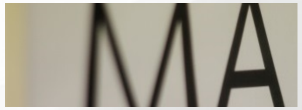
INSTITUT FÜR DEUTSCHE SPRACHE

Sandra Hansen-Morath
Sascha Wolfer



STATISTIK MIT R

t-Test

Mitglied der Leibniz-Gemeinschaft

WIEDERHOLUNG

- Mit dem **Chi-Quadrat-Test** können wir testen, ob eine beobachtete Häufigkeitsverteilung von einer erwarteten Verteilung abweicht.
- Der Chi-Quadrat-Test ist somit geeignet für **nominalskalierte** Variablen.
- Die erwarteten Häufigkeiten in einer Kreuz- oder Kontingenztabelle ergeben sich dabei aus den **Randsummen** der Tabelle.
- Ergebnisse des Chi-Quadrat-Tests sollten über die **Effektstärke** (Phi / Cramérs V) evaluiert werden.

EINORDNUNG T-TEST

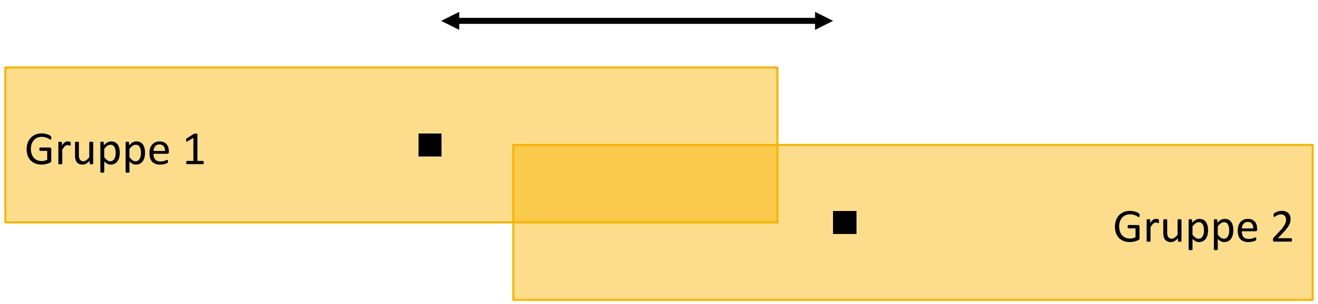
- Der Chi-Quadrat-Test kann als **Anpassungstest** verstanden werden, weil eine bestimmte Verteilung gegen eine andere getestet wird.
- Mit dem t-Test testen wir, ob zwischen zwei Gruppen ein signifikanter **Unterschied** besteht.
  - Die "Gruppen" können dabei auch zwei Messzeitpunkte derselben Individuen sein (Vorher-Nachher-Messung).
  - Es werden die Mittelwerte und deren Konfidenzintervalle verglichen.
- Die abhängige Variable muss beim t-Test **intervallskaliert** und normalverteilt sein.
- Die Gruppeneinteilung kann nur **binär** sein.

BEISPIELE

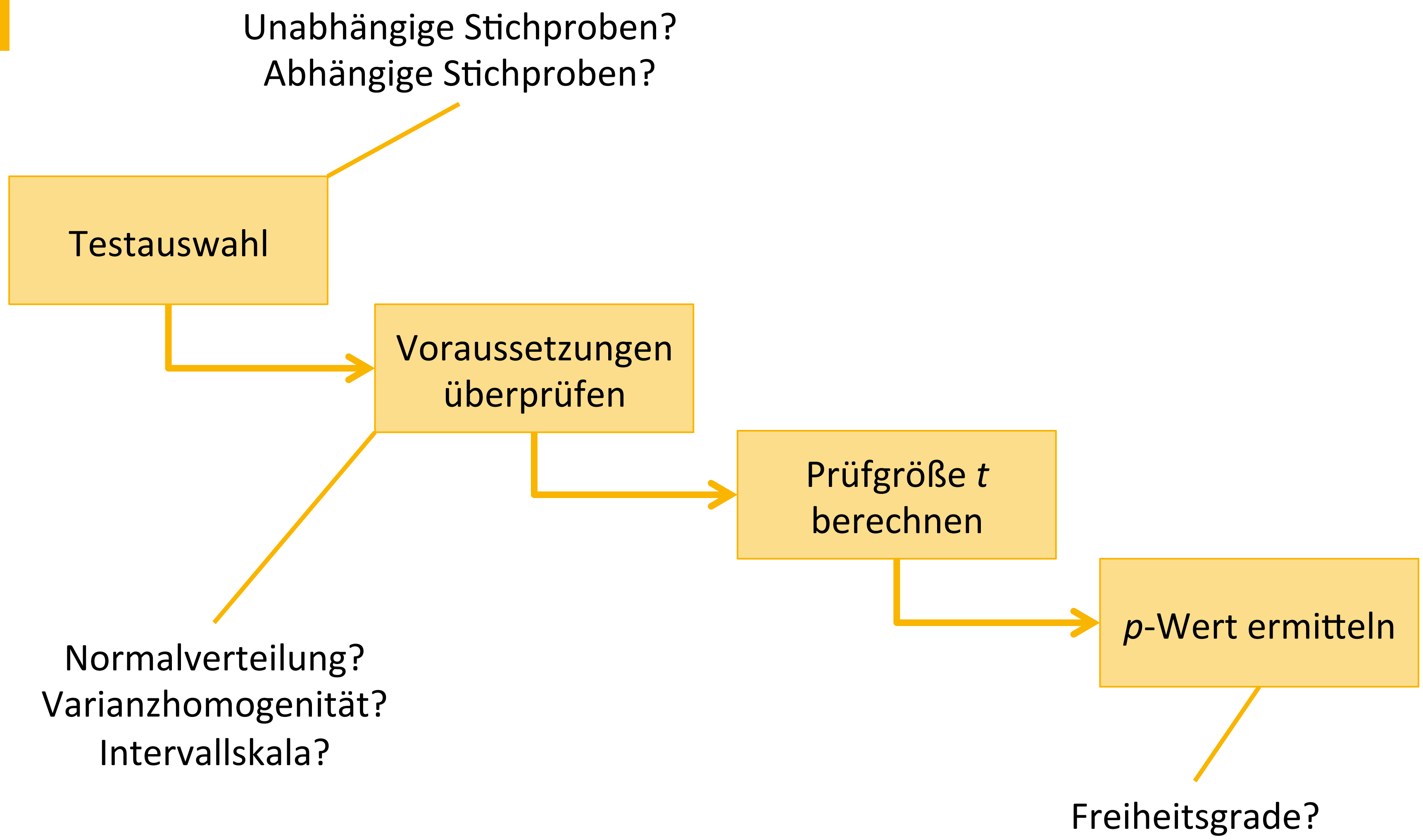
- Sind Print-Zeitungstexte im Schnitt länger als Online-Zeitungstexte?
- Können VersuchsteilnehmerInnen besser mit Wörterbüchern umgehen, nachdem sie geschult wurden? (Im Vergleich zu einer Kontrollgruppe / im Vergleich zum ungeschulten Zustand).
- Unterscheiden sich die Reaktionszeiten von VersuchsteilnehmerInnen bezüglich der Korrektheit der Antwort? (Wird schneller geantwortet, wenn korrekt geantwortet wird?)
- Unterscheiden sich die Lesezeiten von Nomen und Verben? In welche Richtung geht der Unterschied?

DER T-TEST

- Inferenzstatistischer Test
- Idee:
  - Es existieren zwei Stichproben bzw. Gruppen, von denen uns die Mittelwerte bekannt sind.
  - Liegen die Mittelwerte weit genug auseinander und sind die assoziierten Konfidenzintervalle klein genug, um einen Unterschied annehmen zu können?



VORGEHEN


Unabhängige Stichproben?
Abhängige Stichproben?
Testauswahl
Voraussetzungen überprüfen
Prüfgröße *t* berechnen
p-Wert ermitteln
Normalverteilung?
Varianzhomogenität?
Intervallskala?
Freiheitsgrade?

UNABHÄNGIGE UND ABHÄNGIGE STICHPROBEN

- t-Test für unabhängige Stichproben
  - Stichproben sind nicht miteinander verbunden bzw. „gepaart“
  - z.B. unterschiedliche Versuchsteilnehmer, Nominalphrasenlänge in zwei Korpora
  - in R Argument: `paired = F` (default)

- t-Test für abhängige Stichproben
  - Stichproben sind miteinander verbunden bzw. „gepaart“
  - z.B. Vorher-Nachher-Messungen, Trainingsstudien
  - höhere Teststärke, wird eher signifikant
  - in R Argument: `paired = T`

VORAUSSETZUNGEN

- Da Mittelwerte verglichen werden, muss die abhängige Variable logischerweise **intervallskaliert** sein.
- Die abhängige Variable (AV) muss **normalverteilt** sein.
  - Überprüfen über Plotten (`hist, truehist, density`) oder einen Test auf Normalverteilung (Package `nortest`).
  - Wenn die AV nicht normalverteilt ist, kann sie **transformiert** werden (log, sqrt, 1/x ...).
  - Wenn die AV durch keine Transformation in eine Normalverteilung überführt werden kann, kann ein nicht-parametrischer Test verwendet werden (bspw. der **Mann-Whitney U-Test**).
- Zwischen den beiden Gruppen / Stichproben muss **Varianzhomogenität** herrschen.
  - Überprüfen über `leveneTest` oder `var.test`.
  - Kann korrigiert werden (Parameter `var.equal = F`, default!)

BEISPIEL:
UNTERSCHEIDET SICH DIE GEDÄCHTNISLEISTUNG IN EINEM GEDÄCHTNISTEST ZWISCHEN SCHULKLASSE A UND SCHULKLASSE B?

	Gedächtnistests											Mittelwert	Standardabweichung
Schulklasse A	31	29	29	24	32	19	26	19	21	19	28	25.18	5.016
Schulklasse B	20	26	19	21	24	22	23	23	22	19	18	21.55	2.423

- **in R:** `t.test(<vektor1>, <vektor2>)`
  - vektor1: c(31, 29, 29, 24, …)
  - vektor2: c(20, 26, 19, 21,…)

- Testergebnis:
  - Prüfgröße t bei *df* Freiheitsgraden
  - Irrtumswahrscheinlichkeit *p*
  - Konfidenzintervall des Unterschieds
  - Mittelwerte der Stichproben

<u>Beispiel R-Output:</u>

data: x and y
t = 2.1649, df = 14.427, p-value = 0.04762
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
0.04367174 7.22905554
sample estimates:
mean of x mean of y
25.18182 21.54545

Beispiel: http://www.methodenberatung.uzh.ch/datenanalyse/unterschiede/zentral/ttestunabh.html

T-TEST FÜR UNABHÄNGIGE STICHPROBEN

- H0: Zwei Mittelwerte zweier unabhängiger Stichproben stammen aus einer Grundgesamtheit.
- Wenn die H0 abgelehnt wird, nehmen wir an, dass die beiden Stichproben aus unterschiedlichen Populationen stammen.

→ **„Die Stichproben unterscheiden sich signifikant voneinander."**

Mittelwertunterschied

$$\boxed{t} = \frac{\boxed{\bar{x}_1 - \bar{x}_2}}{\boxed{s}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

Prüfgröße

Rauschen: kombinierte Standardabweichung

df = n1 + n2 - 2

KOMBINIERTE STANDARDABWEICHUNG

- Was genau ist die „kombinierte“ Standardabweichung?
- Da die Streuung für die beiden Gruppen unterschiedlich sein kann, muss eine kombinierte Standardabweichung berechnet werden.
- Dafür werden die Varianzen der beiden Gruppen an der jeweiligen Gruppengröße gewichtet.
  - Somit erhält die Varianz der größeren Gruppe mehr "Gewicht" in der Berechnung der kombinierten Standardabweichung.

An Stichprobengröße gewichtete Varianzen der Gruppen

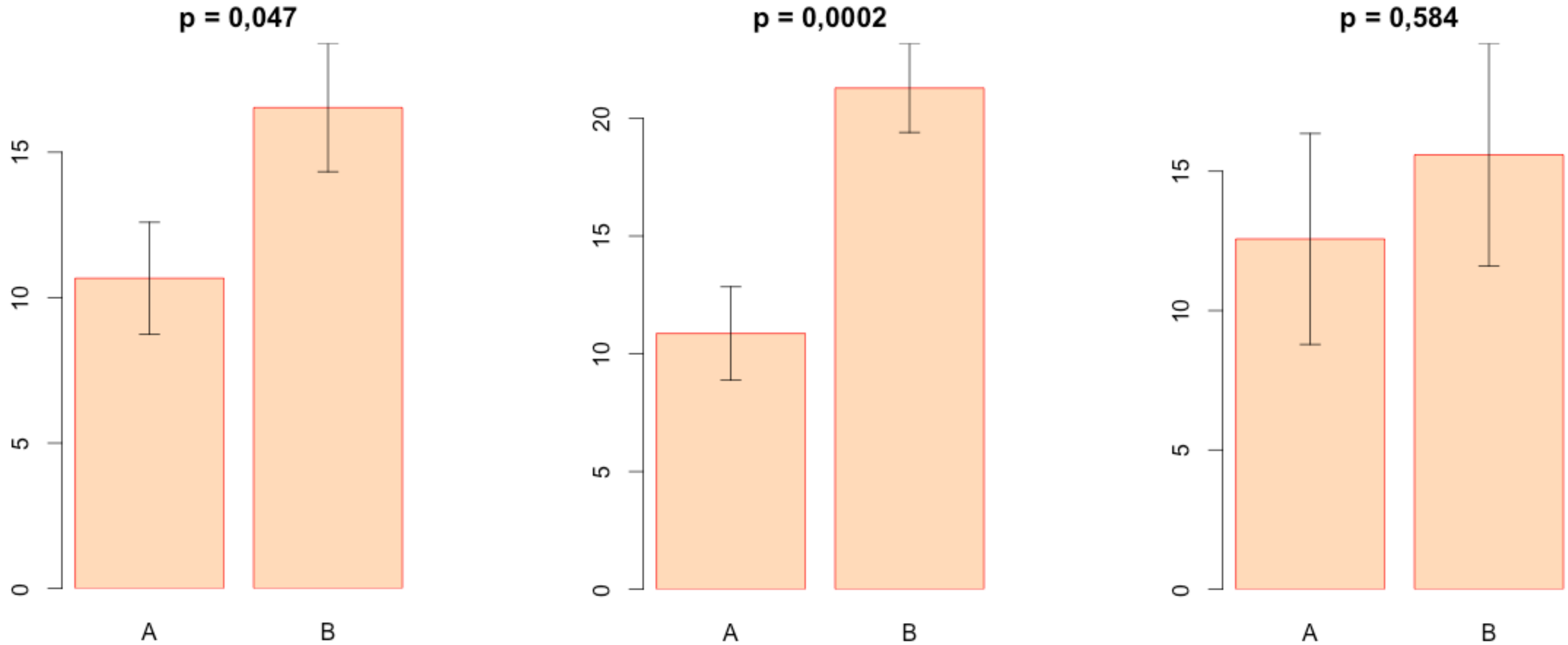
$$s^2 = \frac{\boxed{(n_1 - 1)s_1^{\,2}} + \boxed{(n_2 - 1)s_2^{\,2}}}{n_1 + n_2 - 2}$$

GERICHTETE HYPOTHESEN

- Hat man eine Hypothese über die **Richtung** des Unterschieds, kann diese im Aufruf der Funktion `t.test` angegeben werden.
  - Parameter **`alternative`** kann die Werte `"two.sided"` (default), `"less"` und `"greater"` annehmen.
  - `t.test(x, y, alternative = "less")` bedeutet: "Ich habe die Hypothese, dass der Mittelwert von *x* kleiner als jener von *y* ist."
  - Es wird dann nur auf einer Seite der Verteilung nach einem Unterschied gesucht.
- **Einseitige**, d.h. gerichtete t-Tests sind teststärker und werden somit eher signifikant als **zweiseitige** t-Tests.
  - $p_{\text{einseitig}} = p_{\text{zweiseitig}} / 2$

T-TESTS UND KONFIDENZINTERVALLE

- Ein t-Test zeigt praktisch immer dann einen signifikanten Unterschied an, wenn die **Standardfehler** zweier Stichproben sich nicht überschneiden.



MULTIPLE VERGLEICHE

- Mit dem t-Test können **zwei** Gruppen miteinander verglichen werden.
- Will man mehr als zwei Gruppen untereinander vergleichen (und trotzdem einen t-Test verwenden), kann man **multiple Vergleiche** durchführen.
  - 1 vs. 2; 2 vs. 3; 3 vs. 1, ...
- Das führt zu einer **Inflation des Alpha-Fehlers**.
- Daher muss eine Korrektur vorgenommen werden!
  - bspw. Korrektur nach Holm
- In R: pairwise.t.test(..., p.adjust.method = "holm")

ABHÄNGIGE STICHPROBEN

- Abhängige oder gepaarte Stichproben: Mehrere Messzeitpunkte von einem Individuum (i.w.S.).
  - klassisch: Vorher-Nachher-Studien, Trainingsstudien
- Grundidee: Statt mit zwei Stichproben / Gruppen wird mit **einer** Stichprobe gerechnet: den **Differenzen** zwischen Messzeitpunkten.
- H0: Differenzen stammen aus einer Grundgesamtheit mit dem Mittelwert 0. ("Es gibt keine Unterschiede zwischen Messzeitpunkt 1 und 2.")
- Varianz, die an das Individuum gebunden ist, wird systematisch vernachlässigt.
  - z.B. körperliche Fitness, Intelligenz, Körpergröße, ...

ABHÄNGIGE STICHPROBEN

- Zunächst: Berechnung des Differenzwerts *d* für jedes Individuum: $d_i = x_{i1} – x_{i2}$ wobei i1 und i2 die beiden Messzeitpunkte sind.
- Berechnung des t-Werts: Verhältnis des Mittelwerts und des Standardfehlers der Differenzen.
  - Freiheitsgrade: $df = n_d - 1$

$$t_{n-1} = \frac{\bar{x}_d}{\hat{\sigma}_{\bar{x}_d}} \Rightarrow t_{n-1} = \frac{\bar{x}_d}{s_d} \cdot \sqrt{n}$$

VERGLEICH DER FORMELN

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

$$t_{n-1} = \frac{\bar{x}_d}{s_d} \cdot \sqrt{n}$$

NICHT-PARAMETRISCHE VERSIONEN

- Wenn die abhängige Variable **nicht normalverteilt** und nicht transformierbar ist, können nicht-parametrische Versionen des t-Tests angewendet werden.
  - Keine Verteilungsvoraussetzung!
- Unabhängige Stichproben: **Mann-Whitney-U-Test**
- Abhängige Stichproben: **Wilcoxon-Test**
- In R: `wilcox.test()`
- Es wird mit den **Rangplätzen** der Messwerte gerechnet.
  - Im Allgemeinen sind nicht-parametrische Tests **konservativer**, d.h. Effekte müssen im Vergleich zu parametrischen Tests größer sein, um signifikant zu werden.

BEGRIFFE

Unabhängige Stichprobe
Abhängige (gepaarte) Stichprobe
Intervallskala
Varianzhomogenität
Normalverteilung
t-Test
Prüfgröße
Freiheitsgrade

kombinierte Standardabweichung
multiple Vergleiche
Signifikanzniveau-Korrektur
einseitiger Test
zweiseitiger Test
Standardfehler
Konfidenzintervall
Mann-Whitney-U-Test
Wilcoxon-Test