Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

DMATH
Department of Mathematics

Hypothesentests für Erwartungswert und Median

für D-UWIS, D-ERDW, D-USYS und D-HEST – SS15

Normalverteilung

- $X \sim N(\mu, \sigma^2)$:
  «*X ist normalverteilt mit Erwartungswert μ und Varianz σ^2*»
- pdf:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- cdf: ziemlich umständlich
- Zentraler Grenzwertsatz (CLT):
  - $X_i \sim F \;\; i.i.d.$ mit $E(X_i) = \mu$ und $Var(X_i) = \sigma^2$, dann gilt…

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right), \text{ falls } n \to \infty$$

$$S_n = \sum_{i=1}^{n} X_i \sim \mathcal{N}(n\mu, n\sigma^2), \text{ falls } n \to \infty$$

CLT: Normalapproximation des Binomialtests

1. **Modell**: n Lose kaufen, gleiche Gewinnchance, unabh.

   jedes Los X_i: 1 mit W'keit π, 0 mit W'keit $1 - \pi$

   $$E(X_i) = \boxed{\pi,} Var(X_i) = \boxed{\pi(1-\pi)}$$

   X: Anzahl Gewinne, $X = X_1 + X_2 + \cdots + X_n$

2. $\mathcal{H}_0: \pi = \pi_0$; $\mathcal{H}_A: \pi < \pi_0$

3. **Teststatistik** T: CLT → $T \sim \mathcal{N}(n\pi_0, n\pi_0(1-\pi_0))$

4. **Signifikanzniveau**: $\alpha = 0.05$

CLT: Normalapproximation des Binomialtests

5. **Verwerfungsbereich**: $K = [0, c]$

   Finde c, sodass $P[T \leq c] = 0.05$ (mit Computer oder…

   Standardisiere & verwende Tabelle:

   $P[T \leq c] = P[Z \leq \tilde{c}] = 0.05$ mit $\tilde{c} = \frac{c - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}}$

   aus Tabelle: $\tilde{c} = -1.64$

   nach c auflösen: $c = n\pi_0 - 1.64\sqrt{n\pi_0(1-\pi_0)}$

6. **Testentscheid**

Lernziele heute

- z-Test
- t-Test
- Vorzeichentest
- Wilcoxon-Test

Hausaufgaben

- Skript: Kapitel 4.7 lessen
- Serie 9 lösen
- Quiz 9 bearbeiten
- etutoR 7

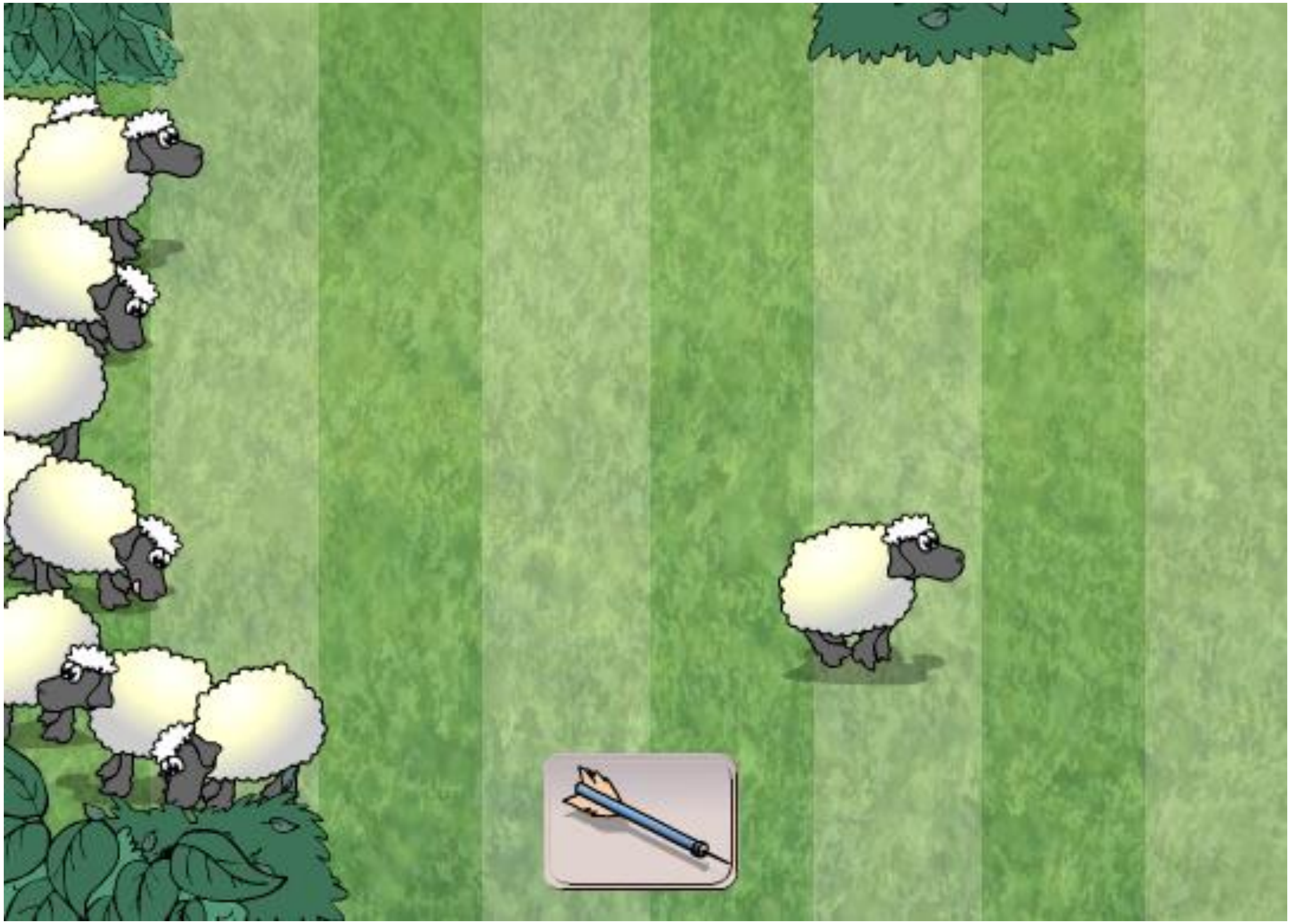
Reaktionszeit

Reagiert man mit der Haupthand schneller, wie mit der Nebenhand?

- Experiment:
  - **Population**: Alle StudentInnen der Vorlesung
  - **Stichprobe**: 70 zufällig ausgewählte StudentInnen
- Messmethode:
  - Reaktionszeittest auf dem Internet
- Testlauf mit beiden Händen (Reihenfolge randomisiert)
- Messung mit beiden Händen (5 Messungen)
- Robustheit:
  - jeweils bestes und schlechtestes Resultat streichen, Rest mitteln
- Differenz aus HH und NH berechnen
- Anreiz:
  - Verlosung eines Kinogutscheins

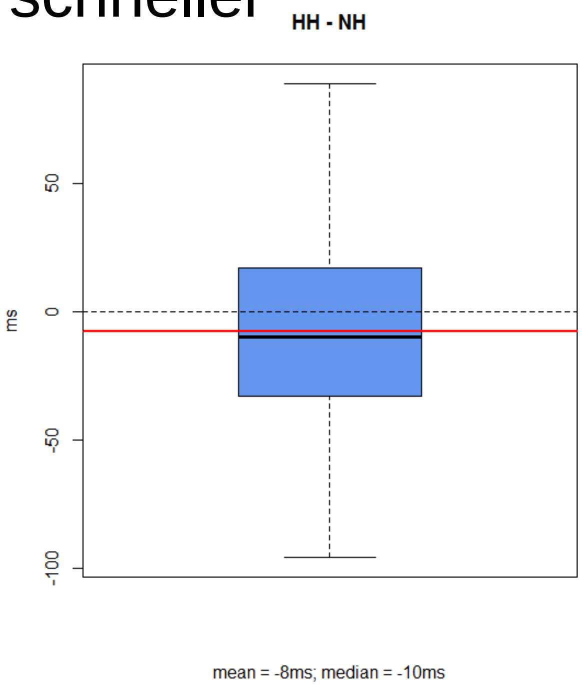
http://www.bbc.co.uk/science/humanbody/sleep/sheep/

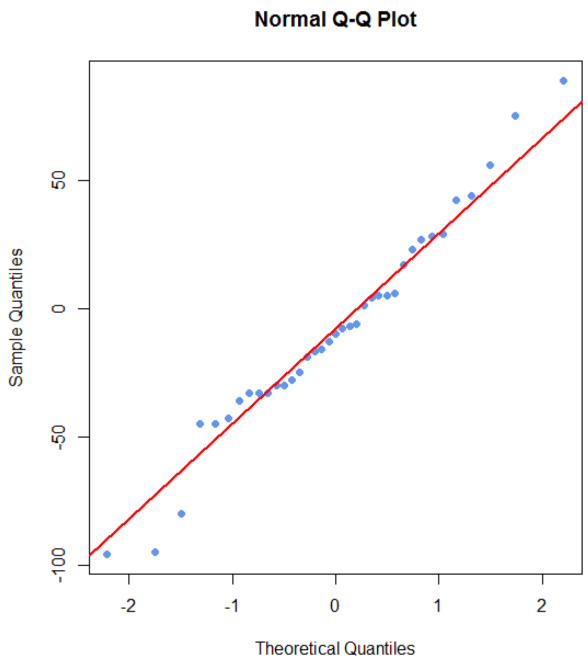
Daten sammeln mit Schafen…



Ergebnis

- 70 StudentInnen angeschrieben
- Rücklauf: 37
- Haupthand ist im Mittel 8 ms <u>schneller</u>, der Median liegt bei 10 ms schneller





Stichprobe versus Population

- In der Stichprobe war die Haupthand 8 ms schneller
- Können wir daraus schliessen, dass die Haupthand in der ganzen Population im Mittel schneller ist?
- Eine Antwort liefern:
  - **z-Test**
  - **t-Test**
  - Wilcoxon-Test (Mann-Whitney-U-Test)
  - Vorzeichen-Test

AND THE WINNER IS...

z-Test (σ_X bekannt)

1. **Modell**: X_i kontinuierliche Messgrösse;
   $X_1, X_2, \ldots, X_n$ $i.i.d.$, $\boldsymbol{\mathcal{N}(\mu, \sigma_X^2)}$, $\boldsymbol{\sigma_X}$ **bekannt**
2. **Nullhypothese**: $\mathcal{H}_0: \mu = \mu_0$
   Alternative: $\mathcal{H}_A: \mu \neq \mu_0$ (oder < oder >)
3. **Teststatistik**:
   $$Z = \frac{(\bar{X}_n - \mu_0)}{\sigma_{\bar{X}_n}} = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma_X} = \frac{\text{beobachtet } - \text{erwartet}}{\text{Standardfehler}}$$
   Verteilung unter $\mathcal{H}_0$: $Z \sim \mathcal{N}(0,1)$
4. **Signifikanzniveau**: α
5. **Verwerfungsbereich** für die Teststatistik:
   $$K = (-\infty, -\Phi^{-1}(1-\alpha/2)] \cup [\Phi^{-1}(1-\alpha/2), \infty)$$
   $$K = (-\infty, -\Phi^{-1}(1-\alpha)] \text{ bei } \mathcal{H}_A: \mu < \mu_0$$
   $$K = [\Phi^{-1}(1-\alpha), \infty) \text{ bei } \mathcal{H}_A: \mu > \mu_0$$
6. **Testentscheid**: Liegt beobachteter Wert z der Teststatistik in K

Problem in der Praxis: σ_X ist nicht bekannt!

- Schätze die Varianz:

$$\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)$$

- Neue Teststatistik:

$$T = \frac{\bar{X}_n - \mu_0}{\frac{\hat{\sigma}_X}{\sqrt{n}}}$$

- Verteilung von T, falls $\mathcal{H}_0$ stimmt:

$$T \sim t_{n-1}$$

William
Sealy
Gosset

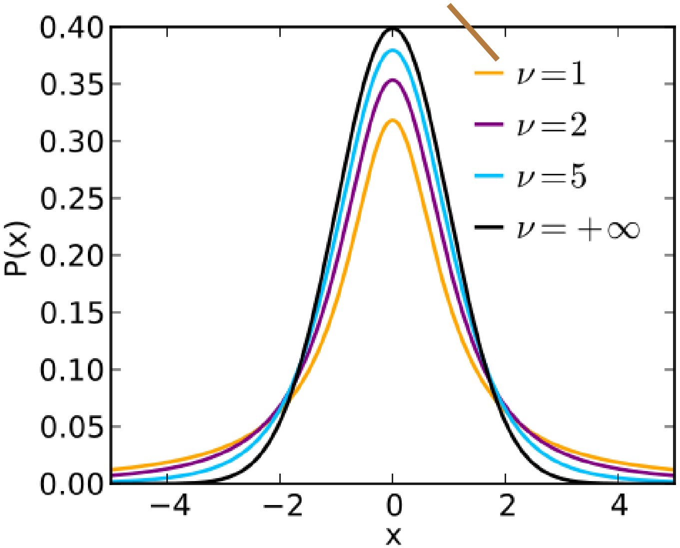
«Student's» t-Verteilung – kleiner Abstecher im Verteilungszoo!

- Annahme:
  - $X_1, X_2, \dots, X_n \sim \mathcal{N}(\mu, \sigma_X^2)$ und unabhängig
- $\hat{\sigma}_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$ ist die geschätzte Varianz
- Die Teststatistik

$$T = \frac{(\bar{X}_n - \mu)}{\left(\frac{\hat{\sigma}_X}{\sqrt{n}}\right)} \sim t_n$$

  folgt einer
  «t-Verteilung mit n Freiheitsgraden»
- Falls $n = \infty$, dann ist $t_\infty = \mathcal{N}(0,1)$

Umso weniger df,
umso meht Streuung



t-Test (σ_X unbekannt)

1. **Modell**: X_i kontinuierliche Messgrösse;
   $X_1, X_2, \ldots, X_n$ $i.i.d., \boldsymbol{\mathcal{N}(\mu, \sigma_X^2)}, \sigma_X$ wird mit $\hat{\sigma}_X$ geschätzt
2. **Nullhypothese**: $\mathcal{H}_0: \mu = \mu_0$
   Alternative: $\mathcal{H}_A: \mu \neq \mu_0$ (oder < oder >)
3. **Teststatistik**:
   $$\mathrm{T} = \frac{(\bar{X}_n - \mu_0)}{\hat{\sigma}_{\bar{X}_n}} = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}_X} = \frac{\text{beobachtet} - \text{erwartet}}{\text{geschätzter Standardfehler}}$$
   Verteilung unter $\mathcal{H}_0$: $\mathrm{T} \sim t_{n-1}$
4. **Signifikanzniveau**: α
5. **Verwerfungsbereich** für die Teststatistik:
   $$K = (-\infty, -t_{n-1;1-\frac{\alpha}{2}}] \cup [t_{n-1;1-\frac{\alpha}{2}}, \infty)$$
   $$K = (-\infty, -t_{n-1;1-\alpha}] \text{ bei } \mathcal{H}_A: \mu < \mu_0$$
   $$K = [t_{n-1;1-\alpha}, \infty) \text{ bei } \mathcal{H}_A: \mu > \mu_0$$
6. **Testentscheid**:
   Liegt beobachteter Wert t der Teststatistik in K

t.test
power.t.test

Beispiel t-Test

1. Modell: X_i Differenz in der Reaktionszeit von HH und NH von StudentIn i
2. Nullhypothese: $\mathcal{H}_0: \mu = 0\ ms$
   Alternative: $\mathcal{H}_A: \mu \neq 0\ ms$
3. Teststatistik:
$$T = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\hat{\sigma}_X} \Rightarrow t = \frac{\sqrt{37}(-8.03 - 0)}{41.13} = -1.19$$
4. Signifikanzniveau: $\alpha = 0.05$
5. Verwerfungsbereich:
$$K = \left(-\infty, -t_{36;0.975}\right] \cup \left[t_{36;0.975}, \infty\right) = (-\infty, -2.03] \cup [2.03, \infty)$$
6. Testentscheid: $t \notin K \Rightarrow \mathcal{H}_0$ kann nicht verworfen werden

P-Wert

- «Kleinstes Signifikanzniveau, bei dem $\mathcal{H}_0$ gerade noch verworfen wird.»
  - z.B. P-Wert = $0.03 \rightarrow \alpha = 0.05$ ✔ $\alpha = 0.01$ ✘
- $\mathcal{H}_A: \mu \neq \mu_0$ und der beobachtete Wert $t = \frac{\sqrt{n}|\bar{X}_n - \mu_0|}{\hat{\sigma}_X}$
- P-Wert berechnet sich…
  - $P[|T| > |t|] = P[T < -|t|] + P[T > |t|] = 2 \cdot P[T > |t|] =$
  - $= 2 \cdot (1 - P[T \leq |t|]) =$
  - $= 2 \cdot \left(1 - F_{t_{n-1}}(|t|)\right) = 2 \cdot \left(1 - F_{t_{n-1}}\left(\frac{\sqrt{n}|\bar{X}_n - \mu_0|}{\hat{\sigma}_X}\right)\right)$

wobei $F_{t_{n-1}}$ die kumulative Verteilungsfunktion der t-Verteilung mit $n - 1$ Freiheitsgraden

$(1-\alpha)$-Vertrauensintervall für μ

- Äquivalente Definitionen:
  - Enthält wahren Wert μ mit Wahrscheinlichkeit $1-\alpha$
  - Enthält alle Werte μ_0, bei denen $\mathcal{H}_0: \mu = \mu_0$ vs $\mathcal{H}_A: \mu \neq \mu_0$ mit Signifikanzniveau α **nicht** verworfen wird
- im t-Test Schritt 5: Nicht verwerfen, falls…

$$\left|\frac{\bar{X}_n - \mu}{\frac{\hat{\sigma}_X}{\sqrt{n}}}\right| < t_{n-1;1-\alpha/2}$$

  … und das nach μ auflösen.
- CI: $\left[\bar{x}_n - t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{\widehat{\sigma_x}}{\sqrt{n}}; \bar{x}_n + t_{n-1;1-\frac{\alpha}{2}} \cdot \frac{\widehat{\sigma_x}}{\sqrt{n}}\right]$
- Bsp. Reaktionszeit:
  $\left[-8.03 - 2.03 \cdot \frac{41.1}{\sqrt{36}}; -8.03 + 2.03 \cdot \frac{41.1}{\sqrt{36}}\right] = [-22.2; 5.61]$ ms

Vorzeichentest = Binomialtest

1. **Modell**: $X_1, X_2, \dots, X_n$ $i.i.d.$, die X_i können beliebig verteilt sein
2. **Nullhypothese**: $\mathcal{H}_0: \mu = \mu_0$, μ ist der Median
   Alternative: $\mathcal{H}_A: \mu \neq \mu_0$ (oder einseitig)
3. **Teststatistik**:
   V: Anzahl X_i's mit $X_i > \mu_0$
   Verteilung unter $\mathcal{H}_0$: $\mathrm{V} \sim \mathrm{Bin}(n, \pi_0)$ mit $\pi_0 = 0.5$
4. **Signifikanzniveau**: α
5. **Verwerfungsbereich** für die Teststatistik:
   $$K = [0, c_u] \cup [c_o, n]$$
   Die Grenzen c_u und c_o müssen mit der Binomialverteilung oder der Normalapproximation berechnet werden.
6. **Testentscheid**:
   Liegt beobachteter Wert v der Teststatistik in K

Bsp. Vorzeichentest

- Angenommen: $\mathcal{H}_0: \mu = \mu_0 = 10, \mathcal{H}_A: \mu \neq 10$
- Beobachtet: $x_1 = 13, x_2 = 9, x_3 = 17, x_4 = 8, x_5 = 14$
- Vorzeichen von $x_i - \mu_0$: +, -, +, -, +
- Mache Binomialtest mit

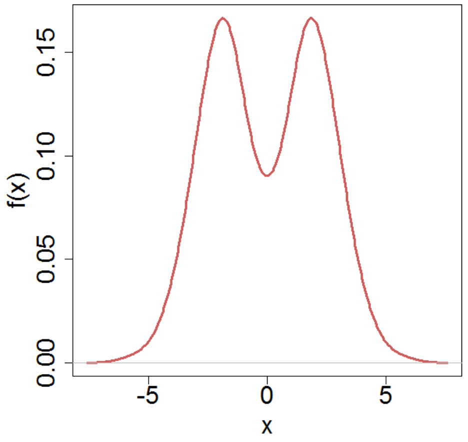
$$\mathcal{H}_0: \pi = 0.5, \mathcal{H}_A: \pi \neq 0.5$$
$$n = 5, v = 3$$

- Der Vorzeichentest kann genau dann verworfen werden, wenn der entsprechende Binomialtest verworfen wird.

Keine Annahme an die Verteilung

Kleinere Macht

Wilcoxon-Test



- Mischung von Vorzeichen- und t-Test
- Annahme: $X_i \sim \mathcal{F}\ i.i.d., \mathcal{F}$ ist symmetrisch
- Teste Median $\mu = \mu_0$
  (einseitig oder zweiseitig)

Bsp. Wilcoxon-Test

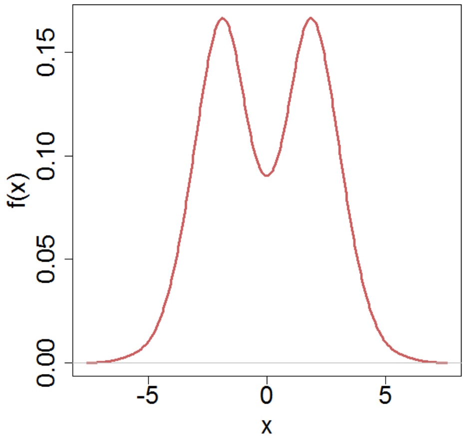
- $\mathcal{H}_0: \mu_0 = 0$
- Beobachtet: -1.9, 0.2, 2.9, -4.1, 3.9
- Absolutbeträge: 1.9, 0.2, 2.9, 4.1, 3.9
- Ränge der Absolutbeträge: 2, 1, 3, 5, 4
- Rangsumme der positiven Gruppe: 1+3+4=8
  Minimale Rangsumme: 0
  Maximale Rangsumme: 1+2+3+4+5=15

- Mit R:

```
 Wilcoxon signed rank test

data:  c(-1.9, 0.2, 2.9, -4.1, 3.9)
V = 8, p-value = 1
alternative hypothesis: true location is not equal to 0
```

Wilcoxon-Test



- Mischung von Vorzeichen- und t-Test
- Annahme: $X_i \sim \mathcal{F}\ i.i.d., \mathcal{F}$ ist symmetrisch
- Teste Median μ: $\mathcal{H}_0: \mu = \mu_0$
  (einseitig oder zweiseitig)

- Intuition der Teststatistik
  - Sortiere $|x_i - \mu_0| \rightarrow r_i$
  - Rängen ursprüngliches Vorzeichen von $(x_i - \mu_0)$ geben
    (*engl. signed ranks*)
  - Teststatistik V: Summe aller Ränge mit $(x_i - \mu_0)$ positiv
- Falls $\mathcal{H}_0$ stimmt, sollte die Rangsumme nicht zu gross und nicht zu klein sein

Übersicht der Tests

	Annahmen				n_{min} bei $\alpha = 0.05$	Macht für Beispiel
	σ_X bekannt	$X_i \sim \mathcal{N}$	symm. Verteilung	i.i.d.		
z-Test	●	●	●	●	1	89%
t-Test		●	●	●	2	79%
Wilcoxon			●	●	6	79%
Vorzeichen				●	5	48%

Verwendetes Beispiel:

- $X_i \sim \mathcal{N}(\mu, \sigma^2), n = 10$
- $\mathcal{H}_0: \mu = 0;\ \mathcal{H}_A: \mu \neq 0; \alpha = 0.05$
- Macht berechnet mit konkreter Alternative: $X_i \sim \mathcal{N}(1,1)$

Stichprobengrösse

- Annahme:
  - $X_i \sim \mathcal{N}(\mu, \sigma^2)\ i.i.d.$
  - σ_X aus Pilotstudie bekannt
- Forderung:
  - Breite von CI kleiner gleich $2 \cdot \rho$
- Gesucht:
  - $n =?$
- Faustregel für 95%-CI:
  - $n \geq 4 \cdot \left(\frac{\sigma}{\rho}\right)^2$

- Bsp. Reaktionszeit:
  - $\sigma = 41.1\ ms$
  - $\rho = 10\ ms$

$$n \geq 4 \cdot \left(\frac{41.1}{10}\right)^2 = 4 \cdot 16.9 \approx 68$$

Zusammenfassung

- z-Test - σ_X bekannt
- t-Test - σ_X unbekannt
- Vorzeichentest - teste Median!
- Wilcoxon-Test - egal welche Verteilung

Hausaufgaben

- Skript: Kapitel 4.7 lessen
- Serie 9 lösen
- Quiz 9 bearbeiten
- etutoR 7