

Kapitel 8

Normalverteilung

8.1 Gauß-Glocke und Gauß'sche Φ -Funktion

Die Zufallsvariable $x = f(u_1, \dots, u_k, v_1, v_2, \dots)$ habe als Wertebereich die Menge aller reellen Zahlen \mathbb{R} oder aber die Menge aller reellen Zahlen, die im Bereich $a \leq x \leq b$ liegen. Sorgt man bei der Durchführung eines Experiments dafür, dass die Einflussgrößen v_1, v_2, \dots , die nicht auf vorgeschriebene konstante Werte gesetzt werden, **die ganze Fülle ihrer** im Rahmen der gewählten Messmethode **zulässigen Werte annehmen**, d.h. einen repräsentativen Querschnitt ihres jeweils zugelassenen Wertebereichs, so ergibt sich sehr oft folgender Effekt:

Der Graph der Dichtefunktion des Experiments hat glockenförmige Gestalt.

Der Einfluss der nicht konstant gehaltenen Größen ist also standardisiert worden. Er ist hierdurch gleichsam neutralisiert, nicht weiter interpretierbar, als erklärender Kausalfaktor für die Messergebnisse weitestgehend ausgeschaltet. Und dieser Effekt stellt sich unabhängig davon ein, auf welchen Werten das Experiment die kontrollierten Einflussgrößen u_1, \dots, u_k konstant hält, er ist also spezifisch für die Messmethode und das heißt für die Zufallsvariable x selbst.¹

Bezeichnung: *Ergibt sich für eine Zufallsvariable x als Dichtefunktion zur idealen Messreihe eine Funktion $y = p(x)$ mit symmetrischem, glockenförmigem Graphen, so nennt man die Zufallsvariable **normalverteilt** oder **zufallsverteilt**.*

Der Erwartungswert \hat{x} wird in diesem Fall allgemein mit μ bezeichnet.

Regel 65 (Berechnungsformel der Dichtefunktion bei Normalverteilung):

R 65

Ist eine Zufallsvariable x normalverteilt, so hängt die Dichtefunktion $y = p(x)$ zur idealen Messreihe/zum Experiment nur vom Erwartungswert μ des Experiments und der Streuung σ der Messmethode ab, und zwar nach der Formel

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-0,5 \cdot \left(\frac{x - \mu}{\sigma}\right)^2\right)$$

*Diese Dichtefunktion heißt **die zu μ und σ gehörige Glockenfunktion**.*

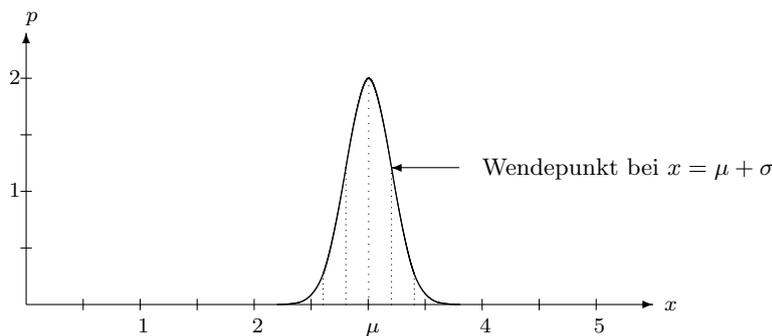
¹Zum untrennbaren Zusammenhang zwischen Messmethode und Zufallsvariable siehe 7.1, S.147

Ihr Graph gleicht der Silhouette einer auf der x -Achse aufsitzenden symmetrischen Glocke. Die senkrechte Symmetrieachse verläuft bei $x = \mu$ durch den höchsten Punkt der Glocke.

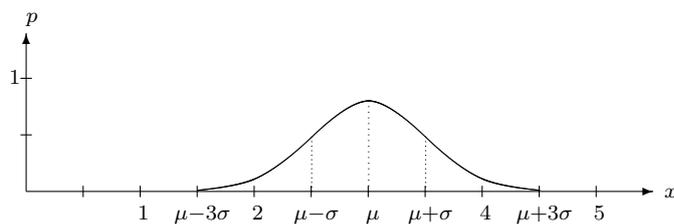
Bei $x = \mu - \sigma$ und $x = \mu + \sigma$ hat der Graph Wendepunkte von konvex zu konkav bzw. umgekehrt.

Da die Exponentialfunktion nur positive Werte annimmt, gilt strenggenommen $p(x) > 0$ für alle $x \in \mathbb{R}$, aber für $x < \mu - 3\sigma$ und für $x > \mu + 3\sigma$ ist $p(x)$ schon recht nahe Null. Deshalb kann man, vergrößert gesprochen, sagen, dass die Breite der Glocke an der Basis etwas mehr als 6σ beträgt.

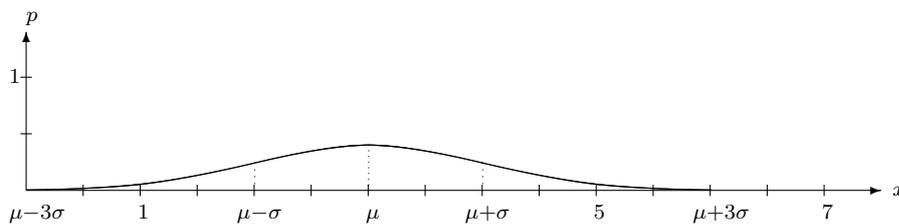
Da die Gesamtfläche zwischen der Glockenkurve und der x -Achse stets $=1$ ist,² gilt: Je kleiner die Streuung σ , umso schmaler und folglich umso höher die Glocke, je größer die Streuung σ , umso breiter und flacher die Glocke:



Glockenfunktion zu $\mu = 3$ und $\sigma = 0.2$



Glockenfunktion zu $\mu = 3$ und $\sigma = 0.5$



Glockenfunktion zu $\mu = 3$ und $\sigma = 1$

Da die Gesamtfläche unter dem Graphen $=1$ ist, ist die Fläche unter dem Graphen links der Symmetrieachse $= \frac{1}{2}$. Somit kennen wir von der zugehörigen Verteilungsfunktion $y = F(x)$ schon den Anfangswert $F(\mu) = \frac{1}{2}$, denn $F(\mu) = \int_{-\infty}^{\mu} p(t)dt =$ die Fläche unter der Glocke linksseits von μ .

²siehe 7.3, S.131

Bezeichnung: Die zu einer Glockenfunktion gehörige Stammfunktion $y = F(x)$ mit dem Anfangswert $F(\mu) = \frac{1}{2}$ heißt Verteilungsfunktion zur Normalverteilung mit dem Erwartungswert μ und der Streuung σ , kurz:

Verteilungsfunktion zur $N(\mu; \sigma)$ -Verteilung.

Entsprechend heißt die Glockenfunktion auch die Dichtefunktion zur $N(\mu; \sigma)$ -Verteilung.

Für die Verteilungsfunktion zu einer Glockenfunktion gibt es außer der Formel

$$F(x) = \int_{-\infty}^x p(t)dt \quad \text{für alle } x \in \mathbb{R}$$

keine andere Berechnungsformel. Da aber der Anfangswert $F(\mu) = \frac{1}{2}$ bekannt ist, kann für sie durch numerische Integration³ eine beliebig genaue Wertetabelle berechnet werden.

Die einfachste Berechnungsformel für die Glockenfunktion ergibt sich im Spezialfall $\mu = 0$ und $\sigma = 1$:

Bezeichnung: Die Dichtefunktion zur $N(0; 1)$ -Verteilung ist

$$z = \varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right) \tag{8.1}$$

Sie wird traditionell als die **Gauß-Glocke**, auch **Gaußsche Fehlerfunktion** oder **Gaußsche Dichtefunktion** bezeichnet.

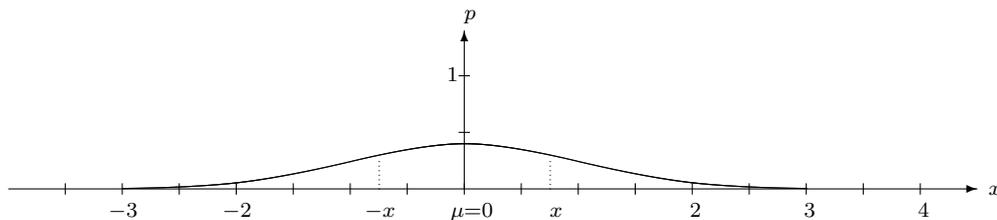
Die zugehörige Verteilungsfunktion zur $N(0; 1)$ -Verteilung

$$y = \Phi(x) = \int_{-\infty}^x \varphi(t)dt = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{t^2}{2}\right)dt \tag{8.2}$$

heißt **Gaußsche-Normalverteilung** oder **Gaußsche Φ -Funktion**.

Sie ist sehr genau tabelliert worden (sog. **Φ -Tabelle**⁴).

Bei der Φ -Tabelle macht man sich zunutze, dass die Gauß-Glocke wegen $\mu = 0$ symmetrisch zur senkrechten Achse liegt:



Die Gauß-Glocke ($\mu = 0, \sigma = 1$)

Die Fläche unter der Gauß-Glocke im Bereich zwischen $-\infty$ und $-x$ ist genauso groß wie die zwischen x und ∞ . Also gilt

$$\begin{aligned} \int_{-\infty}^{-x} \varphi(t)dt &= \int_x^{\infty} \varphi(t)dt \\ \Phi(-x) - \Phi(-\infty) &= \Phi(\infty) - \Phi(x) \\ \Phi(-x) - 0 &= 1 - \Phi(x) \end{aligned}$$

³siehe 6.5.1, S.141

⁴siehe Tabellen zur Statistik, S.218

Symmetrieregeln der Φ -Funktion:

$$\Phi(-x) = 1 - \Phi(x) \quad \text{für alle } x \in \mathbb{R}. \quad (8.3)$$

Wegen dieser Symmetrieregeln sind in der Φ -Tabelle nur die Funktionswerte für positive x ausgedrückt.

Der eigentliche Nutzen der Φ -Tabelle besteht darin, dass sie auch zur Berechnung von Werten aller anderen Verteilungsfunktionen im Falle beliebiger μ und σ benutzt werden kann, denn es besteht folgender

Zusammenhang zwischen der Normalverteilung $y = F(x)$ zur $N(\mu; \sigma)$ -Verteilung und der Gaußschen Normalverteilung $y = \Phi(x)$:

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Beweis : Die Normalverteilung $y = F(x)$ zur $N(\mu; \sigma)$ -Verteilung ist nach *Regel 65*⁵ eine Stammfunktion zur Glockenfunktion

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-0,5 \cdot \left(\frac{x - \mu}{\sigma}\right)^2\right) \quad (8.4)$$

und hat für $x = \mu$ den Wert $F(\mu) = \frac{1}{2}$. Wenn wir zeigen, dass auch die Funktion $u = \Phi\left(\frac{x - \mu}{\sigma}\right)$ eine Stammfunktion zu dieser Glockenfunktion ist und an der Stelle $x = \mu$ den Wert $\frac{1}{2}$ hat, dann folgt nach den *Regeln für Stammfunktionen* (S.3)⁶, dass beide Funktionen gleich sind.

$$\begin{aligned} u' &= \left(\Phi\left(\frac{x - \mu}{\sigma}\right)\right)' && | \text{ Kettenregel} \\ &= \Phi'\left(\frac{x - \mu}{\sigma}\right) \cdot \left(\frac{1}{\sigma} \cdot (x - \mu)\right)' && | \Phi'(x) = \varphi(x) \\ &= \varphi\left(\frac{x - \mu}{\sigma}\right) \cdot \frac{1}{\sigma} \cdot 1 && | \text{ für } \varphi \text{ einsetzen} \\ &= \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-0,5 \cdot \left(\frac{x - \mu}{\sigma}\right)^2\right) \\ &\stackrel{(8.4)}{=} p(x), \end{aligned}$$

also ist $u = \Phi\left(\frac{x - \mu}{\sigma}\right)$ eine Stammfunktion zur vorgeschriebenen Glockenfunktion.

Der Erwartungswert der $N(0; 1)$ -Verteilung ist 0, $y = \Phi(x)$ ist die zugehörige Verteilungsfunktion, daher gilt $\Phi(0) = \frac{1}{2}$. Daraus folgt

$$u(\mu) = \Phi\left(\frac{\mu - \mu}{\sigma}\right) = \Phi(0) = \frac{1}{2}.$$

Damit ist alles gezeigt. □

Daraus gewinnt man mit *Regel 64*⁷ als vielgebrauchte praktische Anwendung

⁵siehe 8.1, S.156

⁶siehe 6.2, S.135

⁷siehe 7.3, S.154

Regel 66 (Berechnung von Wahrscheinlichkeiten bei Normalverteilung):**R 66**

Ist x eine normalverteilte Zufallsvariable mit Erwartungswert μ und Streuung σ , so gilt für die Wahrscheinlichkeit, dass eine Messung von x einen Wert im Bereich $a \leq x \leq b$ liefert,

$$P(a \leq x \leq b) = \int_a^b p(x)dx = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \quad (8.5)$$

Da $\Phi(-\infty) = 0$ und $\Phi(\infty) = 1$ ist, folgen als Spezialfälle die Formeln

$$P(x \leq b) = \int_{-\infty}^b p(x)dx = \Phi\left(\frac{b-\mu}{\sigma}\right) \quad (8.6)$$

$$P(a \leq x) = \int_a^{\infty} p(x)dx = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right) \quad (8.7)$$

WARNUNG: Ist eine Zufallsvariable x normalverteilt, so ist die Wahrscheinlichkeit dafür, dass eine Messung **exakt** einen bestimmten Wert a annimmt, immer gleich Null ($\hat{=}$ Fall $a = b$).

Um ein sinnvolles Ergebnis zu bekommen, muss man die Messungenauigkeit $|\Delta x|$ (=den absoluten Fehler bei der empirischen Messung) miteinbeziehen und die Wahrscheinlichkeit

$$P(a - |\Delta x| \leq x \leq a + |\Delta x|)$$

berechnen.

Beispiel: Man kann also nicht fragen, wie wahrscheinlich es ist, dass eine 170cm große Zwanzigjährige 60kg wiegt, sondern nur, wie wahrscheinlich es ist, dass sie $60\text{kg} \pm 0,5\text{kg}$ oder $60\text{kg} \pm 100\text{g}$... wiegt.

8.2 Sicherheitsvereinbarungen, Fehler 1. und 2. Art

In den folgenden Paragraphen dieses Kapitels geht es immer wieder um Beurteilungen, die jeweils (nur) mit 95%-iger, 99%-iger oder 99,9%-iger Sicherheit gemacht werden können. Dabei ist folgende Sprechweise gebräuchlich:

Bezeichnung: Ergebnisse, die mit 95%-iger Sicherheit ausgesprochen werden, heißen **wahrscheinlich**, solche, die auf 99%-iger Sicherheit basieren, heißen **signifikant**, solche mit 99,9%-iger Sicherheit **hochsignifikant**.

In verschiedenen Anwendungsbereichen sind, wenn nicht ausdrücklich etwas anderes vereinbart wurde, jeweils die folgenden charakteristischen Sicherheitsanforderungen üblich:

Regel 67 (Sicherheitsbestimmungen):**R 67**

Bei Produktbeurteilungen/Qualitätskontrollen werden Urteile mit 95% Sicherheit getroffen, bei wissenschaftlichen Untersuchungen trifft man nur Urteile mit 99% Sicherheit, bei patentrechtlichen und sehr kapitalintensiven Fragestellungen (z.B. bei hohen Investitionen) werden Entscheidungen nur auf Basis 99,9%-iger Sicherheit gefällt.

Diese Praxis birgt zwangsläufig ein unvermeidliches **Restrisiko**:

Bezeichnung: *Stützt man sich auf ein Urteil, weil es mit der vereinbarten Sicherheit getroffen wurde, und es entspricht trotzdem nicht der Realität, so spricht man von einem Fehler*

1. Art. *(Ein Fehler 1. Art ist also ein übereifriges Urteil.)*

Wird ein Urteil verworfen, weil es nicht mit der geforderten Sicherheit ausgesprochen werden kann, und es entspricht trotzdem der Realität, so spricht man von einem Fehler **2. Art.** *(Ein Fehler 2. Art ist also übergroße Reserve beim Beurteilen und Stellungnehmen.)*

Konsequenz: Je niedriger die vereinbarte Sicherheit ist (= bei Produktbeurteilungen), umso eher passieren Fehler 1. Art (also vorschnelle Urteile), aber umso seltener passieren Fehler 2. Art (also keine Stellungnahme). Je höher die vereinbarte Sicherheit ist (= bei wissenschaftlichen Untersuchungen oder sehr kapitalintensiven Fragestellungen), umso seltener passieren Fehler 1. Art (vorschnelle Urteile), aber umso eher passieren Fehler 2. Art (Verweigerung angebrachter Stellungnahme).

Die unterschiedlichen Sicherheitsvereinbarungen der verschiedenen Anwendungsbereiche erklären sich aus dem jeweiligen Interesse des Urteilenden:

Bei Qualitätskontrollen besteht das fragliche Urteil in einer Mängelfeststellung. Für den Hersteller/Lieferanten einer Ware ist es geschäftlich gesehen aber unvergleichlich viel günstiger, eine Ware schon bei wahrscheinlichen Mängeln rasch vorsorglich aus dem Verkehr zu nehmen und/oder zu ersetzen, als das Risiko einzugehen, dass ihm von anderer Seite in einem Prozess nachgewiesen wird, wahrscheinlich mangelhafte Produkte vorsätzlich weiter zu liefern. Sowohl in Hinsicht auf den guten Ruf der Firma wie auf die Kostenseite ist jede noch so umfangreiche Rückrufaktion der Mühe wert. Man nimmt dabei ein Risiko von 5% in Kauf, dass die eingezogene Ware doch einwandfrei ist.

Bei wissenschaftlichen Aussagen und sehr kapitalintensiven Entscheidungen hingegen besteht allerhöchster Anspruch, dass das Urteil auch den Tatsachen entspricht (= signifikant ist). Deshalb gilt es hier als oberstes Gebot, lieber gar keine Aussage zu machen als eine, die tatsächlich unzutreffend ist. Man nimmt nur ein Risiko von 1% bzw. 1 Promille in Kauf, dass die Aussage unzutreffend ist.

Folgerung: Hersteller, die ihre Warentests nicht mit 95%-iger Sicherheit durchführen (sondern mit höherer) und Wissenschaftler, die ihre Aussagen nicht mit 99%-iger Sicherheit machen (sondern mit niedrigerer), gelten als wenig seriös.

8.3 Entfernung fehlerhafter Messdaten: Der Ausreißertest

Die entscheidenden Schlüssel zur Kenntnis einer normalverteilten Zufallsvariablen x sind der Erwartungswert μ und die Streuung σ .⁸ Liegen nur kürzere Messreihen vor, so benutzt man \bar{x} und s behelfsweise als Näherungswerte für μ und σ . In späteren Abschnitten wird auch gezeigt, wie man die absoluten Fehler $|\bar{x} - \mu|$ und $|s - \sigma|$ abschätzen kann.⁹ Dazu müssen aber zunächst einmal sowohl \bar{x} als auch s korrekt aus den Messdaten x_1, \dots, x_n berechnet

⁸Das liegt nicht nur in der Eigenbedeutung beider Größen (siehe 7.2, S.148 f), begründet, sondern auch darin, dass ihre Kenntnis genügt, um die Dichtefunktion und die Verteilungsfunktion zu berechnen (siehe 7.3, Regeln 65, 66).

⁹siehe die Abschnitte über den Vertrauensbereich für \bar{x} (8.5, S.150) bzw. für s (8.7, S.153)

worden sein. Es tritt nun das Problem auf, dass einzelne unglaubwürdige Messdaten, die durch fehlerhafte Messung oder fehlerhafte Notierung des Messergebnisses zustande gekommen sein können, sowohl bei der Berechnung von \bar{x} wie bei der von s gravierende Verzerrungen des Ergebnisses verursachen können, und zwar umso mehr, je kleiner die Anzahl n der Messwerte ist. Es ist also erforderlich, solche Fehldaten aus der Messreihe zu entfernen.

Genauere wissenschaftliche Untersuchungen darüber, wie die empirische Dichtefunktion einer Messreihe sich mit wachsendem n der charakteristischen Glockenfunktion $y = p(x)$ des Experiments annähert, haben zu einem Instrumentarium geführt, das es gestattet, auch und gerade bei kürzeren Messreihen solche Daten herauszufinden, die allzu weit abseits liegen, um mit der Normalverteilung von x verträglich zu sein.

Der folgende Test setzt also voraus, dass die Zufallsvariable x normalverteilt ist, und entscheidet unter dieser Prämisse, ob Daten innerhalb einer Messreihe unglaubwürdig sind (sog. **Ausreißer**).

Der Ausreißertest (nach Nalimov):

Ziel: Entfernung unglaubwürdiger Daten aus einer Messreihe bei normalverteilter Zufallsvariabler x .

Man benötigt: Eine empirische Messreihe mit n , \bar{x} und s ($n > 3$) und die **r-Tabelle**.¹⁰

Durchführung:

1. Schritt: Ermittle einen Messwert mit maximalem Abstand zu \bar{x} (in Frage kommen der kleinste oder der größte Messwert der Messreihe) und nenne diesen Messwert x^* (d.h. der Abstand $|x^* - \bar{x}|$ ist maximal).

2. Schritt: Berechne

$$r^* = \frac{|x^* - \bar{x}|}{s} \cdot \sqrt{\frac{n}{n-1}} \text{ und } f = n - 2.$$

3. Schritt: Schlage für dieses f den Eintrag in der r-Tabelle nach und vergleiche r^* mit dem Tabellenwert $\text{tab-}r$.

Auswertung:

- Ist $r^* < \text{tab-}r(95\%)$, so ist durch den Test ein Ausreißer "nicht feststellbar".
 - Ist $r^* \geq \text{tab-}r(95\%)$, so ist x^* **wahrscheinlich** ein Ausreißer und wird aus der Messreihe entfernt.
 - Ist $r^* \geq \text{tab-}r(99\%)$, so ist x^* **signifikant** ein Ausreißer und wird aus der Messreihe entfernt.
 - Ist $r^* \geq \text{tab-}r(99,9\%)$, so ist x^* **hochsignifikant** ein Ausreißer und wird aus der Messreihe entfernt.
4. Schritt: Nach Entfernung eines Ausreißers müssen für die verkürzte Messreihe n , \bar{x} und s neu berechnet werden und der Ausreißertest an der verkürzten Messreihe wiederholt werden. Das Verfahren bricht ab, sobald der Test zum ersten Mal keinen Ausreißer feststellen konnte.

¹⁰siehe Tabellen zur Statistik, S.230

Der Test ist sehr scharf, besonders bei größerem n .

Bezeichnung: Nach der Entfernung aller Ausreißer heißt eine Messreihe **statistisch homogen** oder **ausreißerfrei**.

WARNUNG: Ausreißer sind “unglaubliche” Daten unter der Hypothese der Normalverteilung. Trotzdem müssen Messdaten, die mit diesem Test aus der Messreihe entfernt werden, nicht unbedingt fehlerhaft sein, sondern können eventuell hochwichtige Einzelphänomene anzeigen, die nicht “normal” (d.h. durch bloßen Zufall) erklärlich sind, sondern auf besondere Ursachen hindeuten, denen nachzugehen wäre. Über die **Weiterbehandlung der Ausreißer nach ihrer Entfernung aus der Messreihe** kann nur der Fachmann des Arbeitsgebietes entscheiden.

8.4 Der Streubereich und seine Schätzung

Bezeichnung: Sei x eine normalverteilte Zufallsvariable. Ein Bereich $a \leq x \leq b$, in dem ein Messwert im Rahmen eines gegebenen Experiments mit (mindestens) 95%iger Wahrscheinlichkeit zu liegen kommt, heißt ein **95%-Streubereich der Variablen x** .

In einem solchen Bereich liegen bei langen Messreihen also mindestens 95% aller Messergebnisse.

Analog spricht man vom **99%-Streubereich** (ist größer) und vom **99,9%-Streubereich** (ist noch größer).

Der 95%-Streubereich einer Variablen x hat keine eindeutig bestimmten Eckwerte a und b : Wenn man a noch verkleinert und b noch vergrößert, erhält man erst recht einen 95%-Streubereich von x . Auch leichte seitliche Verschiebungen sind im Allgemeinen erlaubt. Vorrangig interessiert aber, den Streubereich möglichst eng einzugrenzen, also ein Intervall möglichst kleiner Länge zu kennen, welches ein 95%-Streubereich von x ist.

Gibt es zu einer normalverteilten Zufallsvariablen Tausende von Messdaten (also eine ideale Messreihe), so lassen sich kleinstmögliche Streubereiche leicht ermitteln. Dann ist nämlich der Zahlwert von μ und σ konkret bekannt (weil in sehr guter Näherung $\approx \bar{x}$ bzw. $\approx s$) und man kann folgende Rechnung anstellen: Sei $k = 1, 2$ oder 3 und $a = \mu - k\sigma$, $b = \mu + k\sigma$, dann gilt

$$\begin{aligned}
 P(\mu - k\sigma \leq x \leq \mu + k\sigma) &= P(a \leq x \leq b) && | \text{ Regel 66} \\
 &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) && | \text{ für } a \text{ und } b \text{ einsetzen} \\
 &= \Phi\left(\frac{\mu+k\sigma-\mu}{\sigma}\right) - \Phi\left(\frac{\mu-k\sigma-\mu}{\sigma}\right) \\
 &= \Phi(k) - \Phi(-k) && | \text{ Symmetrieregeln für } \Phi \text{ (8.3)} \\
 &= \Phi(k) - (1 - \Phi(k)) \\
 &= 2\Phi(k) - 1 && k = 1, 2, 3
 \end{aligned}$$

Nachschlagen in der Φ -Tabelle für $k = 1, 2, 3$ liefert hieraus folgende Wahrscheinlichkeiten:

Regel 68 (Grobe Abschätzung von Streubereichen mittels der Streuung σ):

R 68

$$P(\mu - \sigma \leq x \leq \mu + \sigma) = 0,6826 = 68,26\%,$$

$$P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = 0,9544 = 95,44\%,$$

$$P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = 0,9974 = 99,74\%,$$

d.h. bei normalverteilten Zufallsvariablen liegen mehr als zwei Drittel aller Messergebnisse im Bereich $\mu - \sigma \leq x \leq \mu + \sigma$, gut 95% im Bereich $\mu - 2\sigma \leq x \leq \mu + 2\sigma$ und 99,7% im Bereich $\mu - 3\sigma \leq x \leq \mu + 3\sigma$.

Das erklärt, warum man die Streuung σ auch den “mittleren Fehler“ der Messmethode nennt: Erst nach Kenntnis des Streubereichs kann der Informationswert des Erwartungswertes μ eines Experiments richtig beurteilt werden.¹¹

Regel 67 macht schärfstmögliche Angaben, basierend auf genauer Kenntnis von μ und σ . Sie liefert Streubereiche der Bauart

$$\mu - t \cdot \sigma \leq x \leq \mu + t \cdot \sigma$$

mit $t = 2$ (im Fall 95,44%) bzw. $t = 3$ (im Fall 99,74%).

Es gibt aber zwei Einwände: Einerseits sind die Prozentangaben nicht exakt gleich 95%, 99% und 99,9%, andererseits ist man auch stark daran interessiert, Streubereiche ebenfalls von solchen normalverteilten Zufallsvariablen zu kennen, zu denen nur kleinere Messreihen existieren und damit \bar{x} und s nur grobe Schätzwerte für μ und σ liefern, und zwar umso größere, je kleiner die Anzahl n der Messungen.

Eine berühmte wissenschaftlichen Studie (die ein junger Wissenschaftler unter dem Pseudonym “student“ veröffentlichte) lieferte eine Tabelle, die angibt, wie man in Abhängigkeit von n den Faktor t möglichst vorsichtig zu verändern hat, wenn man für μ und σ die Schätzwerte \bar{x} und s einsetzt. Dabei wird der Faktor einerseits so korrigiert, dass die drei gefragten Prozentbereiche (95%, 99% bzw. 99,9%) genau erfasst werden, zum andern wird der Faktor t (und damit auch der Streubereich) jeweils so vergrößert, dass die mit immer kleinerem n sich verschlimmernde Ungenauigkeit der Schätzwerte \bar{x} und s gerade noch abgefangen wird. Die Tabelle ist als **t-Tabelle**¹² bekannt, der tabellierte Faktor t heisst auch **student-Faktor**.

Regel 69 (Schärfstmögliche Angabe der 95%-, 99%- und 99,9%-Streubereiche):

R 69

Sind von einer statistisch homogenen Messreihe n , \bar{x} und s berechnet worden, so ist die hiernach bestmögliche Schätzung des Streubereichs gegeben durch $\bar{x} - ts \leq x \leq \bar{x} + ts$, kurz:

$$\bar{x} \pm ts,$$

wobei der student-Faktor t aus der t-Tabelle zu entnehmen ist, und zwar in Abhängigkeit von dem sog. Freiheitsgrad $f = n - 1$ und der gewünschten Sicherheit in %.

Liegt die ideale Messreihe vor, so benutzt man entsprechend μ , σ und $f = \infty$.

In der **Medizin** benutzt man Streubereiche zur raschen Einschätzung des Seltenheitswertes einer Einzelmessung am Patienten.

¹¹siehe 7.2, S.148. Vergleiche auch das Konzept des “mittleren Fehlers“ bei der Linearen Regression, 4.7, S.59, Formel (4.5)

¹²siehe Tabellen zur Statistik, S.230

Beispiel: Kinderärzte tragen bei den Vorsorgeuntersuchungen für Kleinkinder die gemessenen Werte für Kopfumfang, Körperlänge und Gewicht jeweils in vorgedruckte Graphiken ein, aus denen die 95%- und 99%-Streubereiche dieser Zufallsvariablen in Abhängigkeit vom Lebensalter ablesbar sind. (Die Daten basieren auf idealen Messreihen für die Bevölkerung, könnten aber nach Jahrzehnten veraltet sein.)

Da die Vorsorgeuntersuchung der Kleinkinder in den wissenschaftlichen Bereich fällt, wird man ein Messergebnis (nur) dann als medizinisch auffällig, als außerhalb des normalen Bereichs liegend einstufen, wenn es nicht im 99%-Streubereich liegt.

8.5 Der Vertrauensbereich für \bar{x} und seine Schätzung

Regel 70 (Schätzung des Abstandes zwischen \bar{x} und μ):

R 70

Ist eine statistisch homogene Messreihe mit n , \bar{x} und s gegeben, so gilt mit 95% bzw. 99% bzw. 99,9% Sicherheit folgende Abschätzung für den Abstand zwischen \bar{x} und μ :

$$|\bar{x} - \mu| \leq \frac{t \cdot s}{\sqrt{n}},$$

wobei der student-Faktor t aus der t -Tabelle zu entnehmen ist, abhängig vom Freiheitsgrad $f = n - 1$ und der gewünschten Sicherheit in %.

Anwendungen:

[1] Ist der Erwartungswert μ bekannt, so kann man den Abstand $|\bar{x} - \mu|$ direkt ausrechnen und mit der Aussage der Regel vergleichen:

Wenn er tatsächlich $\leq \frac{t \cdot s}{\sqrt{n}}$ ist, ist mit der jeweiligen Sicherheit die Erhebung der Messdaten korrekt erfolgt und **der Zahlwert von \bar{x} vertrauenswürdig**, ist der Abstand hingegen größer, so hat man mit der jeweiligen Sicherheit **einen systematischen Messfehler aufgedeckt**.

[2] Ist der Erwartungswert μ unbekannt, so kann man hiermit **den Wert von μ schätzen**: Mit der jeweiligen Sicherheit gilt

$$\bar{x} - \frac{t \cdot s}{\sqrt{n}} \leq \mu \leq \bar{x} + \frac{t \cdot s}{\sqrt{n}}.$$

Bezeichnung: *Der Bereich*

$$\bar{x} - \frac{t \cdot s}{\sqrt{n}} \leq x \leq \bar{x} + \frac{t \cdot s}{\sqrt{n}}.$$

heißt, je nach der gewählten Sicherheit, der **95%-**, **99%-** bzw. **99,9%-Vertrauensbereich für \bar{x}** .

8.5.1 Anwendung: Qualitätstest für quantitative Eigenschaften

Eine dritte Anwendungsvariante von *Regel 70* ergibt sich wie folgt:

Wird für eine Ware zugesichert, daß eine Variable x einen gewissen Sollwert X annimmt (z.B. sei x das reale Nettogewicht einer abgepackten Portion und $X = 500g$ das Sollgewicht), und wird diese Zusicherung auch praktisch eingehalten, so liefern immer weiter verlängerte Messreihen zu x Mittelwerte \bar{x} , die für $n \rightarrow \infty$ gegen X streben, d.h. dann (und nur dann) erweist sich X als identisch mit μ und nach *Regel 70* gilt

$$|\bar{x} - X| \leq \frac{t \cdot s}{\sqrt{n}} \quad \text{oder auch} \quad \frac{|\bar{x} - X| \cdot \sqrt{n}}{s} \leq t.$$

Damit erhält man einen vielbenutzten

Qualitätstest an Waren auf Einhaltung eines Sollwertes X einer Variablen x

Ziel: Überprüfung, ob eine gewisse quantitative Eigenschaft der Ware eingehalten wird.

Man benötigt: Eine **statistisch homogene** empirische Messreihe mit n , \bar{x} und s , den Sollwert X und die **t-Tabelle**.¹³

Durchführung:

1. Schritt: Berechne

$$TAU = \frac{|\bar{x} - X| \cdot \sqrt{n}}{s} \quad \text{und} \quad f = n - 1.$$

2. Schritt: Schlage für dieses f und 95% Sicherheit den Eintrag in der t-Tabelle nach und vergleiche TAU mit dem Tabellenwert t .

Auswertung:

- Ist $TAU < t$, so ist durch den Test eine Abweichung vom Sollwert X "nicht feststellbar".
- Ist $TAU \geq t$, so gilt die Ware als mangelhaft, da der Sollwert X nicht eingehalten wird.

8.6 Der F-Test zum Vergleich zweier Streuungen s_1, s_2

In 7.1 wurde dargelegt, dass zwei Messreihen nur dann wirklich von derselben Zufallsvariablen x handeln, wenn sie dieselbe Messmethode benutzen. Und nur dann können ihre inhaltlichen Ergebnisse (= Mittelwerte) sinnvollerweise auf Gleichheit oder Verschiedenheit untersucht werden. Es ist daher entscheidend wichtig, einen Test auf Gleichheit der Messmethode zu besitzen.

Da die Streuung σ eine spezifische Konstante der Messmethode ist und die empirische Streuung s ein Näherungswert für σ , können zwei Messreihen, denen dieselbe Messmethode zu

¹³siehe Tabellen zur Statistik, S.230

Grunde liegt, sich in ihren empirischen Streuungen s_1, s_2 nicht beliebig stark unterscheiden, und zwar umso weniger, je größer die Anzahl der Messungen n_1, n_2 ist.

Sollen mehr als zwei Streuungen gleichzeitig auf Gleichheit getestet werden, so muss statt des F-Tests ein anderer Test verwendet werden, z.B. der Bartlett-Test¹⁴.

Der F-Test:

Ziel: Überprüfung der Streuungen zweier Messreihen auf Gleichheit.

Man benötigt: Zwei **statistisch homogene** Messreihen mit n_1, s_1 und n_2, s_2 und die **F-Tabelle**.¹⁵ (Die Mittelwerte werden nicht gebraucht.)

Durchführung:

1. Schritt: Richte die Nummerierung der Messreihen so ein, dass $s_1 \geq s_2$ ist (notfalls Nummerierung bei s und(!) n vertauschen).

2. Schritt: Berechne

$$F = \left(\frac{s_1}{s_2}\right)^2, \quad f_1 = n_1 - 1 \text{ und } f_2 = n_2 - 1.$$

(Achtung! Wenn der 1. Schritt richtig durchgeführt wurde, ist $F \geq 1$.)

3. Schritt: Schlage für diese f_1 und f_2 den Eintrag in der F-Tabelle nach und vergleiche F mit dem Tabellenwert $\text{tab-}F$.

Auswertung:

- Ist $F < \text{tab-}F(95\%)$, so ist durch den Test ein Unterschied zwischen den Streuungen s_1, s_2 "nicht feststellbar".
- Ist $F \geq \text{tab-}F(95\%)$, so sind s_1 und s_2 **wahrscheinlich** verschieden.
- Ist $F \geq \text{tab-}F(99\%)$, so sind s_1 und s_2 **signifikant** verschieden.
- Ist $F \geq \text{tab-}F(99,9\%)$, so sind s_1 und s_2 **hochsignifikant** verschieden.

Anwendungen:

- [1] Stimmt es, dass **zwei empirische** Messreihen **dasselbe Experiment** durchgeführt zu haben? Nur wenn ein Unterschied zwischen s_1 und s_2 nicht feststellbar ist, darf man sich der nächstfolgenden Frage zuwenden, ob die Ergebnisse die gleichen sind.
- [2] Stimmt es, dass **zwei empirische** Messreihen unterschiedliche Experimente zu **derselben Zufallsvariablen** x durchgeführt zu haben? Nur wenn ein Unterschied zwischen s_1 und s_2 nicht feststellbar ist, darf man sich der Frage zuwenden, ob die Ergebnisse sich unterscheiden.
- [3] Stimmt es, dass **eine empirische** Messreihe mit n_1 und s_1 eine wohlbekannte Methode benutzt (und also eine wohlbekannte Zufallsvariable x gemessen) hat? Wenn ja, muss der Test an ihr und der idealen Messreihe mit $n_2 = \infty, s_2 = \sigma$ ($\Rightarrow f_2 = \infty$) ergeben, dass ein Unterschied zwischen s_1 und σ nicht feststellbar ist. (Im 1. Schritt ggf. unnummerieren!)

8.6.1 Anwendung: Test auf gleichmäßige Qualität einer Ware

Eine praktische Anwendung des F-Tests ergibt sich wie folgt: Besteht eine quantitative Eigenschaft x einer Ware darin, dass sie einen bestimmten Sollwert X erfüllt¹⁶ (bei Medikamenten z.B. ist für die

¹⁴siehe 8.9.1, S.172

¹⁵siehe Tabellen zur Statistik, S.220

¹⁶siehe 8.5.1, S.166

Menge x eines enthaltenen Wirkstoffs stets eine Sollmenge X vorgeschrieben), so ist es oft dennoch unvermeidlich, dass bei Kontrollmessungen von x das Messergebnis in einem gewissen Maße um den Sollwert X herum schwankt. Man kann jedoch verlangen, dass diese Schwankung ein produktionsbedingtes Mindestmaß nicht überschreitet. Dieses zulässige Mindestmaß nennt man die **Sollstreuung** S . Wünschenswert und erstrebt ist also stets eine möglichst kleine Sollstreuung.

Ist für ein Produkt eine Sollstreuung S zugesichert und wird diese Zusicherung auch praktisch eingehalten, so liefern immer weiter verlängerte Messreihen zu x empirische Streuungen s , die für $n \rightarrow \infty$ gegen S streben, d.h. dann (und nur dann) erweist sich S als identisch mit σ .

Damit erhält man einen vielbenutzten

Qualitätstest an Waren auf Einhaltung einer Sollstreuung

Ziel: Gewährleistung gleichbleibender Qualität einer Ware.

Man benötigt: Eine **statistisch homogene** empirische Messreihe zu x mit n und s , die Sollstreuung S und die **F-Tabelle**.¹⁷

Durchführung:

1. Schritt: Ist $s \geq S$, so nummeriere $n_1 = n$, $s_1 = s$ sowie $n_2 = \infty$, $s_2 = S$, sonst umgekehrt.
2. Schritt: Berechne

$$F = \left(\frac{s_1}{s_2} \right)^2, \quad f_1 = n_1 - 1 \text{ und } f_2 = n_2 - 1.$$

(Achtung! Wenn der 1. Schritt richtig durchgeführt wurde, ist $F \geq 1$. Einer der f -Werte ist $= \infty$.)

3. Schritt: Schlage für diese f_1 und f_2 und 95% Sicherheit den Eintrag in der F-Tabelle nach und vergleiche F mit dem Tabellenwert $\text{tab-}F(95\%)$.

Auswertung:

- Ist $F < \text{tab-}F(95\%)$, so ist durch den Test eine Abweichung von der Sollstreuung S "nicht feststellbar".
- Ist $F \geq \text{tab-}F(95\%)$, so gilt die Ware als mangelhaft, da die Sollstreuung S nicht eingehalten wurde.

Bemerkung 1: Ist $s \leq S$, so wird man im Allgemeinen hoch zufrieden sein, die Ware nicht als mangelhaft sondern als über Erwarten gleichmäßig gelungen betrachten und den Test schon im 1. Schritt abbrechen. Anders steht es, wenn man probeweise zu einer neuen Produktionsmethode übergegangen ist und an deren Produkten nun testen will, ob mit der neuen Methode echt geringere Streuungen als die alte Sollstreuung S erzielt werden können. Dann wird man den obigen Qualitätstest weiter durchführen, allerdings eventuell mit 99% Sicherheit, wenn die Produktionsumstellung im großen Stil sehr kapitalintensiv wäre, sogar mit 99,9% Sicherheit.

Bemerkung 2: Dem Abnehmer der Ware wird die Sollstreuung von x in der Regel nicht ausdrücklich deklariert, sondern die zugesicherte Qualitätsgarantie versteckt sich in der Genauigkeit, mit der der Sollwert X bekanntgemacht wird.¹⁸ Wird im Beipackzettel eines Medikaments angegeben, dass Wirkstoff A in der Konzentration $X = 0,5\text{g}/100\text{ml}$ enthalten ist, so wird damit indirekt eine Sollstreuung $S = 0,05\text{g}/100\text{ml}$ zugesichert. Lautet die Inhaltsangabe hingegen $X = 500\text{mg}/100\text{ml}$, so beträgt die zugesicherte Sollstreuung nur noch $S = 0,5\text{mg}/100\text{ml}$.

¹⁷siehe Tabellen zur Statistik, S.220

¹⁸siehe die *Zahlenverlässlichkeitsregel* in 4.5.1, S.52

8.7 Der Vertrauensbereich für s und seine Schätzung

Der F-Test basiert auf genauen wissenschaftlichen Untersuchungen darüber, wie nah, in Abhängigkeit von n , die Streuung s einer Messreihe ihrem Grenzwert σ , der Streuung der Messmethode, kommt. Dies kann man dazu verwenden, um bei gegebener empirischer Messreihe und unbekanntem σ unter Ausnutzung des F-Tests eine Schätzung für σ zu erhalten:

Regel 71 (Schätzung des Wertes von σ und des Abstandes zwischen s und σ):

R 71

Gegeben sei eine **statistisch homogene empirische Messreihe** mit n und s , eine vereinbarte Sicherheit (wahlweise 95%, 99% oder 99,9%) und die F-Tabelle.

Schlage unter der vereinbarten Sicherheit für $f_1 = n - 1$, $f_2 = \infty$ den Eintrag in der F-Tabelle nach und nenne ihn F_u .

Schlage unter der vereinbarten Sicherheit für $f_1 = \infty$, $f_2 = n - 1$ den Eintrag in der F-Tabelle nach und nenne ihn F_o .

Dann liegt mit der vereinbarten Sicherheit σ zwischen folgenden Eckwerten

$$\frac{s}{\sqrt{F_u}} < \sigma < s\sqrt{F_o}$$

und für den Abstand $|s - \sigma|$ gilt die Abschätzung

$$|s - \sigma| \leq s\sqrt{F_o} - \frac{s}{\sqrt{F_u}}$$

Bezeichnung: Der Bereich

$$\frac{s}{\sqrt{F_u}} < x < s\sqrt{F_o}$$

heißt, je nach der gewählten Sicherheit, der **95%-**, **99%-** bzw. **99,9%-Vertrauensbereich für s** .

Der Name dieses Intervalls erklärt sich wie folgt: Ist die Methodenstreuung σ bekannt, so kann man den Abstand $|s - \sigma|$ direkt ausrechnen und mit der Aussage der Regel vergleichen: Wenn er tatsächlich der Abschätzung der Regel entspricht, ist mit der jeweiligen Sicherheit die Erhebung der Messdaten methodisch korrekt erfolgt und der Zahlwert von \bar{x} entsprechend brauchbar, ist der Abstand hingegen größer, so liegt mit der jeweiligen Sicherheit ein Methodenfehler vor. (Bequemer testet man dies aber gemäss Anwendung [3] des F-Tests, S.167). Ist der Erwartungswert σ unbekannt, so kann man darauf vertrauen, dass s den Wert von σ in der angegebenen Weise eingrenzt.

Beweis von Regel 71: Wir wählen die gewünschte Sicherheit und wenden den F-Test an auf die gegebene empirische Messreihe und die ideale Messreihe mit $n = \infty$ und einem σ , das zwar zahlenmäßig unbekannt ist, aber tatsächlich als Grenzwert zu s gehört.

1. Fall: $s \geq \sigma$:

Laut 1. Schritt bekommt die empirische Messreihe die Nr.1, die ideale die Nr. 2.

Laut 2. Schritt berechnen wir $F = \left(\frac{s}{\sigma}\right)^2$, $f_1 = n - 1$ und $f_2 = \infty$.

Wir schlagen im 3. Schritt den tab-F-Wert nach und nennen ihn F_u .

Da σ und s zusammengehören, ist im 4. Schritt ein Unterschied zwischen beiden nicht feststellbar, d.h. es gilt $F < \text{tab-F}$, eingesetzt:

$$\begin{aligned} \left(\frac{s}{\sigma}\right)^2 &< F_u \\ \frac{s}{\sigma} &< \sqrt{F_u} \\ \frac{s}{\sqrt{F_u}} &< \sigma \end{aligned} \tag{8.8}$$

Nach Fallannahme ist $\sigma \leq s$. Weil alle Einträge der F-Tabelle > 1 sind, auch F_o , ist weiterhin $s < s\sqrt{F_o}$, zusammengenommen daher $\sigma \leq s < s\sqrt{F_o}$. In Kombination mit (8.8) ergibt sich mit der gewählten Sicherheit

$$\frac{s}{\sqrt{F_u}} < \sigma \leq s < s\sqrt{F_o}. \quad (8.9)$$

2. Fall: $s < \sigma$:

Laut 1. Schritt bekommt die ideale Messreihe die Nr.1, die empirische die Nr. 2.

Laut 2. Schritt berechnen wir $F = \left(\frac{\sigma}{s}\right)^2$, $f_1 = \infty$ und $f_2 = n - 1$.

Wir schlagen im 3. Schritt den tab- F -Wert nach und nennen ihn F_o .

Da σ und s zusammengehören, ist im 4. Schritt ein Unterschied zwischen beiden nicht feststellbar, d.h. es gilt $F < \text{tab-}F$, eingesetzt:

$$\begin{aligned} \left(\frac{\sigma}{s}\right)^2 &< F_o \\ \frac{\sigma}{s} &< \sqrt{F_o} \\ \sigma &< s\sqrt{F_o} \end{aligned} \quad (8.10)$$

Weil auch $F_u > 1$ ist, ist $s < s\sqrt{F_u}$, also $\frac{s}{\sqrt{F_u}} < s$, außerdem nach Fallannahme $s < \sigma$, zusammengenommen daher $\frac{s}{\sqrt{F_u}} < s < \sigma$. In Kombination mit (8.10) ergibt sich mit der gewählten Sicherheit

$$\frac{s}{\sqrt{F_u}} < s < \sigma < s\sqrt{F_o}. \quad (8.11)$$

Damit ist die in *Regel 71* behauptete Lage von σ in dem angegebenen Vertrauensbereich in jedem Fall bewiesen.

Da nach (8.9) und (8.11) mit σ stets auch s in diesem Vertrauensbereich liegt, kann der Abstand $|s - \sigma|$ höchstens so groß sein wie der zwischen den Eckdaten dieses Bereichs. Damit ist auch die Abschätzung für den Abstand bewiesen. \square

8.8 Der t-Test zum Vergleich zweier Mittelwerte \bar{x}_1, \bar{x}_2

Die Ergebnisse (= Mittelwerte) zweier empirischer Messreihen darf man bekanntlich nur miteinander in Beziehung bringen/vergleichen, wenn sie **dieselbe** Zufallsvariable x gemessen haben, d.h. wenn sie dieselbe Messmethode benutzt haben, was wiederum unbedingt dadurch abgesichert werden muss, dass der F-Test (Tabelle für 95%) keinen Unterschied zwischen den empirischen Streuungen feststellt.

Der t-Test:

Ziel: Prüfung, ob die Mittelwerte \bar{x}_1, \bar{x}_2 **zweier empirischer** Messreihen sich unterscheiden.

Man benötigt: Zwei **statistisch homogene** empirische Messreihen mit n_1, \bar{x}_1, s_1 und n_2, \bar{x}_2 und s_2 (also n_1, n_2 beide $\neq \infty$), **die den F-Test bestanden haben**, d.h. dass ein Unterschied zwischen s_1 und s_2 nicht feststellbar war, sowie die **t-Tabelle**.

Durchführung:

1. Schritt: Berechne

$$s_d = \sqrt{\frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}}.$$

2. Schritt: Berechne

$$T = \frac{|\bar{x}_1 - \bar{x}_2|}{s_d} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \quad \text{sowie} \quad f = n_1 + n_2 - 2.$$

3. Schritt: Schlage für dieses f den Eintrag in der t-Tabelle nach und vergleiche T mit dem Tabellenwert t .

Auswertung:

- Ist $T < t(95\%)$, so ist durch den Test ein Unterschied zwischen \bar{x}_1 und \bar{x}_2 "nicht feststellbar".
- Ist $T \geq t(95\%)$, so sind \bar{x}_1 und \bar{x}_2 **wahrscheinlich** verschieden.
- Ist $T \geq t(99\%)$, so sind \bar{x}_1 und \bar{x}_2 **signifikant** verschieden.
- Ist $T \geq t(99,9\%)$, so sind \bar{x}_1 und \bar{x}_2 **hochsignifikant** verschieden.

Sollen mehr als zwei Mittelwerte simultan auf Gleichheit getestet werden, so muss statt des t-Tests ein anderer Test verwendet werden, z.B. die Einfache Varianzanalyse¹⁹.

8.8.1 Anwendung 1: Zusammenfassung zweier Messreihen zu einer

Wenn zwei **statistisch homogene** empirische Messreihen zur Messung derselben Zufallsvariablen x mit n_1, \bar{x}_1, s_1 und n_2, \bar{x}_2 und s_2 (also n_1, n_2 beide $\neq \infty$), **sowohl den F-Test als auch den t-Test bestanden haben**, d.h. dass ein Unterschied weder zwischen s_1 und s_2 noch zwischen \bar{x}_1 und \bar{x}_2 feststellbar war, so darf man sie zu einer einzigen, längeren Messreihe zusammenfassen. Der Mittelwert \bar{x} und die Streuung s der zusammengefassten Reihe ergeben verbesserte Näherungswerte für μ und σ .

Die Berechnung der statistischen Daten n, \bar{x} und s der zusammengefassten Messreihe geschieht wie folgt:

1. Schritt: Berechne $n = n_1 + n_2$.

2. Schritt: Berechne

$$\bar{x} = \frac{n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2}{n}.$$

3. Schritt: Berechne

$$s = \sqrt{\frac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n - 1} + \frac{n_1 \cdot n_2 \cdot (\bar{x}_1 - \bar{x}_2)^2}{n \cdot (n - 1)}}.$$

Wiederholte Anwendung gestattet es, viele kleine Datensammlungen zu einer nahezu idealen Messreihe zusammenzufügen und somit μ und σ möglichst genau zu berechnen. Dieses ermöglicht dann einerseits die Berechnung von Wahrscheinlichkeiten mit der Φ -Tabelle²⁰, andererseits eine scharfe Berechnung von Streubereichen²¹.

Wissenschaftliche Arbeiten auf dem Gebiet der Medizin etwa bemühen sich, Datensammlungen, die aus den verschiedensten Kliniken und Publikationen stammen, zu immer größeren Datenpools zusammenzufassen.

¹⁹siehe 8.9.2, S.173

²⁰siehe 8.1, *Regel 66*, S.160

²¹siehe 8.4, *Regel 69*, S.164, und Anwendungsbeispiel dazu

8.8.2 Anwendung 2: Untersuchung von Einflüssen auf die normalverteilte Zufallsvariable x

Will man die Abhängigkeitsbeziehungen der Zufallsvariablen

$$x = f(u_1, \dots, u_k, v_1, v_2, \dots)$$

von den u_i genauer untersuchen, so kann man verschiedene Ansätze verfolgen:

Entweder man betrachtet x künstlich als Funktion einer einzigen Variablen (z.B. u_1) und konzipiert verschiedene Experimente, in denen alle u_2, \dots, u_k auf immer dieselben konstanten Werte gesetzt werden sowie alle v_1, v_2, \dots immer innerhalb derselben Bandbreiten variieren, nur u_1 setzt man in den verschiedenen Experimenten auf verschiedene Werte.

Oder man betrachtet x als Funktion von zwei Variablen u_1, u_2 (oder noch mehr) und plant Experimente, in denen die zu untersuchenden Einflussgrößen u_1, u_2 (oder mehr) auf unterschiedlichen interessierenden Wertekombinationen konstant gehalten werden, während für alle anderen u_i und für die v_i in allen Experimenten dieselben Bedingungen gelten.

Vergleicht man nun zwei Messreihen zu zwei verschiedenen derartigen Experimenten mittels des t-Tests, so wird erkennbar, ob die Experimente zu unterschiedlichen Ergebnissen führen.

8.9 Simultaner Vergleich von mehr als zwei empirischen Messreihen

Sollen die Streuungen und Mittelwerte von mehr als zwei Messreihen auf Gleichheit miteinander verglichen werden, so besteht das genaueste Verfahren darin, alle Teilmengen von je zwei dieser Messreihen mittels des F-Tests und anschließenden t-Tests zu untersuchen. Das ergibt aber rasch einen extrem hohen Arbeitsaufwand: Ist m die Anzahl der Messreihen, so sind $\binom{m}{2} = \frac{m \cdot (m-1)}{2}$ verschiedene Paare von je zwei Messreihen zu überprüfen.

Will man diesen Arbeitsaufwand vermeiden und ist bereit, dafür eine geringere Ergebnisschärfe in Kauf zu nehmen, so kann man alle m Messreihen auf einmal vergleichen mit den zwei nachfolgenden Tests.

8.9.1 Der Bartlett-Test zum Vergleich von Streuungen s_1, \dots, s_m (Verallgemeinerter F-Test)

Ziel: Simultane Überprüfung von mehr als zwei Messreihen auf Gleichheit der Streuungen.

Man benötigt: m statistisch homogene Messreihen mit n_1, s_1 bis n_m, s_m und die χ^2 -Tabelle²² (Die Mittelwerte werden nicht gebraucht.)

Durchführung:

1. Schritt: Berechne $n = n_1 + \dots + n_m = \sum n_i$
2. Schritt: Berechne $f_i = n_i - 1$ ($i = 1, \dots, m$) und $f = n - m$
3. Schritt: Berechne

$$s_{\text{mitt}} = \sqrt{\frac{f_1 s_1^2 + \dots + f_m s_m^2}{f}} = \sqrt{\frac{\sum f_i s_i^2}{f}}$$

²²siehe Tabellen zur Statistik, S.228

4. Schritt: Berechne

$$c = \frac{(\sum \frac{1}{f_i}) - \frac{1}{f}}{3(m-1)} + 1$$

5. Schritt: Berechne die Prüfgröße

$$PB = \frac{(f \cdot \ln(s_{\text{mitt}}) - (\sum f_i \cdot \ln s_i)) \cdot 2}{c}$$

6. Schritt: Schlage für f in der χ^2 -Tabelle nach und vergleiche PB mit dem Tabellenwert.

Auswertung:

- Ist $PB < \chi^2(95\%)$, so sind durch den Test Unterschiede der Streuungen s_1, \dots, s_m "nicht feststellbar".
- Ist $PB > \chi^2(95\%)$, so sind die Streuungen s_i **wahrscheinlich** verschieden (= nicht alle gleich).
- Ist $PB > \chi^2(99\%)$, so sind die Streuungen s_i **signifikant** verschieden (= nicht alle gleich).
- Ist $PB > \chi^2(99,9\%)$, so sind die Streuungen s_i **hochsignifikant** verschieden (= nicht alle gleich).

Bezeichnung: m Messreihen, bei denen Unterschiede der Streuungen nicht erkennbar sind, heißen **bezüglich der Streuung homogen**.

Hat der Bartlett-Test ergeben, dass die Streuungen wahrscheinlich, signifikant oder hochsignifikant verschieden sind, und möchte man daraufhin solche Messreihen aussondern, die für dieses Ergebnis verantwortlich sind, so verfährt man wie folgt:

Während die einzelnen s_i die mittlere Abweichung der Messergebnisse innerhalb der i -ten Messreihe angeben, wird bei s_{mitt} nachträglich und zusätzlich noch über alle Messreihen gemittelt, s_{mitt} gibt also eine Art von mittlerem Wert der s_i an.

Bezeichnung: $s_{\text{mitt}} = \sqrt{\frac{\sum f_i s_i^2}{f}}$ heißt die **mittlere Streuung INNERHALB der Messreihen**.

Aus allen s_i bestimmt man nun ein solches, für das der Abstand $|s_i - s_{\text{mitt}}|$ maximal wird, (s_i ist sozusagen ein "Streuungs-Ausreißer") und sondert die zugehörige Messreihe aus. Für die verbleibenden $m - 1$ Messreihen führt man sämtliche Schritte des Bartlett-Tests erneut durch. Kann der Test nun keine Unterschiede der Streuungen mehr erkennen, so ist man fertig. Andernfalls wiederholt man das Aussonderungsverfahren so lange, bis die verbleibenden Messreihen bezüglich der Streuung homogen sind.

8.9.2 Die Einfache Varianzanalyse zum Vergleich von Mittelwerten $\bar{x}_1, \dots, \bar{x}_m$ (Verallgemeinerter t-Test)

Die Ergebnisse (= Mittelwerte) von m ausreißerfreien Messreihen darf man bekanntlich nur dann in Beziehung zueinander bringen/vergleichen, wenn sie **dieselbe** Zufallsvariable x gemessen haben, wenn sie also bezüglich der Streuung homogen sind. Die sog. **Nullhypothese** für solche Messreihen lautet:

Nullhypothese: Unterschiede zwischen den Mittelwerten \bar{x}_i sind nicht feststellbar, d.h. alle Messreihen haben denselben Erwartungswert μ .

Die Einfache Varianzanalyse:

Ziel: Prüfung, ob die sog. **Nullhypothese** verworfen werden muss.

Man benötigt: m **statistisch homogene** empirische Messreihen mit den statistischen Daten

$$n_i, \bar{x}_i, s_i \quad (i = 1, \dots, m),$$

die den Bartlett-Test oder $\binom{n}{2}$ F-Tests bestanden haben, d.h. dass Unterschiede zwischen den Streuungen s_i nicht feststellbar waren, sowie die **F-Tabelle**.

Durchführung:

1. Schritt: Im Bartlett-Test schon berechnet worden sind die Größen

$$n = \sum n_i,$$

$$f_i = n_i - 1 \text{ für } i = 1, \dots, m \quad \text{und} \quad f = n - m \text{ sowie}$$

$$s_{\text{mitt}} = \sqrt{\frac{\sum f_i s_i^2}{f}}.$$

2. Schritt: Berechne

$$\bar{x} = \frac{\sum n_i \bar{x}_i}{n}$$

3. Schritt: Berechne

$$s_{\bar{x}} = \sqrt{\frac{\sum n_i (\bar{x}_i - \bar{x})^2}{m - 1}}$$

4. Schritt: Berechne

$$F = \left(\frac{s_{\bar{x}}}{s_{\text{mitt}}} \right)^2 \text{ sowie } f_1 = m - 1 \text{ und } f_2 = f.$$

5. Schritt: Ist $F > 1$, so schlage für f_1 und f_2 in der F-Tabelle nach und vergleiche F mit tab- F .

Auswertung:

- Ist $F \leq 1$ oder $F < \text{tab-}F(95\%)$, so ist ein Unterschied der \bar{x}_i "nicht feststellbar", die Nullhypothese ist anzunehmen.
- Ist $F \geq \text{tab-}F(95\% \text{ bzw. } 99\% \text{ bzw. } 99,9\%)$, so sind die \bar{x}_i wahrscheinlich bzw. signifikant bzw. hochsignifikant verschieden (= nicht alle gleich), d.h. die Nullhypothese ist abzulehnen. Die Messreihen gehören mit der entsprechenden Sicherheit zu Experimenten, die nicht alle denselben Erwartungswert besitzen.

Bezeichnung: $s_{\bar{x}} = \sqrt{\frac{\sum n_i (\bar{x}_i - \bar{x})^2}{m - 1}}$ wird auch als **mittlere Streuung ZWISCHEN den Messreihen** bezeichnet.

Ist die Nullhypothese richtig, so ist die mittlere Streuung **zwischen** den Messreihen $s_{\bar{x}}$ kleiner oder kaum größer als die mittlere Streuung **innerhalb** der Messreihen s_{mitt} . Ist das Größenverhältnis $F = \frac{s_{\bar{x}}^2}{s_{mitt}^2}$ besonders klein, so ist die Nullhypothese besonders glaubwürdig, übersteigt dieses Größenverhältnis ein gewisses Maß, so ist die Nullhypothese zu verwerfen. Da in Zähler und Nenner von F zwei Varianzen stehen, heißt der Test "Varianzanalyse".

Anwendungen:

[1] **Aufbau einer größeren Datensammlung zwecks Gewinnung einer idealen Messreihe:**

m ausreißerfreie Messreihen zum selben Experiment, die sowohl den Bartlett-Test als auch die Einfache Varianzanalyse bestanden haben, d.h. dass Unterschiede weder bei den Streuungen s_1, \dots, s_m noch bei den Mittelwerten $\bar{x}_1, \dots, \bar{x}_m$ feststellbar waren, dürfen zu einer einzigen, längeren Messreihe mit den statistischen Daten n , \bar{x} und s zusammengefasst werden. Dabei gilt

$$n = \sum n_i,$$

$$\bar{x} = \frac{\sum n_i \bar{x}_i}{n},$$

$$s = \sqrt{\frac{(n-m)s_{mitt}^2 + (m-1)s_{\bar{x}}^2}{n-1}}.$$

Hinweis: n , \bar{x} , s_{mitt} und $s_{\bar{x}}$ wurden bereits im Verlauf der Einfachen Varianzanalyse berechnet.

[2] **Auswertung verschiedener Experimente zur selben Zufallsvariablen x**

Die Einfache Varianzanalyse zeigt ganz generell, ob die verschiedenen Experimente überhaupt zu unterschiedlichen Ergebnissen (= Erwartungswerten) führen.

Eine besonders häufige Versuchsanordnung ist die Folgende: Man will feststellen, ob die Zufallsvariable

$$x = f(u_1, \dots, u_k, v_1, v_2, \dots)$$

von einem einzelnen u_i , z.B. von u_1 , überhaupt abhängig ist, oder ob u_1 aus der Liste der Einflussgrößen gestrichen werden kann. Dazu betrachtet man x künstlich als Funktion der einen Variablen u_1 und entwirft m verschiedene Experimente, in denen alle u_2, \dots, u_k jedesmal auf denselben Werten konstant gehalten werden sowie alle v_1, v_2, \dots jedesmal innerhalb der gleichen Bandbreiten variieren, nur u_1 setzt man in den m Experimenten auf m verschiedene Werte. Zu jedem Experiment veranstaltet man eine Messreihe.

Ergibt nun die Einfache Varianzanalyse, dass die Nullhypothese anzunehmen ist, so bedeutet das nicht, dass sie richtig ist, sondern lediglich, dass weniger als 95% Wahrscheinlichkeit für die These spricht, dass u_1 irgendeinen Einfluss auf x hat. Ein anderer Ausgang der Einfachen Varianzanalyse gibt an, wie stark/hochgradig der Einfluss von u_1 auf x ist.

Will man das Verhalten von x als Funktion **zweier** Variabler, etwa u_1 und u_2 , studieren, lässt man also in verschiedenen Experimenten beide Variable verschiedene Werte annehmen, so bescheinigt die Einfache Varianzanalyse lediglich, ob sämtliche Experimente zum gleichen Erwartungswert gehören oder nicht. Erweiterte Möglichkeiten ergeben sich mit folgender Versuchsplanung:

Man lasse für **zwei** Variable u_1 und u_2 jeweils m_1 bzw. m_2 verschiedene Werte zu. Durch Kombination dieser zulässigen Werte ergeben sich insgesamt $m_1 \cdot m_2$ verschiedene mögliche Experimente. Besitzt man zu jedem davon eine ausreißerfreie Messreihe, sind all diese Messreihen bezüglich der Streuung homogen und wendet man auf ihre statistischen Daten dann einen anderen Test an, die sog. **Doppelte oder Zweifache Varianzanalyse**, so erhält man in einem Arbeitsgang einen Entscheid, ob keine, eine oder jede der zwei folgenden Nullhypothesen zu verwerfen ist: a) Der Wert von u_1 ist ohne Einfluss auf x , b) der Wert von u_2 ist ohne Einfluss auf x .