Wilfried Rohm, HTL Saalfelden wilfried.rohm@schule.at

Zufallsstreubereiche und Vertrauensbereiche

- Mathematische / Fachliche Inhalte in Stichworten:

  Zufallstreubereiche , Vertrauensbereiche = Konfidenzintervalle, Binomialverteilung, Normalverteilung, Chiquadratverteilung, t-Verteilung.

- Kurzzusammenfassung

  Es erfolgt Berechnung und Darstellung der Zufallsstreubereiche für fehlerhafte Einheiten (Binomialverteilung) als Beispiel einer diskreten Verteilung sowie normalverteilter Größen (Messwerte: Mittelwert und Stichproben-Streuung). Ausserdem erfolgen Berechnung und Darstellung der Vertrauensbereiche für den Anteil fehlerhafter Einheiten (Binomialverteilung) sowie für den Erwartungswert μ und die Standardabweichung σ bzw. Varianz σ^2 bei Vorliegen einer normalverteilten Grundgesamtheit. Diese Datei eignet sich vorallem zu Demonstrationszwecken.

- Didaktische Überlegungen / Zeitaufwand:

  Erfahrungsgemäß haben Schüler immer wieder Schwierigkeiten bei der Unterscheidung der Begriffe "Vertrauensbereich" und "Zufallsstreubereich", was zum Teil auch an der eher kochrezeptartigen Ermittlung dieser Größen für die einzelnen Verteilungen liegt.Die Verwendung eines Computeralgebrasystems wie Mathcad erlaubt speziell bei diskreten Verteilungen eine sehr anschauliche nummersiche Ermittlung dieser Werte sowie eine entsprechende Veranschaulichung, was wesentlich dazu beitragen kann, diese wichtigen Begriffe besser zu verstehen!
  Zur Verwendung als Demonstrationsfile blende man die Regionen für die Berechnungen aus und variiere die farblich hervorgegehobenen Eingabebereiche für Stichprobenumfang (n), Irrtumswahrscheinlichkeit α bzw. den anderen, spezifischen Parametern.

- Lehrplanbezug (bzw. Gegenstand / Abteilung / Jahrgang):

  Angewandte Mathematik, 5.Jahrgang, alle Abteilungen

- Mathcad-Version: **erstellt in der Version 11.**

INHALT (gewünschten Bereich anklicken!):

Begriffsdefinitionen

->Inhalt

Zufallsstreubereiche : Schluß von der Grundgesamtheit auf mögliche Stichprobenergebnisse.
d.h.: Man geht von (gegebenen) Parametern einer bestimmten Verteilung aus und bestimmt daraus ein- oder zweiseitige Bereiche, in denen die interessierenden Stichprobenwerte mit vorgegebener Wahrscheinlichkeit liegen.
Üblicherweise werden in der Praxis dazu statistische Sicherheiten von 90%, 95%, 99% und/oder 99,9% verwendet. In den folgenden Beispielen wird nur mit zweiseitigen Zufallsstreubereichen gearbeitet!

Anwendungen: Shewart-Regelkarten, d.h. Regelkarten, die überprüfen sollen, ob sich eine Verteilung (bezüglich Lage, Streuung, ...) im Laufe der Zeit **verändert!**

Hinweise zur Bezeichnungsweise:

Binomialverteilung: **Parameter der Grundgesamtheit:**
p = fester (gleichbleibender) Prozentsatz für den Anteil fehlerhafter Einheiten.
Stichprobenwerte:
n = Stichprobenumfang
x = Anzahl der gefunden fehlerhaften Einheiten in der Stichprobe.

Normalverteilung: **Parameter der Grundgesamtheit**
μ = Parameter, der die LAGE der Verteilung beschreibt
σ = STANDARDABWEICHUNG als Maß für die Streuung

Stichprobengrößen:
x ... Einzelwerte (Messwerte)
$\bar{x}$ ... Mittelwert aus n Einzelwerten (Messwerten)
s ... Stichprobenstandardabweichung aus n Einzelwerten (Messwerten)

Vertrauensbereiche / Konfidenzintervalle:

Schluß von Stichprobenwerten auf die Parameter der Grundgesamtheit.
d.h.: Ausgehend von einem Stichprobenergebnis soll bestimmt werden, in welchem ein- oder zweiseitigen Bereich der jeweilige Parameter der Grundgesamtheit mit vorgegebener Wahrscheinlichkeit liegt.
Üblicherweise werden in der Praxis dazu statistische Sicherheiten von 90%, 95%, 99% und/oder 99,9% verwendet. In den folgenden Beispielen wird nur mit zweiseitigen Vertrauensbereichen gearbeitet!

Anwendungen: Auswertung von Messserien, Interpretation von Meinungsumfragen, ...

->Inhalt

Zufallsstreubereich einer Diskreten Verteilung - Beispiel Binomialverteilung

->Inhalt

$\alpha := 0.20$ $n := 100$ $p := 0.05$

Berechnung

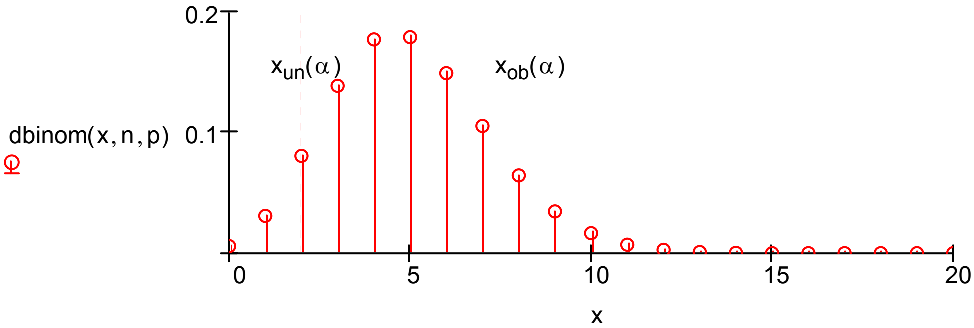
$x_{un}(\alpha) := \text{qbinom}\left(\frac{\alpha}{2}, n, p\right)$ $x_{un}(0.05) = 1$ $x_{un}(0.01) = 0$ $x_{un}(0.20) = 2$

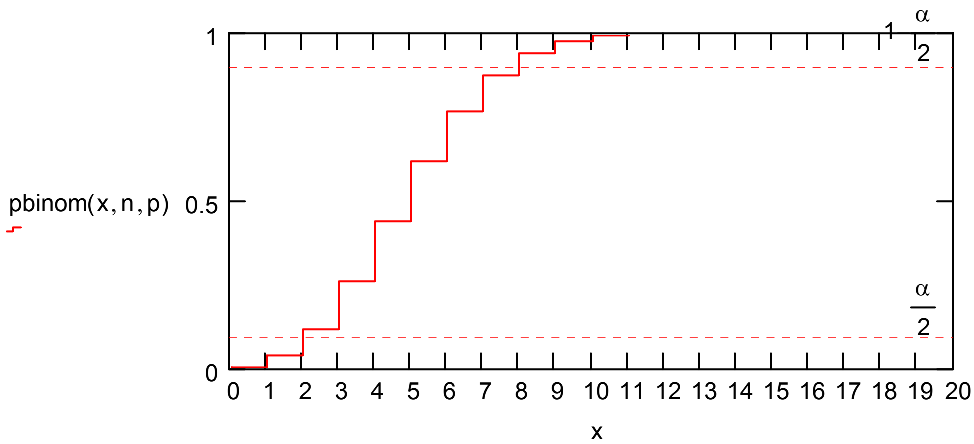
$x_{ob}(\alpha) := \text{qbinom}\left(1 - \frac{\alpha}{2}, n, p\right)$ $x_{ob}(0.05) = 10$ $x_{ob}(0.01) = 11$ $x_{ob}(0.20) = 8$

Zur Erläuterung: Zur Berechnung des Zufallsstreubereiches bei einer diskreten Verteilung muss die Festlegung berücksichtigt werden, dass sowohl an der oberen Grenze wie auch an der unteren Grenze der **größtmögliche Bereich kleiner oder gleich $\alpha/2$** abgeschnitten wird. Daher ist der $(1-\alpha)$-Zufalsstreubereich einer diskreten Verteilung im allgemeinen größer als $(1-\alpha)$.
Diese Festlegung berücksichtigt der Befehl der inversen Binomialverteilung zur Berechung des jeweiligen Quantils der Binomialverteilung (**qbinom**) automatisch, da die Funktion **qbinom(G,n,p)** die kleinste ganze Zahl **k** zurückliefert, für die $\text{pbinom}(k, n, p) \geq G$ gilt.

Berechnung

$x_{un}(\alpha) = 2$ $x_{ob}(\alpha) = 8$ $x := 0..20$





Interpretation: Bei einer Stichprobe des Umfangs n aus einem Los mit (gleichbleibendem!) Schlechtanteil p ist die Wahrscheinlichkeit zumindest $1-\alpha$, dass die Anzahl der "schlechten" Stücke in der Stichprobe zwischen $x_{un}(\alpha)$ und $x_{ob}(\alpha)$ liegt!

Vertrauensbereich einer Diskreten Verteilung - Beispiel Binomialverteilung

->Inhalt

Beispiel:
Bei einer Wahlumfrage werden n=500 gefragt, "welche Partei sie am nächsten Sonntag wählen würden". 52 der gerfagten Personen entschieden sich für die "GRAUE PARTEI".
Wie groß ist demnach (mit einer Irrtumswahrscheinlichkeit <= 5%) der Anteil der GRAU-Wähler in der gesamten Bevölkerung?

Lösung über den Vertrauensbereich der Binomialverteilung

$n := 500$ $\quad x := 52$ $\quad \alpha := 10\%$ $\quad$ Schätzwert aus der Stichprobe $\frac{x}{n} = 0.104$

Berechnung

$p := 0.10$ $\quad x_{geg} := x$

Vorgabe

$$\text{pbinom}(x, n, p) = 1 - \frac{\alpha}{2}$$

$p_{unten} := \text{Suchen}(p)$ $\quad p_{unten} = 0.0841$

$p := 0.50$

Vorgabe

$$\text{pbinom}(x, n, p) = \frac{\alpha}{2}$$

$p_{oben} := \text{Suchen}(p)$ $\quad p_{oben} = 0.12923$

Zur Erläuterung: Die Berechnung des Vertrauensbereiches der Binomialverteilung führt auf eine recht komplizierte Formel, aufbauend auf der F-Verteilung.
Mit Hilfe eines Computeralgebrasystemes kann die Berechnung **numerisch** erfolgen und wird dadurch wesentlich verständlicher! Wir lösen einfach obige Gleichungen.

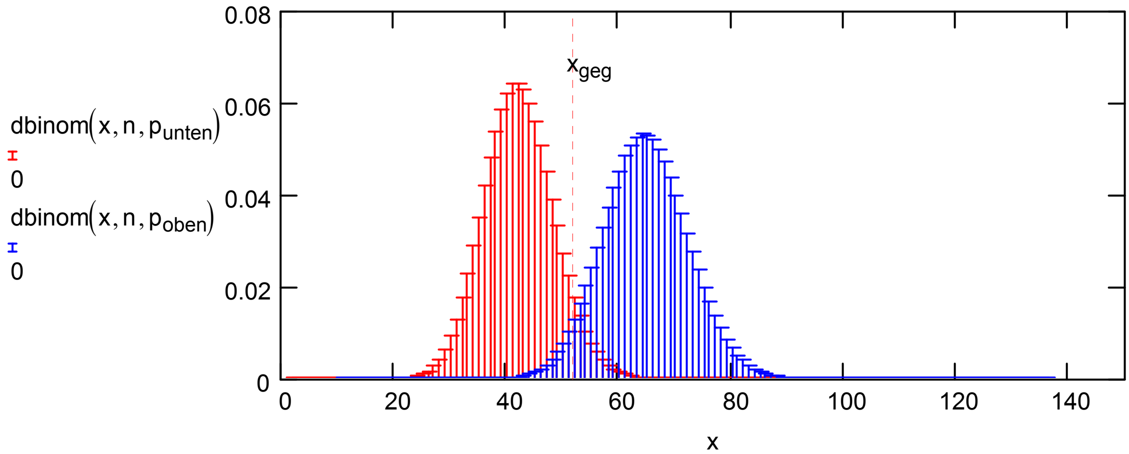
Damit wird $\mathbf{p_{unten}}$ so bestimmt, dass der Wert x der Stichprobe gerade am **OBEREN RAND** des Zufallstreubereiches der Verteilung für p_{unten} zu liegen kommt.
Daher ist die Gleichung $\mathbf{G_{bi}(p_{unten}) = 1 - a/2}$ nach p_{unten} aufzulösen.

Und $\mathbf{p_{oben}}$ wird so bestimmt, dass der Wert x der Stichprobe gerade am **UNTEREN RAND** des Zufallstreubereiches der Verteilung für p_{oben} zu liegen kommt.
Daher ist die Gleichung $\mathbf{G_{bi}(p_{oben}) = a/2}$ nach p_{oben} aufzulösen.

$\mu := n \cdot \frac{x}{n}$ $\quad x := 0 .. 4 \cdot \mu$

Berechnung

$p_{unten} = 0.0841$ $\quad p_{oben} = 0.12923$



Interpretation : Derr Anteil der "Grau-Wähler" in der Grundgesamtheit bzw. der "wahre Parameter" p liegt - ausgehend von unserem Stichprobenergebnis - mit einer Wahrscheinlichkeit von (zumindest) 1-α im Bereich

$p_{unten} \le p \le p_{oben}$

Die Grenzverteilungen für p_{unten} bzw. für p_{oben} sind in der Zeichnung rot bzw. blau eingezeichnet!

nhalt

Normalverteilung - Zufallsstreubereich für Einzelwerte x

->Inhalt

$\alpha := 0.10$

$\mu := 100$ $\sigma := 0.1$

Berechnung

$$x_{un} := \mu - \text{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) \cdot \sigma$$

$x_{un} = 99.83551$

$$x_{ob} := \mu + \text{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) \cdot \sigma$$

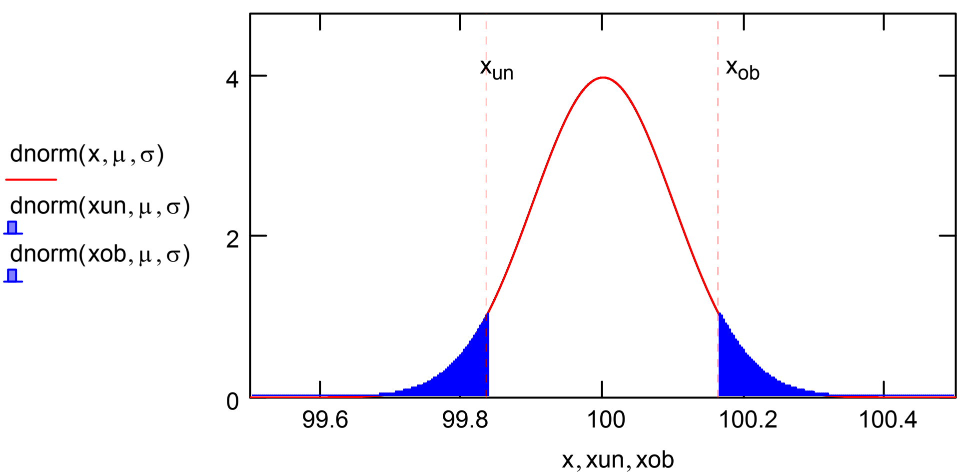
$x_{ob} = 100.16449$

$$x := \mu - 6\sigma, \mu - 6\sigma + \frac{\sigma}{1000} .. \mu + 6\sigma$$

$$xun := \mu - 6 \cdot \sigma, \mu - 6 \cdot \sigma + \frac{\sigma}{1000} .. x_{un}$$

$$xob := x_{ob}, x_{ob} + \frac{\sigma}{1000} .. \mu + 6 \cdot \sigma$$

Berechnung



Interpretation : Mit einer Wahrscheinlichkeit von (1-α) liegt ein Merkmalswert aus einer normalverteilten Grundgesamtheit mit den Parametern μ und σ im Bereich $x_{un} \le x \le x_{ob}$

->Inhalt

Normalverteilung: Zufallsstreubereich für $\bar{x}$ (bei gegebenem σ)

->Inhalt

$n := 10$ $\alpha := 0.05$ $\mu := 100$ $\sigma := 0.1$

Berechnung

$$xq_{un} := \mu - \text{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) \cdot \frac{\sigma}{\sqrt{n}}$$ $$xq_{un} = 99.93802$$

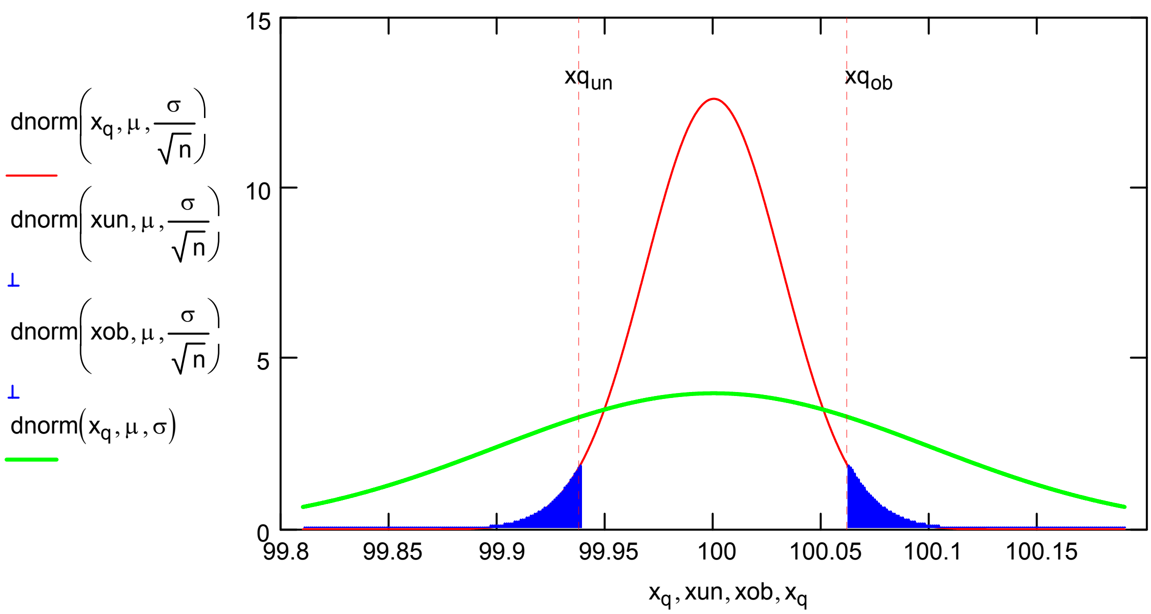
$$xq_{ob} := \mu + \text{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) \cdot \frac{\sigma}{\sqrt{n}}$$ $$xq_{ob} = 100.06198$$

$$x_q := \mu - 6\frac{\sigma}{\sqrt{n}}, \mu - 6\frac{\sigma}{\sqrt{n}} + \frac{\sigma}{1000} .. \mu + 6\frac{\sigma}{\sqrt{n}}$$

$$xun := \mu - 6 \cdot \frac{\sigma}{\sqrt{n}}, \mu - 6 \cdot \frac{\sigma}{\sqrt{n}} + \frac{\sigma}{1000} .. xq_{un}$$

$$xob := xq_{ob}, xq_{ob} + \frac{\sigma}{1000} .. \mu + 6 \cdot \frac{\sigma}{\sqrt{n}}$$

Berechnung



Interpretation : Mit einer Wahrscheinlichkeit von 1-α liegt der Mittelwert $\bar{x}$ von n Messwerten (Merkmalswerten) aus einer normalverteilten Grundgesamtheit (μ,σ) im folgenden Bereich: $xq_{un} \le x_q \le xq_{ob}$

Der Vergleich mit der grün eingezeichneten Verteilung für die Einzelwerte zeigt, dass die Mittelwerte mit der Standardabweichung $\sigma_{xq} = \frac{\sigma}{\sqrt{n}}$ entsprechend weniger streuen als die Einzelwerte!

Die Fläche unter beiden Kurven muss 1 ergeben!

->Inhalt

Normalverteilung: Vertrauensbereich für μ (bei gegebenem σ)

->Inhalt

$n := 10$ $\alpha := 5\%$ $x_{quer} := 100$ $\sigma := 0.1$

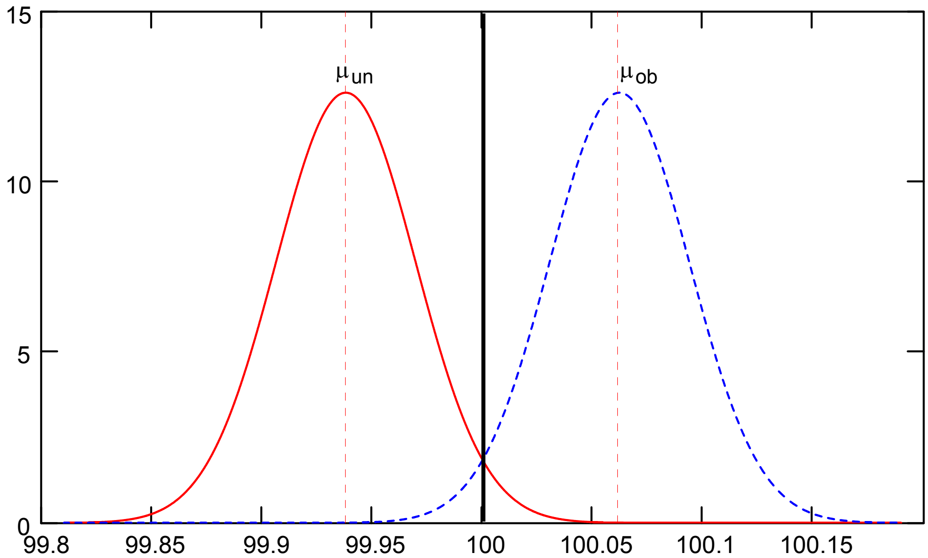
Berechnung

$$\mu_{un} := x_{quer} - \text{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) \cdot \frac{\sigma}{\sqrt{n}} \qquad \mu_{un} = 99.93802$$

$$\mu_{ob} := x_{quer} + \text{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) \cdot \frac{\sigma}{\sqrt{n}} \qquad \mu_{ob} = 100.06198$$

$$x_q := x_{quer} - 6\frac{\sigma}{\sqrt{n}}, x_{quer} - 6\frac{\sigma}{\sqrt{n}} + \frac{\sigma}{1000} .. x_{quer} + 6\frac{\sigma}{\sqrt{n}}$$

Berechnung



Interpretation : Bei einer Messung einer Merkmalsgrösse durch Mittelung von n Einzelwerten stellt sich die Frage, wie genau diese Schätzung für den Parameter μ einer normalverteilten Grundgesamtheit ist. Der Vertrauensbereich (bzw. das Konfidenzintervall) sagt nun aus, dass das "wahre μ" der Grundgesamtheit mit einer Wahrscheinlichkeit 1-α im folgenden Bereich liegt: $\mu_{un} \le \mu \le \mu_{ob}$

Im Diagramm sind die Grenzverteilungen rot bzw. blau eingezeichnet, der Schätzwert gemäß der Stichprobe ist der mittlere Schwarze Strich!

->Inhalt

Normalverteilung: Vertrauensbereich für μ bei gegebenem $\bar{x}$ und s (Vergleich mit σ gegeben: grüne Linie)

->Inhalt

$n := 10$ $\alpha := 5\%$ $x_{quer} := 100$ $s := 0.1$

Berechnung

$t(x, f) := qt(x, f)$

$$\mu_{un} := x_{quer} - t\left(1-\frac{\alpha}{2}, n-1\right)\cdot\frac{s}{\sqrt{n}} \qquad \mu_{un} = 99.92846$$

$$\mu_{ob} := x_{quer} + t\left(1-\frac{\alpha}{2}, n-1\right)\cdot\frac{s}{\sqrt{n}} \qquad \mu_{ob} = 100.07154$$

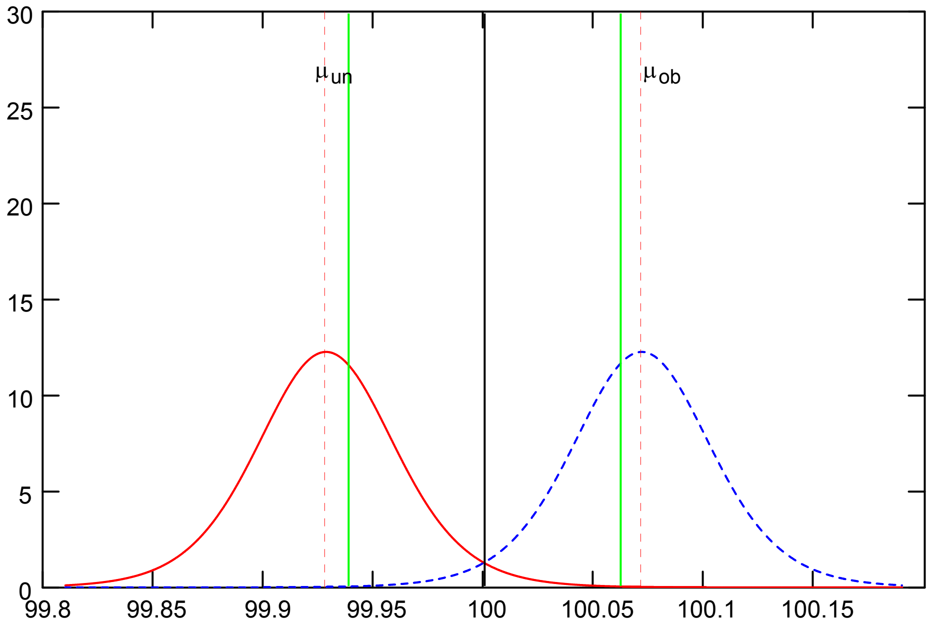
$$\mu\sigma_{un}(n) := x_{quer} - \mathrm{qnorm}\left(1-\frac{\alpha}{2}, 0, 1\right)\cdot\frac{\sigma}{\sqrt{n}} \qquad \mu\sigma_{un}(n) = 99.93802$$

$$\mu\sigma_{ob}(n) := x_{quer} + \mathrm{qnorm}\left(1-\frac{\alpha}{2}, 0, 1\right)\cdot\frac{\sigma}{\sqrt{n}} \qquad \mu\sigma_{ob}(n) = 100.06198$$

$$x_q := x_{quer} - 6\frac{s}{\sqrt{n}}, x_{quer} - 6\frac{s}{\sqrt{n}} + \frac{s}{1000} .. x_{quer} + 6\frac{s}{\sqrt{n}}$$

$$t_u(x_q) := \frac{x_q - \mu_{un}}{\frac{s}{\sqrt{n}}} \qquad t_o(x_q) := \frac{x_q - \mu_{ob}}{\frac{s}{\sqrt{n}}}$$

Berechnung



Interpretation : Bei einer Messung einer Merkmalsgrösse durch Mittelung von n Einzelwerten stellt sich die Frage, wie genau diese Schätzung für den Parameter μ einer normalverteilten Grundgesamtheit ist. Der Vertrauensbereich (bzw. das Konfidenzintervall) sagt nun aus, dass das "wahre μ" der Grundgesamtheit mit einer Wahrscheinlichkeit 1-α im folgenden Bereich liegt: $\mu_{un} \le \mu \le \mu_{ob}$

Im Diagramm sind die Grenzverteilungen rot bzw. blau eingezeichnet, der Schätzwert gemäß der Stichprobe ist der mittlere Schwarze Strich!
Es ist aber ein Unterschied in der Genauigkeit der Schätzung, ob die Streuung der Grundgesamtheit als bekannt vorausgesetzt werden kann (dies ergibt die grün eingezeichneten Grenzen für μ_{un} bzw. μ_{ob}) oder ob diese Standardabweichung aus den n Stichprobenwerten mit der Stichprobenstandardabweichung s erst abgeschätzt werden muss. Dann muss mit der t-Verteilung gerechnet werden, was zu einem entsprechend größeren Vertrauensbereich für μ führt (gestrichelte rote Linien = größere Unsicherheit!). Der Unterschied wird allerdings immer kleiner, je größer n wird, weil die t-Verteilung für n gegen Unendlich in die Normalverteilung übergeht!

->Inhalt

Zufallsstreubereich für s

->Inhalt

$\alpha := 0.10$ $n := 20$ $\sigma := 0.1$

$\chi_{G,f} = (n-1)\cdot\frac{s^2}{\sigma 2}$

Berechnung

$\mathrm{chi}(G,f) := \mathrm{qchisq}(G,f)$

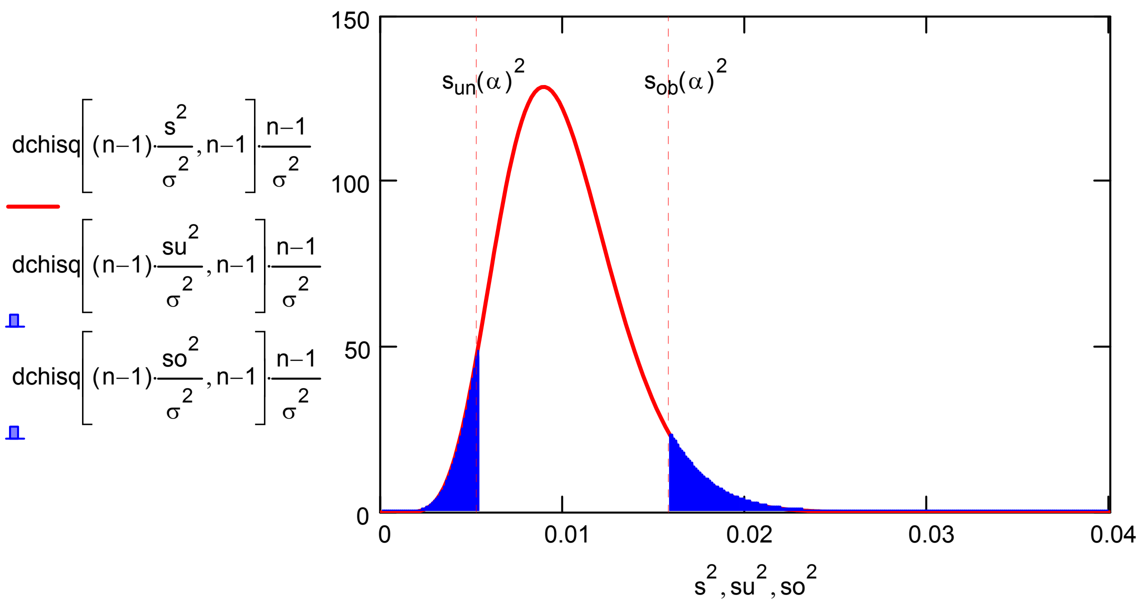
ZSB - Formeln

$$s_{un}(\alpha) := \sqrt{\frac{\mathrm{chi}\left(\frac{\alpha}{2}, n-1\right)}{n-1}}\cdot\sigma \qquad s_{ob}(\alpha) := \sqrt{\frac{\mathrm{chi}\left(1-\frac{\alpha}{2}, n-1\right)}{n-1}}\cdot\sigma$$

$s_{un}(0.05) = 0.06847$ $s_{ob}(0.05) = 0.13149$

$$s := 0, \frac{\sigma}{100} .. 2\cdot\sigma \qquad su := 0, \frac{\sigma}{1000} .. s_{un}(\alpha) \qquad so := s_{ob}(\alpha), s_{ob}(\alpha) + \frac{\sigma}{1000} .. 2\cdot\sigma$$

Berechnung



Interpretation : Mit einer Wahrscheinlichkeit von 1-α liegt die Stichproben-Standardabweichung von n Messwerten (Merkmalswerten) aus einer normalverteilten Grundgesamtheit (μ,σ) im folgenden Bereich: $s_{un} \le s \le s_{ob}$

Da die Verteilung der Stichprobenvarianzen einer Chiquadratverteilung mit (n-1) Freiheitsgraden folgt, ist dieser Bereich NICHT symmetrisch um s gelegen! Allerdings wird dieser Bereich bzw. die Chiquadratverteilung für größer werdendes n immer symmetrischer, weil die Chiquadratverteilung für n gegen Unendlich in die Normalverteilung übergeht!

->Inhalt

Vertrauensbereich für σ

->Inhalt

$\alpha := 0.10$ $n := 10$ $s_{geg} := 0.1$

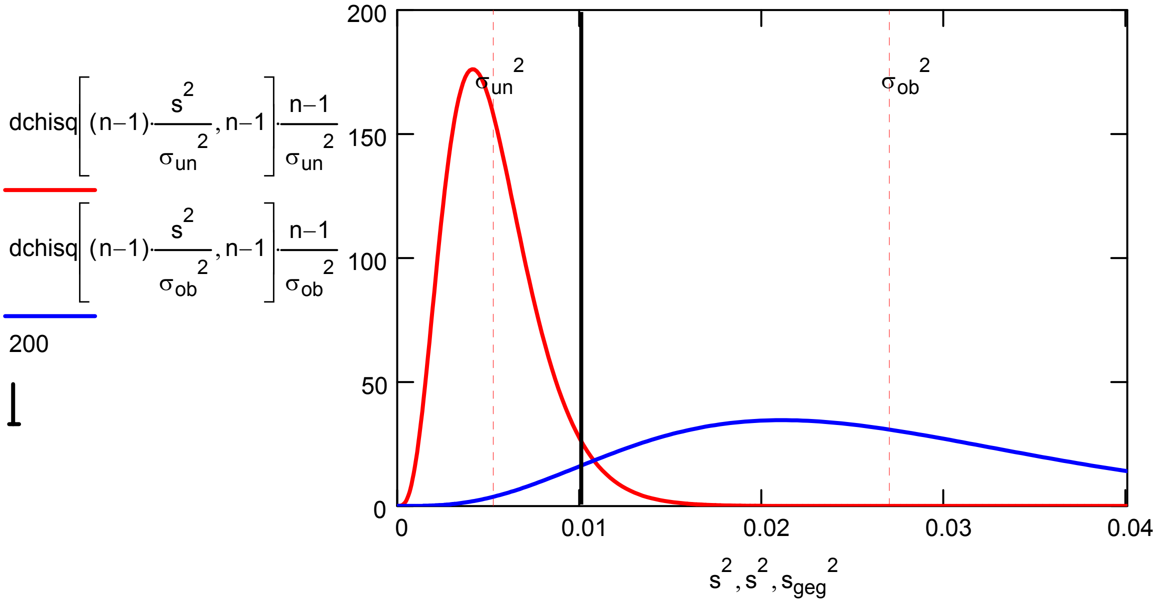
Berechnung

$$\sigma_{un} := \sqrt{\frac{n-1}{\mathrm{chi}\left(1-\frac{\alpha}{2}, n-1\right)}} \cdot s_{geg} \qquad \sigma_{ob} := \sqrt{\frac{n-1}{\mathrm{chi}\left(\frac{\alpha}{2}, n-1\right)}} \cdot s_{geg}$$

$\sigma_{un} = 0.07293$ $\sigma_{ob} = 0.16452$

$$s := 0, \frac{s_{geg}}{100} .. 2 \cdot s_{geg}$$

Berechnung



Interpretation : Bei der Schätzung der Standardabweichung σ aus einer Stichprobe von n Einzelwerten durch Ermittlung der Größe s stellt sich die Frage, wie genau diese Schätzung ist. Der Vertrauensbereich (bzw. das Konfidenzintervall) sagt nun aus, dass das "wahre σ" der Grundgesamtheit mit einer Wahrscheinlichkeit 1-α im folgenden Bereich liegt: $\sigma_{un} \le \sigma \le \sigma_{ob}$

Im Diagramm sind die Grenzverteilungen rot bzw. blau eingezeichnet, der Schätzwert s^2 gemäß der Stichprobe ist der mittlere Schwarze Strich!

->Inhalt