

Statistik

04. z-Test und t-Test

Roland Schäfer

Institut für Germanistische Sprachwissenschaft
Friedrich-Schiller-Universität Jena

stets aktuelle Fassungen: <https://github.com/rsling/VL-Deutsche-Syntax>

1 Übersicht

2 Wiederholungen

- Logik von statistischen Tests

3 t-Test

- t-Test mit einer Stichprobe
- t-Test mit zwei Stichproben

4 Nächste Woche | Überblick

Übersicht

- Wann sind Unterschiede zwischen Mittelwerten signifikant?
- Mittelwerte in Grundgesamtheiten und Stichproben

- Gravetter & Wallnau (2007)
- Bortz & Schuster (2010)
- oder eben gleich Fisher (1935)

Wiederholungen

- 1 **Nullhypothese** (H_0) festlegen: Der theoretisch angenommene Effekt existiert **nicht** (z. B.: Die Versuchsperson [VP] kann **nicht** erkennen, ob Tee oder Milch zuerst in der Tasse war).
- 2 **Stichprobengröße** und **Versuchsaufbau** festlegen (z. B. acht Tassen mit vier *Tee zuerst*-Tassen; VP kennt das Verhältnis)
- 3 **sig-Niveau** festlegen: Wie unwahrscheinlich darf das Ergebnis unter Annahme der H_0 sein, damit wir die H_0 zurückweisen.
- 4 Experiment durchführen, Ergebnis messen.
- 5 **p-Wert** berechnen: Wie wahrscheinlich **war** es, dieses Ergebnis oder ein extremeres Ergebnis zu erreichen, wenn die H_0 die Welt korrekt beschreibt.
- 6 Wenn $p \leq \text{sig}$, dann H_0 zurückweisen: Entweder der Effekt existiert (z. B. die VP kann die Reihenfolge des Einschenkens erkennen) **oder ein seltenes Ereignis ist eingetreten.**

- Voraussetzung: **echte Zufallsstichprobe**
- Ergebnis: **kein Beweis**
- keine Auskunft darüber, wie „wahrscheinlich“ der Effekt ist
- keine Auskunft darüber, wie stark wir von der Existenz des Effekts überzeugt sein sollten (= *inverse probability*)
- jede Ho-Zurückweisung: nur ein kleinteiliger Hinweis auf einen Effekt
- **substantielle** theoretische Hypothese oft und hart testen!

- **Sensitivity**: keine Auskunft über die **Stärke** des Effekts
 - ▶ große Stichprobe → hohe Sensitivität
 - ▶ kleine Stichprobe → niedrige Sensitivität
 - ▶ je sensibler desto leichter werden schwache Effekte signifikant
 - ▶ Abhilfe bei Neyman-Pearson: **Power** (Teststärke) vor dem Experiment
 - ▶ quasi-kompatibel zu Fisher: **Effektstärke** nach dem Experiment

Und beim Konfidenzintervall?

Am Beispiel des 95%-Konfidenzintervalls (KI)

- **Falsch:** Wir können zu 95% sicher sein, dass der wahre Wert im KI liegt.
- **Falsch:** Der wahre Wert liegt mit 95% Wahrscheinlichkeit im KI.
- Warum? Wenn der wahre Wert nicht im geschätzten KI liegt, ist die Wahrscheinlichkeit 1, dass er nicht im KI liegt.
- Fakten haben die Wahrscheinlichkeit 1.
- Richtig: Entweder liegt der wahre Wert im KI oder ein seltenes Ereignis ist eingetreten
- „selten“ heißt: nur in 5 von 100 Fällen (im Grenzwert)

- **exakter** Test:

- ▶ Die Wahrscheinlichkeitsverteilung ist bekannt und wird direkt zugrunde gelegt (= Berechnung der exakten Wahrscheinlichkeit).
- ▶ Fisher-Test, Binomialtest
- ▶ hohe Sensitivität
- ▶ geeignet für kleine Stichproben
- ▶ oft rechenintensiv

- **approximativer** oder **asymptotischer** Test:

- ▶ Die Wahrscheinlichkeitsverteilung ist nicht bekannt (oder kann mathematisch nicht effizient zugrundegelegt werden) und es wird ein Differenzwert berechnet, der asymptotisch eine bekannte Verteilung hat.
- ▶ χ^2 -Test, t-Test, ANOVA
- ▶ oft wird Normalverteilung approximiert
- ▶ wegen asymptotischer Natur weniger sensitiv (= größere Stichprobe)

- parametrischer Test:

- ▶ Messung eines Parameters/mehrerer Parameter der Grundgesamtheit
- ▶ (Parameter entsprechen in der Messung einer Variable)
- ▶ zum Beispiel Mittelwert oder Varianz
- ▶ Voraussetzung: **bekannte Wahrscheinlichkeitsverteilung der Variable**
- ▶ z. B. t-Test (mittel), ANOVA (Varianz)

- nichtparametrischer Test:

- ▶ keine direkte Messung eines zufallsverteilten Parameters
- ▶ zum Beispiel Ränge oder Zähldaten
- ▶ keine Verteilungsannahmen (auch: *verteilungsfreier Test*)
- ▶ z. B. χ^2 , Binomialtest, H-Test, U-Test

t-Test

- Mittel μ über X in der Grundgesamtheit bekannt (z. B. mittlere Satzlänge im Korpus).
- Stichprobe (z. B. der Grundriss von PE) zeigt gemessenes Mittel \bar{x} .
- Ist die Abweichung signifikant?
- $H_0: \bar{x} = \mu$

Wäre die **Varianz der GG** als $s^2(X)$ bekannt:

- $SF(X)$ bei Stichprobengröße n ausrechnen, und...
- mit $z = \frac{\bar{x} - \mu}{SF(X)}$ einen Signifikanztest über Normalverteilung rechnen
- Problem aber leider: $SF(X) = \frac{s(X)}{\sqrt{n}}$
- und $s^2(X)$ meist nicht bekannt!

Aufgabe: Mit Ihrer Stichprobe aus NaB und $\mu = 6.8$ sowie $s^2(X) = 10.8$ z-Test rechnen. (Bzw. erstmal die nötigen Werte ausrechnen. Wir besprechen dann die Interpretation als Test.)

Annahme beim t-Test mit einer Stichprobe

- Wir kennen μ oder haben eine Hypothese (z. B. $\mu = 0.5$).
- Wir haben eine Stichprobe x mit n und bekannten \bar{x} und $s^2(x)$.
- anders als bei z-Test: **Wir schätzen $SF(X) \approx SF(x)$!**

$$t = \frac{\bar{x} - \mu}{SF(\bar{x})}$$

Bitte rechnen für Satzlängen (in Wörtern):

$$\mu = 7.3$$

$$x = [6, 3, 12, 16, 8, 15, 9, 9, 2, 11]$$

1 $\bar{x} = 9.1$

2 $s^2(x) = 21.43$

3 $s(x) = 4.63$

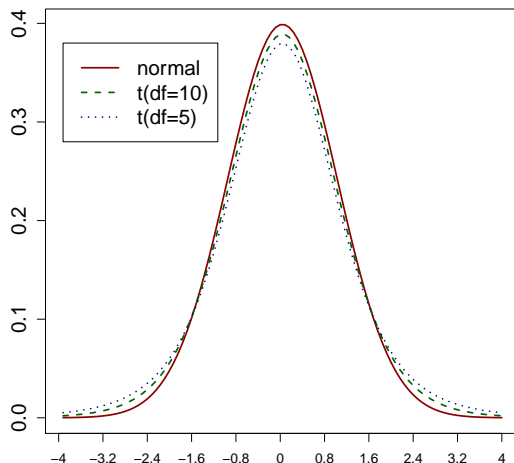
4 $SF(x) = \frac{4.63}{\sqrt{10}} = 1.464$

5 $t = \frac{9.1-7.3}{1.464} = 1.229$

Und was sagt uns $t = 1.229$?

t-Verteilung

Während die z-Werte normalverteilt sind, flacht die Verteilung der t-Werte durch die Schätzung je nach df verglichen mit der Normalverteilung ab.



- $df = n - 1$ (\bar{x} muss für $s^s(x)$ bekannt sein)
- Welche t-Werte machen $1 - \alpha$ der Werte aus?
- `> qt(c(0+0.05/2, 1-0.05/2), df=9)`
 $\Rightarrow 2.262157.. - 2.262157$
- Der errechnete t-Wert ist nicht signifikant.
- $H_0: \mu = \bar{x}$ nicht zurückgewiesen.

- Signifikanz \neq starker Effekt
- Effektstärke beim t-Test für Stichprobe x :

$$\text{Cohens } d = \frac{\bar{x} - \mu}{s(x)}$$

- Herleitung/Erklärung: Gravetter & Wallnau, Kap. 9

- ähnlich der Effektstärke:
Welcher Anteil der Varianz in den Daten
wird durch die Unabhängige erklärt?

$$\text{Cohens } r^2 = \frac{t^2}{t^2 + df}$$

- Herleitung/Erklärung: Gravetter & Wallnau, Kap. 9

- zwei Grundgesamtheiten (z. B. dt. Sätze im 19. und im 20. Jh.)
- dazu: zwei Stichproben (je eine) mit einem Mittelwert (z. B. Länge)
- Interesse: anhand der zwei Stichproben zeigen, dass sie (sehr wahrscheinlich) aus zwei Grundgesamtheiten kommen
- $H_0: \mu_1 - \mu_0 = 0$
- hier also: eine unabhängige Variable (Jahrhundert) und eine abhängige Variable (Satzlänge), gemessen als Mittel

Allgemein funktioniert der t-Test **immer** so:

$$t = \frac{\text{Stichprobenwert} - \text{Grundgesamtheitswert}}{\text{Standardfehler}}$$

Jetzt geht man per Hypothese von zwei GG und zwei Stichproben aus, also:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SF(x_1 - x_2)}$$

- Wir testen also auf die **Differenz der Unterschiede**.
- Per H_0 wird gesetzt: $\mu_1 - \mu_2 = 0$

Für gleichgroße Stichproben:

$$SF(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s^2(x_1)}{n_1} + \frac{s^2(x_2)}{n_2}}$$

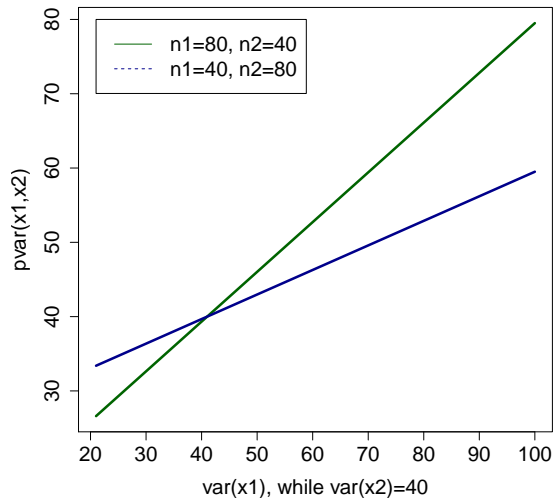
- Problem: Beitrag zum SF von beiden Stichproben gleich.
- Besser: **zusammengefasste Varianz**, und daraus dann SF.

$$s_p^2(x_1, x_2) = \frac{(\sum_{i=1}^n (x_{1,i} - \bar{x}_1)^2) + (\sum_{i=1}^n (x_{2,i} - \bar{x}_2)^2)}{(n_1 - 1) + (n_2 - 1)} = \frac{SQ(x_1) + SQ(x_2)}{(n_1 - 1) + (n_2 - 1)}$$

$$SF(x_1 - x_2) = \sqrt{\frac{s_p^2(x_1, x_2)}{n_1} + \frac{s_p^2(x_1, x_2)}{n_2}}$$

Mehr: Gravetter & Wallnau, Kap. 10

Illustration der zusammengefassten Varianz



t-Wert

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{SF(x_1 - x_2)} = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{SF(x_1 - x_2)} = \frac{\bar{x}_1 - \bar{x}_2}{SF(x_1 - x_2)}$$

Freiheitsgrade

$$df = df(x_1) + df(x_2) = (n_1 - 1) + (n_2 - 1)$$

Effektstärke

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_p^2}}$$

Erklärung der Varianz

$$r^2 = \frac{t^2}{t^2 + df}$$

Bitte „von Hand in R“ t-Test für folgende zwei Stichproben
bei $\alpha = 0.05$ rechnen:

$$\begin{aligned}x_1 &= [11, 11, 8, 8, 11, 9, 8, 11, 9, 8] \\x_2 &= [10, 14, 14, 13, 11, 14, 10, 14, 12, 10]\end{aligned}$$

Und überprüfen mit:
`> t.test(x1, x2)`

Die GGs müssen normalverteilt sein:

```
shapiro.test(x)
```

Wenn $p \leq 0.05$ wird die Nullhypothese des Shapiro-Wilk-Tests verworfen –
Ho: Die Werte stammen aus einer normalverteilten GG.

Die Varianzen müssen homogen sein:

```
var.test(x1, x2)
```

Auch hier: $p \leq 0.05$ weist die Ho zurück (sehr informell) –
Ho: Die Varianzen von x1 und x2 sind homogen.

Solche Tests sind umstritten, weil sie i. d. R. viel zu empfindlich reagieren.
Zuur u. a. (2009) empfehlen z. B. grafische Methoden (bei linearen Modellen).

Wenn Voraussetzungen nicht erfüllt sind:

- steigt das Risiko für Typ 1-Fehler
- nicht-parametrische Alternative nehmen
- Daten transformieren
- sich über Robustheit des Test ggü. verletzten Annahmen informieren (oft schwer zugängliche und kontroverse Spezialliteratur)

Nächste Woche | Überblick

- 1 Statistik, Inferenz und probabilistische Grammatik
- 2 Deskriptive Statistik
- 3 Nichtparametrische Verfahren
- 4 z-Test und t-Test
- 5 ANOVA
- 6 Freiheitsgrade und Effektstärken
- 7 Power
- 8 Lineare Modelle
- 9 Generalisierte Lineare Modelle
- 10 Gemischte Modelle

- Bortz, Jürgen & Christof Schuster. 2010. *Statistik für Human- und Sozialwissenschaftler*. 7. Aufl. Berlin: Springer.
- Fisher, Ronald A. 1935. *The Design of Experiments*. London: Macmillan.
- Gravetter, Frederick J. & Larry B. Wallnau. 2007. *Statistics for the Behavioral Sciences*. 7. Aufl. Belmont: Thomson.
- Zuur, Alain F., Elena N. Ieno, Neil Walker, Anatoly A. Saveliev & Graham M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. Berlin etc.: Springer.

Kontakt

Prof. Dr. Roland Schäfer
Institut für Germanistische Sprachwissenschaft
Friedrich-Schiller-Universität Jena
Fürstengraben 30
07743 Jena

<https://rolandschaefer.net>
roland.schaefer@uni-jena.de

Creative Commons BY-SA-3.0-DE

Dieses Werk ist unter einer Creative Commons Lizenz vom Typ *Namensnennung - Weitergabe unter gleichen Bedingungen 3.0 Deutschland* zugänglich. Um eine Kopie dieser Lizenz einzusehen, konsultieren Sie

<http://creativecommons.org/licenses/by-sa/3.0/de/> oder wenden Sie sich brieflich an Creative Commons, Postfach 1866, Mountain View, California, 94042, USA.