991), Rinne (1997) und auf den Seiten 26 und 27 zu finden.

Wir können $q_{1-\alpha,p,n}$ aber auch numerisch bestimmen. Hierzu müssen wir zuerst den Wert von r bestimmen, der die Gleichung (52) erfüllt. Lösen wir die Gleichung (52) nach 0 auf, so erhalten wir

$$\Phi\left(\frac{1}{\sqrt{n}}+r\right)-\Phi\left(\frac{1}{\sqrt{n}}-r\right)-p=0$$

Wir suchen also die Nullstelle von

$$f(r)=\Phi\left(\frac{1}{\sqrt{n}}+r\right)-\Phi\left(\frac{1}{\sqrt{n}}-r\right)-p$$

Ein numerisches Verfahren ist das Newton-Raphson-Verfahren. Ist die Nullstelle der Funktion $f(r)$ gesucht, so iteriert man hier beginnend mit dem Startwert r_0

$$r_n=r_{n-1}-\frac{f(r_{n-1})}{f'(r_{n-1})} \tag{53}$$

für $n=1,1,\ldots$ so lange, bis r_n sich stabilisiert.

Als Startwert r_0 kann man den Wert von r wählen, der für $n\to\infty$ die Gleichung (52) erfüllt.

Für gegebenes p wird der Wert r_0 gesucht mit

$$\Phi\left(r_0\right)-\Phi\left(-r_0\right)=p$$

Wegen

$$\Phi\left(-r_0\right)=1-\Phi\left(r_0\right)$$

muss also gelten

$$2\,\Phi\left(r_0\right)-1=p$$

Somit folgt

$$r_0=\Phi^{-1}\left(\frac{p+1}{2}\right)$$

Mit

$$f'(r)=\phi\left(\frac{1}{\sqrt{n}}+r\right)+\phi\left(\frac{1}{\sqrt{n}}-r\right)$$

können wir die Iteration aus Gleichung (53) auf Seite 22 also durchführen. Dabei gilt

$$\phi(z)=\frac{1}{\sqrt{2\,\pi}}\,e^{-0.5\,z^2}$$

Beispiel 1 (fortgesetzt von Seit 21)
Der Arbeitnehmer will wissen, in welchem Intervall mindestens 90 Prozent der Fahrzeiten mit einer Wahrscheinlichkeit von 0.95 liegen.
Es gilt also $p = 0.9$ und $1 - \alpha = 0.95$. Wir bestimmen zunächst r. Es gilt

$$r_0 = \Phi^{-1}\left(\frac{0.9 + 1}{2}\right) = \Phi^{-1}(0.95) = 1.645$$

Somit gilt

$$\begin{aligned} r_1 &= 1.645 - \frac{\Phi\left(\frac{1}{\sqrt{9}} + 1.645\right) - \Phi\left(\frac{1}{\sqrt{9}} - 1.645\right) - 0.9}{\frac{1}{\sqrt{2\,\pi}}\, e^{-0.5\,(1/3+1.645)^2} + \frac{1}{\sqrt{2\,\pi}}\, e^{-0.5\,(1/3-1.645)^2}} \\ &= 1.645 - \frac{\Phi(1.98) - \Phi(1.31) - 0.9}{\frac{1}{\sqrt{2\,\pi}}\, e^{-0.5\,(1/3+1.645)^2} + \frac{1}{\sqrt{2\,\pi}}\, e^{-0.5\,(1/3-1.645)^2}} \\ &= 1.645 - \frac{0.9761 - 0.0948 - 0.9}{0.0564 + 0.1688} = 1.645 - \frac{-0.0187}{0.2252} = 1.728 \end{aligned}$$

Ausgehend von r_1 bestimmen wir r_2 und erhalten $r_2 = 1.734$. Auch r_3 ist gleich 1.734. Wir können also $q_{0.95,0.9,9}$ bestimmen. Mit $\chi^2_{0.05,8} = 2.7326$ gilt

$$q_{0.95,0.9,9} = 1.734 \cdot \sqrt{\frac{8}{2.7326}} = 2.967$$

Dieser Wert stimmt mit den Werten in den Tabellen in Hahn and Meeker (1991) und Rinne (1997) überein. Mit $\bar{x} = 1800$ und $s = 135.4$ erhalten wir folgendes Toleranzintervall

$$[1398.3, 2201.7]$$

Wir können auch einseitige Toleranzintervalle bestimmen:

$$[\bar{x} - w_{1-\alpha,p,n}\, s, \infty) \tag{54}$$

$$(-\infty, \bar{x} + w_{1-\alpha,p,n}\, s] \tag{55}$$

Die Werte von $w^{ue}_{1-\alpha,p,n}$ sind bei Hahn and Meeker (1991) und Rinne (1997) tabelliert.

Literatur

Bickel, P. J. and Doksum, K. A. (2001). *Mathematical statistics.* Prentice Hall, Upper Saddle River, NJ, 2 edition.

Conover, W. J. (1999). *Practical nonparametric statistics.* Wiley, New York, 3 edition.

Gossett, W. S. (1908). The probable error of a mean. *Biometrika*, 6:1–25.

Hahn, G. J. (1970). Statistical intervals for a normal population: part 1, tables, examples, and applications. *Journal of Quality Technology*, 2:115–125.

Hahn, G. J. and Meeker, W. Q. (1991). *Statistical intervals : a guide for practitioners.* Wiley, New York, 1 edition.

Kendall, M. G., Stuart, A., and Ord, J. K. (1991). *The advanced theory of statistics.*, volume 2 Classical inference and relationship. Arnold, London, 5 edition.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the theory of statistics.* McGraw-Hill, New York.

Randles, R. H. and Wolfe, D. A. (1979). *Introduction to the theory of nonparametric statistics.* Wiley, New York, 1 edition.

Rinne, H. (1997). *Taschenbuch der Statistik.* Deutsch, Thun, 2 edition.

A Die simulierten Daten

Tabelle 1: 20 Stichproben vom Umfang $n = 9$ aus der Standardnormalverteilung

Stichprobe									
1	1.05	1.07	-0.12	0.03	-1.59	0.35	1.85	0.40	0.36
2	-0.58	0.12	1.27	0.28	0.68	0.08	0.01	-1.57	1.52
3	0.82	-1.10	-0.79	0.48	0.86	0.22	-0.15	-0.61	0.36
4	0.59	-0.24	-0.64	-0.76	0.51	-0.60	-1.51	-1.58	-0.31
5	-2.03	0.60	1.06	-0.14	-0.29	0.11	0.50	-0.19	-0.71
6	1.86	1.71	0.17	0.87	0.67	-0.17	0.75	1.13	-0.30
7	-0.75	-1.12	2.15	-0.88	0.18	-1.27	-0.49	-0.47	-0.01
8	0.39	-0.17	-0.97	-0.74	0.91	-0.55	-0.61	-0.68	1.45
9	2.24	1.73	-1.15	0.20	1.54	-0.57	-0.90	0.24	-0.30
10	-1.26	-0.13	-0.23	0.89	1.35	-0.18	0.95	-0.73	0.62
11	0.53	1.52	0.84	-0.43	0.38	-0.54	-1.57	-0.85	0.71
12	1.25	-1.42	0.29	0.45	-0.55	0.48	-0.17	1.07	-0.51
13	0.41	0.94	1.56	0.17	1.10	0.76	0.20	1.46	0.51
14	-1.24	0.29	-1.07	1.28	0.13	0.82	-0.98	-0.27	-0.06
15	0.27	-0.25	0.44	-0.56	0.79	-0.25	0.10	1.95	0.82
16	1.03	-0.01	-2.44	1.66	-2.17	1.44	0.04	0.20	1.06
17	-0.76	0.06	-1.25	0.25	0.19	-1.43	0.50	-0.18	-0.04
18	-0.42	-1.71	0.62	-0.91	0.48	1.29	1.99	0.39	1.24
19	2.03	0.17	-0.40	-1.84	1.50	0.16	-1.15	-0.64	0.97
20	0.78	-0.02	-1.25	-1.42	-0.85	0.38	0.74	-0.99	0.44

20 standardnormalverteilte Zufallszahlenn

```
-0.87 -1.52  0.49  0.86 0.89  2.38 -0.69 -0.39 1.09  0.61
 0.14  0.13 -0.65 -0.38 0.87 -0.19 -0.79  2.49 0.14 -0.93
```

B Tabellen

Tabelle 2: Werte von $q_{0.95,p,n}$ für $p = 0.9, 0.95, 0.99$ und $n = 2, 3, \ldots, 20$

p n	0.90	0.95	0.99
2	32.019	37.674	48.430
3	8.380	9.916	12.861
4	5.369	6.370	8.299
5	4.275	5.079	6.634
6	3.712	4.414	5.775
7	3.369	4.007	5.248
8	3.136	3.732	4.891
9	2.967	3.532	4.631
10	2.839	3.379	4.433
11	2.737	3.259	4.277
12	2.655	3.162	4.150
13	2.587	3.081	4.044
14	2.529	3.012	3.955
15	2.480	2.954	3.878
16	2.437	2.903	3.812
17	2.400	2.858	3.754
18	2.366	2.819	3.702
19	2.337	2.784	3.656
20	2.310	2.752	3.615

Tabelle 3: Werte von $q_{0.99,p,n}$ für $p = 0.9, 0.95, 0.99$ und $n = 2, 3, \ldots, 20$

n \ p	0.90	0.95	0.99
2	160.194	188.491	242.301
3	18.930	22.401	29.055
4	9.398	11.150	14.527
5	6.612	7.855	10.260
6	5.337	6.345	8.301
7	4.613	5.488	7.187
8	4.147	4.936	6.468
9	3.822	4.550	5.966
10	3.582	4.265	5.594
11	3.397	4.045	5.308
12	3.250	3.870	5.079
13	3.130	3.727	4.893
14	3.029	3.608	4.737
15	2.945	3.507	4.605
16	2.872	3.421	4.492
17	2.808	3.345	4.393
18	2.753	3.279	4.307
19	2.703	3.221	4.230
20	2.659	3.168	4.161

Tabelle 4: Verteilungsfunktion $\Phi(z)$ der Standardnormalverteilung

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.500	0.504	0.508	0.512	0.516	0.520	0.524	0.528	0.532	0.536
0.1	0.540	0.544	0.548	0.552	0.556	0.560	0.564	0.567	0.571	0.575
0.2	0.579	0.583	0.587	0.591	0.595	0.599	0.603	0.606	0.610	0.614
0.3	0.618	0.622	0.626	0.629	0.633	0.637	0.641	0.644	0.648	0.652
0.4	0.655	0.659	0.663	0.666	0.670	0.674	0.677	0.681	0.684	0.688
0.5	0.691	0.695	0.698	0.702	0.705	0.709	0.712	0.716	0.719	0.722
0.6	0.726	0.729	0.732	0.736	0.739	0.742	0.745	0.749	0.752	0.755
0.7	0.758	0.761	0.764	0.767	0.770	0.773	0.776	0.779	0.782	0.785
0.8	0.788	0.791	0.794	0.797	0.800	0.802	0.805	0.808	0.811	0.813
0.9	0.816	0.819	0.821	0.824	0.826	0.829	0.831	0.834	0.836	0.839
1.0	0.841	0.844	0.846	0.848	0.851	0.853	0.855	0.858	0.860	0.862
1.1	0.864	0.867	0.869	0.871	0.873	0.875	0.877	0.879	0.881	0.883
1.2	0.885	0.887	0.889	0.891	0.893	0.894	0.896	0.898	0.900	0.901
1.3	0.903	0.905	0.907	0.908	0.910	0.911	0.913	0.915	0.916	0.918
1.4	0.919	0.921	0.922	0.924	0.925	0.926	0.928	0.929	0.931	0.932
1.5	0.933	0.934	0.936	0.937	0.938	0.939	0.941	0.942	0.943	0.944
1.6	0.945	0.946	0.947	0.948	0.949	0.951	0.952	0.953	0.954	0.954
1.7	0.955	0.956	0.957	0.958	0.959	0.960	0.961	0.962	0.962	0.963
1.8	0.964	0.965	0.966	0.966	0.967	0.968	0.969	0.969	0.970	0.971
1.9	0.971	0.972	0.973	0.973	0.974	0.974	0.975	0.976	0.976	0.977
2.0	0.977	0.978	0.978	0.979	0.979	0.980	0.980	0.981	0.981	0.982
2.1	0.982	0.983	0.983	0.983	0.984	0.984	0.985	0.985	0.985	0.986
2.2	0.986	0.986	0.987	0.987	0.987	0.988	0.988	0.988	0.989	0.989
2.3	0.989	0.990	0.990	0.990	0.990	0.991	0.991	0.991	0.991	0.992
2.4	0.992	0.992	0.992	0.992	0.993	0.993	0.993	0.993	0.993	0.994
2.5	0.994	0.994	0.994	0.994	0.994	0.995	0.995	0.995	0.995	0.995
2.6	0.995	0.995	0.996	0.996	0.996	0.996	0.996	0.996	0.996	0.996
2.7	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997
2.8	0.997	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998
2.9	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.999	0.999	0.999
3.0	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999

Tabelle 5: Quantil z_p der Standardnormalverteilung

p	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
0.50	0.000	0.002	0.005	0.008	0.010	0.012	0.015	0.018	0.020	0.023
0.51	0.025	0.028	0.030	0.033	0.035	0.038	0.040	0.043	0.045	0.048
0.52	0.050	0.053	0.055	0.058	0.060	0.063	0.065	0.068	0.070	0.073
0.53	0.075	0.078	0.080	0.083	0.085	0.088	0.090	0.093	0.095	0.098
0.54	0.100	0.103	0.106	0.108	0.110	0.113	0.116	0.118	0.121	0.123
0.55	0.126	0.128	0.131	0.133	0.136	0.138	0.141	0.143	0.146	0.148
0.56	0.151	0.154	0.156	0.159	0.161	0.164	0.166	0.169	0.171	0.174
0.57	0.176	0.179	0.182	0.184	0.187	0.189	0.192	0.194	0.197	0.199
0.58	0.202	0.204	0.207	0.210	0.212	0.215	0.217	0.220	0.222	0.225
0.59	0.228	0.230	0.233	0.235	0.238	0.240	0.243	0.246	0.248	0.251
0.60	0.253	0.256	0.258	0.261	0.264	0.266	0.269	0.272	0.274	0.277
0.61	0.279	0.282	0.284	0.287	0.290	0.292	0.295	0.298	0.300	0.303
0.62	0.306	0.308	0.311	0.313	0.316	0.319	0.321	0.324	0.327	0.329
0.63	0.332	0.334	0.337	0.340	0.342	0.345	0.348	0.350	0.353	0.356
0.64	0.358	0.361	0.364	0.366	0.369	0.372	0.374	0.377	0.380	0.383
0.65	0.385	0.388	0.391	0.393	0.396	0.399	0.402	0.404	0.407	0.410
0.66	0.412	0.415	0.418	0.421	0.423	0.426	0.429	0.432	0.434	0.437
0.67	0.440	0.443	0.445	0.448	0.451	0.454	0.456	0.459	0.462	0.465
0.68	0.468	0.470	0.473	0.476	0.479	0.482	0.484	0.487	0.490	0.493
0.69	0.496	0.499	0.501	0.504	0.507	0.510	0.513	0.516	0.519	0.522
0.70	0.524	0.527	0.530	0.533	0.536	0.539	0.542	0.545	0.548	0.550
0.71	0.553	0.556	0.559	0.562	0.565	0.568	0.571	0.574	0.577	0.580
0.72 ´	0.583	0.586	0.589	0.592	0.595	0.598	0.601	0.604	0.607	0.610
0.73	0.613	0.616	0.619	0.622	0.625	0.628	0.631	0.634	0.637	0.640
0.74	0.643	0.646	0.650	0.653	0.656	0.659	0.662	0.665	0.668	0.671

Tabelle 6: Quantil z_p der Standardnormalverteilung

p	.000	.001	.002	.003	.004	.005	.006	.007	.008	.009
0.75	0.674	0.678	0.681	0.684	0.687	0.690	0.694	0.697	0.700	0.703
0.76	0.706	0.710	0.713	0.716	0.719	0.722	0.726	0.729	0.732	0.736
0.77	0.739	0.742	0.745	0.749	0.752	0.755	0.759	0.762	0.766	0.769
0.78	0.772	0.776	0.779	0.782	0.786	0.789	0.793	0.796	0.800	0.803
0.79	0.806	0.810	0.813	0.817	0.820	0.824	0.827	0.831	0.834	0.838
0.80	0.842	0.845	0.849	0.852	0.856	0.860	0.863	0.867	0.870	0.874
0.81	0.878	0.882	0.885	0.889	0.893	0.896	0.900	0.904	0.908	0.912
0.82	0.915	0.919	0.923	0.927	0.931	0.935	0.938	0.942	0.946	0.950
0.83	0.954	0.958	0.962	0.966	0.970	0.974	0.978	0.982	0.986	0.990
0.84	0.994	0.999	1.003	1.007	1.011	1.015	1.019	1.024	1.028	1.032
0.85	1.036	1.041	1.045	1.049	1.054	1.058	1.062	1.067	1.071	1.076
0.86	1.080	1.085	1.089	1.094	1.098	1.103	1.108	1.112	1.117	1.122
0.87	1.126	1.131	1.136	1.141	1.146	1.150	1.155	1.160	1.165	1.170
0.88	1.175	1.180	1.185	1.190	1.195	1.200	1.206	1.211	1.216	1.221
0.89	1.226	1.232	1.237	1.243	1.248	1.254	1.259	1.265	1.270	1.276
0.90	1.282	1.287	1.293	1.299	1.305	1.311	1.316	1.322	1.328	1.335
0.91	1.341	1.347	1.353	1.360	1.366	1.372	1.379	1.385	1.392	1.398
0.92	1.405	1.412	1.419	1.426	1.432	1.440	1.447	1.454	1.461	1.468
0.93	1.476	1.483	1.491	1.498	1.506	1.514	1.522	1.530	1.538	1.546
0.94	1.555	1.563	1.572	1.580	1.589	1.598	1.607	1.616	1.626	1.635
0.95	1.645	1.655	1.665	1.675	1.685	1.695	1.706	1.717	1.728	1.739
0.96	1.751	1.762	1.774	1.787	1.799	1.812	1.825	1.838	1.852	1.866
0.97	1.881	1.896	1.911	1.927	1.943	1.960	1.977	1.995	2.014	2.034
0.98	2.054	2.075	2.097	2.120	2.144	2.170	2.197	2.226	2.257	2.290
0.99	2.326	2.366	2.409	2.457	2.512	2.576	2.652	2.748	2.878	3.090

Tabelle 7: Quantile der χ^2-Verteilung mit k Freiheitsgraden

k	$\chi^2_{k;0.025}$	$\chi^2_{k;0.05}$	$\chi^2_{k;0.1}$	$\chi^2_{k;0.9}$	$\chi^2_{k;0.95}$	$\chi^2_{k;0.975}$
1	0.001	0.004	0.016	2.706	3.841	5.024
2	0.051	0.103	0.211	4.605	5.991	7.378
3	0.216	0.352	0.584	6.251	7.815	9.348
4	0.484	0.711	1.064	7.779	9.488	11.143
5	0.831	1.145	1.610	9.236	11.070	12.833
6	1.237	1.635	2.204	10.645	12.592	14.449
7	1.690	2.167	2.833	12.017	14.067	16.013
8	2.180	2.733	3.490	13.362	15.507	17.535
9	2.700	3.325	4.168	14.684	16.919	19.023
10	3.247	3.940	4.865	15.987	18.307	20.483
11	3.816	4.575	5.578	17.275	19.675	21.920
12	4.404	5.226	6.304	18.549	21.026	23.337
13	5.009	5.892	7.042	19.812	22.362	24.736
14	5.629	6.571	7.790	21.064	23.685	26.119
15	6.262	7.261	8.547	22.307	24.996	27.488
16	6.908	7.962	9.312	23.542	26.296	28.845
17	7.564	8.672	10.085	24.769	27.587	30.191
18	8.231	9.390	10.865	25.989	28.869	31.526
19	8.907	10.117	11.651	27.204	30.144	32.852
20	9.591	10.851	12.443	28.412	31.410	34.170
21	10.283	11.591	13.240	29.615	32.671	35.479
22	10.982	12.338	14.041	30.813	33.924	36.781
23	11.689	13.091	14.848	32.007	35.172	38.076
24	12.401	13.848	15.659	33.196	36.415	39.364
25	13.120	14.611	16.473	34.382	37.652	40.646