Wilfried Rohm wrohm@aon.at

Das "Alkomatproblem" - Statistische Tests

Link zur Beispielsübersicht

- Mathematische / Fachliche Inhalte in Stichworten:

  Statistische Tests, t-Test paarweise unverbunden, Vertrauensbereich für Mittelwert und Standardabweichung, Kalibrierung, Wahrscheinlichkeitsrechnung, Statistik.

- Kurzzusammenfassung

  In diesem Beispiel wird Bezug auf ein 2002 im Bundesland Salzburg aufgetretenes Problem der Genauigkeit bzw. Zuverlässigkeit von Messergebnissen mit Alkomaten genommen. Ein statistischer Test für den Vergleich von 2 Meßreihen soll durchgeführt werden. Weitere Fragestellungen aus der Wahrscheinlichkeitsrechung / Statistik ausgehend von diesem Beispiel sind zu beantworten.

- Lehrplanbezug: **Angewandte Mathematik, 5.Jahrgang, alle Abteilungen**

Aufgabenstellung

Im Jahr 2002 wurde bekannt, dass die Alkomaten der Polizei im Bundesland Salzburg nicht zuverlässig genug messen - ein zunächst abgestrafter Autofahrer wurde per Gerichtsentscheid schließlich freigesprochen, weil "nicht mit erforderlicher Sicherheit nachgewiesen werden könne, dass der Mann mit 1,6 Promille oder mehr gefahren sei". Es wurde bekannt, dass Alkomaten im Ausland (Deutschland) nach 2 verschiedenen Methoden messen und daher eine "gesichertere" Aussage über den Promillegehalt eines Autolenkers möglich erscheint. Bei den Salzburger Alkomaten wird nur nach nur einem Verfahren gemessen, "wodurch der Einfluß von Fremdsubstanzen nicht ausgeschlossen werden kann". Dies ist der Hintergrund folgender Fragestellung:

2 verschiedene Verfahren sollen verglichen werden, ob sich ihre Ergebnisse "signifikant" unterscheiden oder ob sie als gleichwertig eingestuft werden können. Dazu werden 10 verschiedene Testpersonen unterschiedlichen Mengen Alkohol "ausgesetzt" und ihre Atem-Alkoholkonzentration unmittelbar hintereinander nach beiden Verfahren gemessen.
Man erhielt folgende Ergebnisse (Alkoholwerte in mg/l Atemluft)

Person – Nr	Verfahren A	Verfahren B
1	0.12	0.17
2	0.19	0.22
3	0.25	0.25
4	0.53	0.64
5	0.66	0.69
6	0.87	0.91
7	0.96	0.94
8	1.12	1.22

a) Wie beurteilst du das Ergebnis? Berücksichtige dabei, dass deine Untersuchungen weitreichende Auswirkungen haben können (Einbau einer 2.Messmethode in die Salzburger Alkomaten). Erläutere dabei den verwendeten Test und die verwendeten Formelsätze.

b) Erkläre an Hand der verwendeten Formeln möglichst genau und ausführlich, wie sich ein höherer Stichprobenumfang auf das Testergebnis auswirken würde.

c) Zur Kalibrierung des Gerätes wird eine mit Alkohol angereicherte "Testluft" mit 1,00 mg/l Alkoholgehalt nach Verfahren A unter "Wiederholbedingungen" 20-mal gemessen. Man erhielt eine Mittelwert von 1.03 mg mit einer Standardabweichung von 0.025 mg. Welche Aussage(n) kannst du auf Grund dieses Ergebnisses treffen? (Vorgangsweise und Ergebnis sollen einem "statistischen Laien" entsprechend erklärt werden)

Teil a) Durchführung eines t-Tests für paarweise verbundene Messwerte

Begründung : Über die Personen-Nummer ist eine paarweise Bindung der Messwerte gegeben. Näheres zu Testauswahl bzw. zu diesem Test im Beitrag "Statistische Tests" von Wilfried ROHM.

$n := 8$

$ORIGIN := 1$ $i := 1 .. n$ $VerfahrenA_i :=$ $VerfahrenB_i :=$

VerfahrenA	VerfahrenB
0.12	0.17
0.19	0.22
0.25	0.25
0.53	0.64
0.66	0.69
0.87	0.91
0.96	0.94
1.12	1.22

Bestimmung von Mittelwert und Standardabweichung der Differenzen

a) mit Hilfe vordefinierter Funktionen

$d_{quer} := \text{mittelwert}(VerfahrenA - VerfahrenB)$ $s_d := \text{Stdev}(VerfahrenA - VerfahrenB)$

$d_{quer} = -0.042$ $s_d = 0.045$

b) Mit Hilfe selbst errechneter Werte $n := \text{länge}(VerfahrenA)$ $n = 8$

$$d_{quer} := \frac{1}{n} \cdot \sum_i \left(VerfahrenA_i - VerfahrenB_i\right) \qquad s_d := \sqrt{\frac{1}{n-1} \cdot \sum_i \left(VerfahrenA_i - VerfahrenB_i - d_{quer}\right)^2}$$

$d_{quer} = -0.042$ $s_d = 0.045$

$g_{nv}(x, \mu, \sigma) := \text{dnorm}(x, \mu, \sigma)$

$g_t(x) := dt(x, n-1)$

Definition der benötigten Dichtefunktionen, Benennung gemäß der an der HTL üblichen Schreibweise

Durchführung des t - Testes

Null-Hypothese H0 $d = \mu_1 - \mu_2 = 0$ Die Differenz der beiden Mittelwerte beträgt 0. (kein systematischer Unterschied zwischen den beiden Verfahren!)

Gegenhypothese H1: $d \neq 0$ Signifikanter Unterschied der Mittelwerte der beiden Verteilungen (zweiseitig)

Stichprobenwerte $d_{quer} = -0.042$ $s_d = 0.045$

Voraussetzung Der t- Test für paarweise verbundene Stichproben hat gegenüber dem t-Test für voneinander unabhängige Stichproben deutlich weniger Voraussetzungen. Über die Varianzen sind keine Aussagen nötig, selbst die Einzelwerte müssen nicht normalverteilt sein, lediglich die Differenzen müssen (annähernd) normalverteilt sein, was grundsätzlich auf Grund der Erfahrung als gegeben angenommen werden kann.

Ermittlung der Prüfgröße $f := n - 1$ $f = 7$ Anzahl der Freiheitsgrade

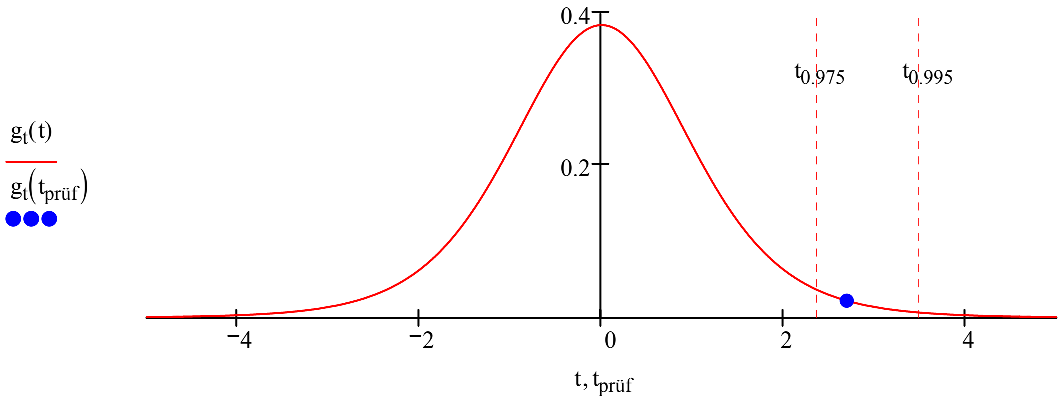
$$t_{prüf} := \frac{|d_{quer}|}{\frac{s_d}{\sqrt{n}}}$$ $t_{prüf} = 2.693$

Kritische Werte

$t_{0.975} := qt(0.975, n - 1)$	$t_{0.975} = 2.365$	$\alpha = 5\%$
$t_{0.995} := qt(0.995, n - 1)$	$t_{0.995} = 3.499$	$\alpha = 1\%$
$t_{0.9995} := qt(0.9995, n - 1)$	$t_{0.9995} = 5.408$	$\alpha = 0.1\%$

Grafische Darstellung

$t := -5, -4.99 .. 5$

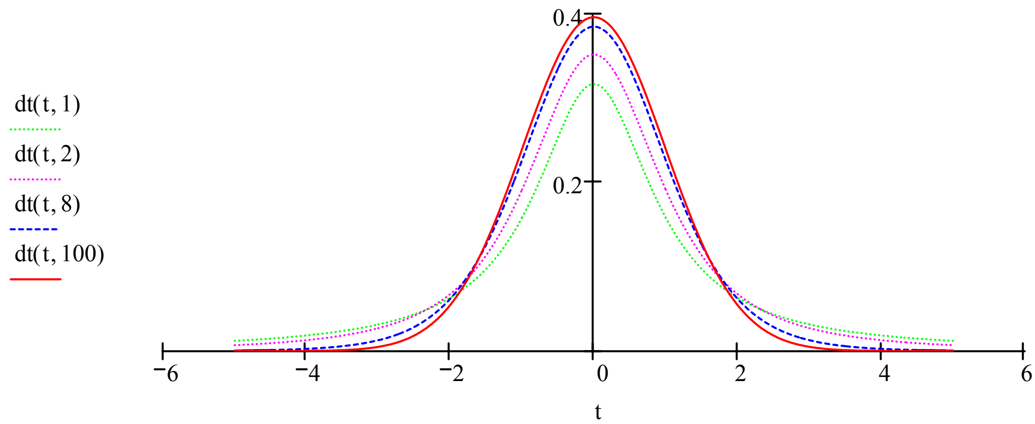


Interpretation des Ergebnisses

Die Ablehnung der Nullhypothese, dass zwischen den beiden Verfahren **kein** systematischer Unterschied besteht, erfolgt auf dem 5%-Niveau -jedoch nicht auf dem 1%-Niveau.
In der technischen Praxis wird diese Situation häufig als **INDIFFERENT** bezeichnet. Da ausserdem laut Angabe das Testergebnis über hohe Investitionen entscheiden soll, wäre zu empfehlen, die Versuchsdurchführung so zu ändern, dass eine "schärfere" Aussage möglich erscheint.
Dies kann durch eine Erhöhunger des Stichprobenumfanges (Untersuchung der Alkoholwerte von **mehr** Personen) erreicht werden - siehe dazu auch Teil b) der Aufgabenstellung!

Teil b) Auswirkungen eines höheren Stichprobenumfanges auf den Test

b) t-Verteilung mit verschiedenen Freiheitsgraden:



Bei höherem Stichprobenumfang geht die t-Verteilung in eine Normalverteilung über und wird daher immer schmäler. Gleichzeitig dazu werden auch die **kritischen Werte immer kleiner.**

Im Gegensatz dazu verhält sich die **Prüfgröße** proportional zu $\sqrt{n}$, dass bedeutet, dass sie mit steigendem n **immer größer** wird:

$$t_{Pr} = \frac{|d_{quer}|}{s_d} \cdot \sqrt{n}$$

Daraus folgt, dass mit steigendem n die Schärfe des Tests immer weiter zunimmt.

Wie einer meiner Schüler bei der Klausurarbeit zeigte, kann man dies anschaulich noch schöner demonstrieren , indem man Prüfgröße und kritische Werte in Abhängigkeit von variablen Stichprobenumfang für die gegebenen Werte vergleicht:

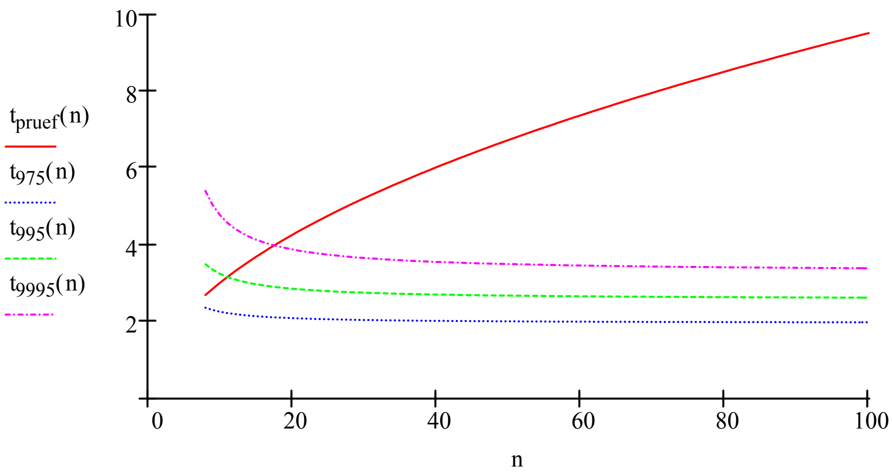
$n := 8 .. 100$

$$t_{pruef}(n) := \frac{|d_{quer}|}{\frac{s_d}{\sqrt{n}}}$$

$$t_{975}(n) := qt(0.975, n-1)$$

$$t_{995}(n) := qt(0.995, n-1)$$

$$t_{9995}(n) := qt(0.9995, n-1)$$



Man sieht in dieser Darstellung, die Veränderung des Ergebnisses bei höheren Stichprobenumfängen.

Würde DASSELBE Ergebnis (xquer und sd) bei höheren Stichprobenumfängen (z.B. 30) gemessen worden sein, so könnte eine hochsignifikante Ablehnung der Nullhypothese auf dem Niveau von 99.9% durchgeführt werden !!

In dieser Darstellung sieht man auch schön, das bereits ab einem Stichprobenumfang von ca. 19 eine signifikante Ablehnung (Irrtumswahrscheinlichkeit < 0.1 %) der Nullhypothese erfolgt.

Teil c) Zur Kalibrierung des Gerätes

Grundsätzlich erlauben die gegebenen Werte 3 Aussagen:

1) In welchem Bereich liegt der Alkoholwert der Testluft auf Grund der Messungen durch den Alkomaten mit vorgegener Genauigkeit (es wird hier gewählt: Irrtumswahrscheinlichkeit α = 1%) - dies kann durch die Ermittlung des 1-α-Vertrauensbereiches für μ erfolgen.

2) Ein Vergleich des Ergebnisses von 1) mit dem tatsächlichen Alkoholgehalt der Testluft (1,00 mg/l = als "wahrer Wert" angenommen) erlaubt die Feststellung, ob ein systematischer Fehlers des Meßgerätes vorliegt oder ob die Abweichungen "im Rahmen des Zufalls liegen"

3) Analog zu 1) kann auch die Ermittlung des 1-α-Vertrauensbereiches für die Standardabweichung σ des Gerätes erfolgen.

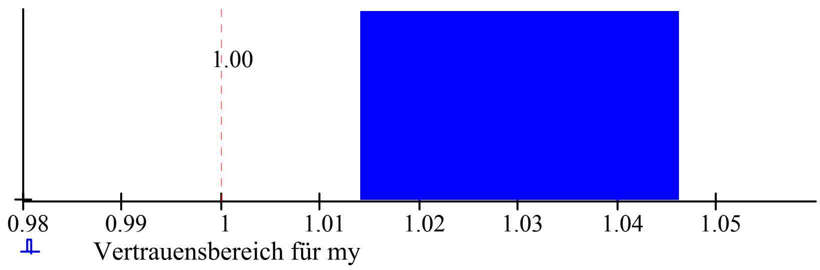
ad 1) und 2) Berechnung des Vertrauensbereiches (99%) für μ

$x_{quer} := 1.03 \qquad s := 0.025 \qquad n := 20 \qquad f := n - 1 \qquad \alpha := 0.01$

$$\mu_{un} := x_{quer} - qt\left(1 - \frac{\alpha}{2}, f\right) \cdot \frac{s}{\sqrt{n}} \qquad \mu_{un} = 1.014$$

$$\mu_{ob} := x_{quer} + qt\left(1 - \frac{\alpha}{2}, f\right) \cdot \frac{s}{\sqrt{n}} \qquad \mu_{ob} = 1.046$$

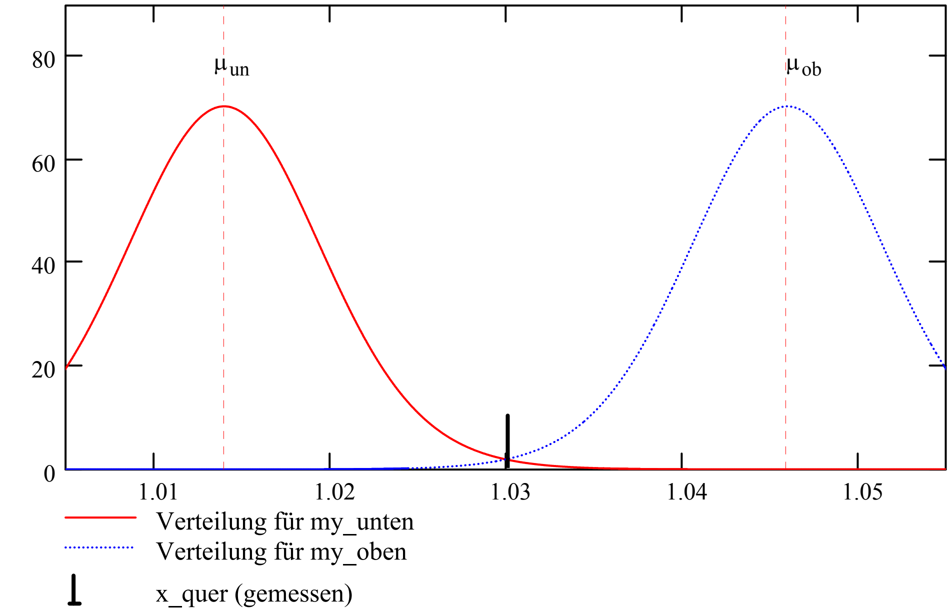
$x := \mu_{un}, \mu_{un} + 0.00001 .. \mu_{ob}$



Der als "wahrer Wert" angegnommene Wert von 1,00 mg Alkoholgehlat pro Liter liegt ausserhalb des Vertrauensbereiches für μ (siehe obige Skizze!). Das bedeutet, dass unser Gerät einen offenbar zu großen Wert misst, d.h. einen systematischen Fehler aufweist. Diese Aussage wird mit einer Irrtumswahrscheinlichkleit kleiner als 1% getätigt.

Im folgenden werden noch die ermittelten "Grenzverteilungen" gezeichnet, also jene Verteilungen, die hinter µ_unten bzw. µ_oben stehen. Der gemessene Mittelwert x_quer ist demnach jener Wert, der gerade an der oberen bzw. unteren Grenze des Zufallsstreubereiches unter der Annahme von µ=µ_unten bzw. µ=µ_oben liegt.
Hinweis: Die Formatierung der Ordinate muss so erfolgen, dass eine Dichtefunktion vorliegt (Fläche unter der Kurve = 1!)

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



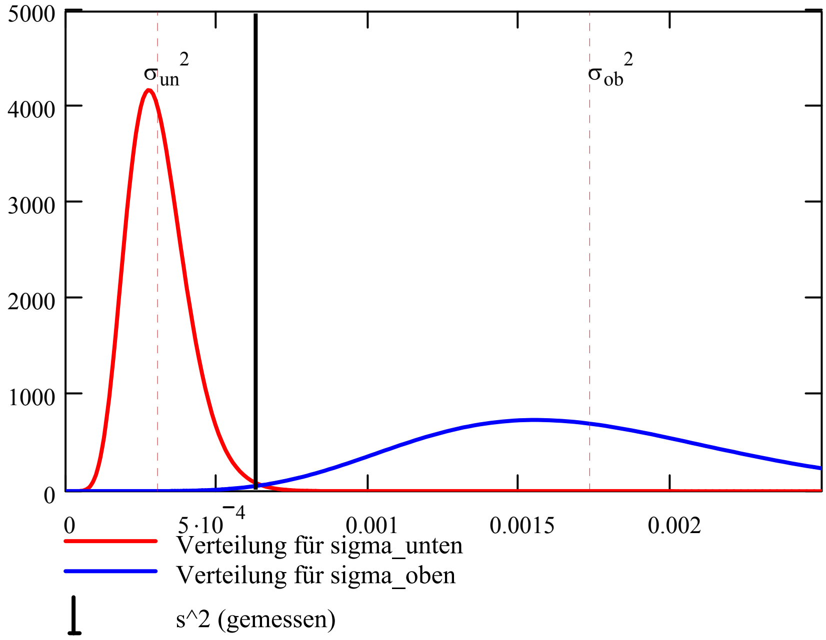
ad 3) 99%-Vertrauensbereich für σ

$$\sigma_{un} := \sqrt{\frac{n-1}{\text{qchisq}\left(1-\frac{\alpha}{2}, f\right)}} \cdot s \qquad \sigma_{un} = 0.018$$

$$\sigma_{ob} := \sqrt{\frac{n-1}{\text{qchisq}\left(\frac{\alpha}{2}, f\right)}} \cdot s \qquad \sigma_{ob} = 0.042$$

Die nachfolgende Skizze zeigt die "Grenzverteilungen" gemäß der Chiquadratverteilung.
Auf der Abszisse wird s^2 unter der Annahme eines σ=σ_un bzw. σ=σ_ob aufgetragen. Eingezeichnet ist auch der Wert für s^2, der aus den Messwerten ermittelt wurde (senkrechter schwarzer Strich).
Von σ_un bis σ_ob reicht der Vertrauensbereich für σ, also jener Bereich, in dem "das wahre σ" mit einer Wahrscheinlichkeit von 1-α (hier: 99%) liegt.

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



[Link zur Beispielsübersicht]