Übungen mit dem Applet „Vertrauensbereich für den Mittelwert“

1 Begriffe und statistischer Hintergrund

Stichprobenergebnisse x_i unterliegen der Zufallsstreuung, daher streuen auch Kennwerte, die aus den Stichprobenergebnissen berechnet werden.

Der **Mittelwert** $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$ der Stichprobenergebnisse ist ein Schätzwert für den wahren, aber normalerweise unbekannten Mittelwert μ der Verteilung (n = Stichprobenumfang = Anzahl der Werte in einer Messreihe).

Zieht man mehrere Stichproben aus derselben Grundgesamtheit (d.h. wiederholt man eine Messreihe unter identischen Bedingungen), so erhält man aufgrund der Zufallsstreuung unterschiedliche Stichprobenmittelwerte $\bar{x}$, die jedoch immer in der Nähe des Mittelwerts μ der Verteilung liegen. Daraus kann man umgekehrt folgern, dass der wahre Mittelwert μ in der Nähe des Stichprobenmittelwerts $\bar{x}$ liegt – man kann den Vertrauensbereich (=Konfidenzintervall) für den Mittelwert μ angeben. Der 95%-Vertrauensbereich für μ ist

$$\bar{x} - t_{f;0,975} \cdot \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{f;0,975} \cdot \frac{s}{\sqrt{n}}$$

Die **Standardabweichung** $s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$ ist ein Schätzwert für die Standardabweichung $\sigma = \sqrt{\sigma^2}$ (d.h. die Breite) der Verteilung. Der t-Wert ist ein „Sicherheitszuschlagsfaktor“, der vom gewünschten Vertrauensniveau (hier 95%) und vom Freiheitsgrad f = n – 1 abhängt (vgl. Applet „Stetige Verteilungen“).

Aus jeder Stichprobe erhält man einen anderen Vertrauensbereich, aber 95% der mit dieser Formel berechneten Vertrauensbereiche enthalten den wahren Mittelwert μ.

Mit dem Applet können Sie μ und σ>0 der Grundgesamtheit beliebig vorgeben, die Verteilung wird als Kurve dargestellt. Aus dieser Grundgesamtheit wird eine Stichprobe mit Umfang n gezogen, die Einzelwerte x_i und der Stichprobenmittelwert $\bar{x}$ werden mit dem 95%- und dem 99%-Vertrauensbereich dargestellt. Die grafischen Darstellungen sollen zeigen, dass

- die Stichprobenergebnisse streuen (trotz aller Sorgfalt bei der Versuchsdurchführung)

- aus jeder Stichprobe erhält man einen anderen Vertrauensbereich, aber 95% der 95%-Vertrauensbereiche enthalten den wahren Mittelwert μ (das Maximum der Kurve)
- die 99%-Vertrauensbereiche sind immer breiter als die 95%-Vertrauensbereiche
- je größer der Stichprobenumfang, desto schmäler sind die Vertrauensbereiche im Allgemeinen, die Breite streut jedoch, weil s streut.

2 Visualisierungen mit dem Applet

2.1 Zufallsstreuung der Einzelwerte und der Mittelwerte

Geben Sie mit R einen beliebigen Mittelwert μ (z.B. 10) und eine Standardabweichung $\sigma>0$ (z.B. 2) vor. Wählen Sie zunächst einen kleinen Wert (z.B. 4) für den Stichprobenumfang n. Starten Sie für jede Einstellung die Simulation mehrmals mit G und beobachten Sie

- für jede Simulation erhält man andere Einzelwerte (=Zufallsstreuung)
- manchmal unterscheiden sich die Werte mehr, manchmal weniger
- auch der Mittelwert streut, er liegt aber immer in der Nähe des Maximums der Verteilung.

Wählen Sie nun einen wesentlich größeren Wert (z.B. 100) für den Stichprobenumfang n. Starten Sie die Simulation wieder mehrmals und beobachten Sie

- die Einzelwerte streuen über einen breiteren Bereich
- die Breite der Bereiche ändert sich jedoch weniger von Simulation zu Simulation
- auch der Mittelwert streut weniger und liegt meist näher am Maximum der Verteilung.

2.2 Vertrauensbereich für den Mittelwert

In dieser Simulation haben Sie Mittelwert μ und Standardabweichung σ vorgegeben, sie sind Ihnen daher bekannt. Wären die Einzelergebnisse jedoch Versuchsergebnisse gewesen, die Sie bei Wiederholungen eines Versuchs unter identischen Bedingungen gewonnen hätten, so wüssten Sie μ und σ nicht. Sie hätten nur die Versuchsergebnisse und könnten daraus den Stichprobenmittelwert $\bar{x}$ und die Stichprobenstandardabweichung s berechnen. Aufgrund Ihrer Erfahrungen in 2.1 erwarten Sie jedoch, dass der wahre Mittelwert μ in der Nähe des Stichprobenmittelwerts $\bar{x}$ liegt – Sie können den Vertrauensbereich berechnen (mit ein/aus).

Starten Sie mit verschiedenen Vorgaben für μ, σ und n die Simulation jeweils mehrmals und beobachten Sie in der Grafik:

- Je sicherer Sie sein wollen, dass der wahre Mittelwert μ im Bereich liegt, desto breiter wird der Bereich – der 99%-Vertrauensbereich ist immer breiter als der 95%-Vertrauensbereich.
- Bei jeder Simulation erhalten Sie einen anderen Vertrauensbereich, allein aufgrund der Zufallsstreuung – auch wenn sich in den Versuchsbedingungen nichts geändert hat, werden Sie bei der Wiederholung einer Messreihe nicht dieselben Ergebnisse erhalten.
- Manchmal ist die Standardabweichung s größer, manchmal kleiner, und damit ändert sich auch die Breite der Vertrauensbereiche – dies bedeutet nicht notwendigerweise, dass eine Messreihe schlampiger durchgeführt wurde als die andere.
- Die meisten Vertrauensbereiche enthalten den wahren Mittelwert, aber manchmal enthält der Vertrauensbereich den wahren Mittelwert nicht (er ist dann rot gekennzeichnet, dies geschieht beim 95%-Vertrauensbereich natürlich häufiger) – bei einer einzelnen Messreihe haben Sie natürlich nur einen Vertrauensbereich und Sie wissen nicht, ob gerade dieser Vertrauensbereich den wahren Mittelwert enthält oder nicht.
- Je größer der Stichprobenumfang n ist, desto schmäler sind die Vertrauensbereiche (bei großem n nimmt die Breite wie $1/\sqrt{n}$ ab) – für mehr Versuchsaufwand werden Sie mit schmäleren Bereichen belohnt.
- **Achtung: Der Vertrauensbereich macht eine Aussage über den Mittelwert der Verteilung (d.h. die Lage des Maximums), nicht über Einzelwerte.** Mit zunehmendem n wird der Vertrauensbereich immer schmäler, der Bereich, in dem Einzelwerte liegen wird dagegen breiter – bei einer großen Stichprobe sind mit höherer Wahrscheinlichkeit auch Extremwerte mit dabei.
- Wenn es Ihnen gelingt, durch mehr Sorgfalt oder bessere Methoden die Standardabweichung zu halbieren, wird die Breite des Vertrauensbereichs ebenfalls halbiert. Wollten Sie diese Halbierung durch größeren Stichprobenumfang n erreichen, bräuchten Sie viermal so viele Einzelversuche – Sorgfalt wird daher

vorausgesetzt, größerer Stichprobenumfang ist nur empfehlenswert, wenn Sie trotz Sorgfalt die Streuung nicht weiter reduzieren können.

2.3 Interpretation des Vertrauensniveaus

Um die Bedeutung des Vertrauensniveaus zu erleben, schalten Sie die Darstellung mit um (Toggle ein/aus). Sie sehen nun Mittelwerte und Vertrauensbereiche für 50 Stichproben gleichzeitig. Für alle Stichproben wurden dieselben Werte für n, μ und σ verwendet, die Ergebnisse unterscheiden sich also nur zufällig.
Starten Sie die Simulation mehrmals und für verschiedene Werte von n (z.B. 4 und 100) und beobachten Sie:

- Aus jeder Stichprobe erhält man andere Vertrauensbereiche.
- Die meisten Vertrauensbereiche überdecken den wahren Mittelwert (senkrechte Linie), aber 5% der 95%-Vertrauensbereiche und 1% der 99%-Vertrauensbereiche überdecken den wahren Mittelwert nicht – sie sind dann rot gezeichnet.
- 5% heißt im Mittel 2,5 der 50 Bereiche, 1% heißt im Mittel 0,5 der 50 Bereiche – die genaue Anzahl schwankt jedoch von Simulation zu Simulation und genügt der Binomialverteilung mit n=50 und p=0,05 bzw. 0,01.
- Bei größerem Stichprobenumfang sind die Vertrauensbereiche schmäler, der Anteil rot gekennzeichneter Bereiche ist jedoch unverändert – das Risiko wird über das Vertrauensniveau gesteuert und ist unabhängig vom Stichprobenumfang.

3 Verallgemeinerung

In dieser Simulation wurde der Vertrauensbereich für den Mittelwert dargestellt. Auch für andere Stichprobenparameter (z.B. Median, Standardabweichung, Steigung einer Regressionsgeraden, Anteil fehlerhafter Einheiten) kann man Vertrauensbereiche berechnen. Die Interpretation des Vertrauensniveaus (Konfidenzniveaus) ist immer dieselbe und auch die Abhängigkeit vom Stichprobenumfang ist ähnlich (bei ausreichend großem n nehmen die Breiten sämtlicher Vertrauensbereiche wie $1/\sqrt{n}$ ab.