ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Hypothesentests für Erwartungswert und Median

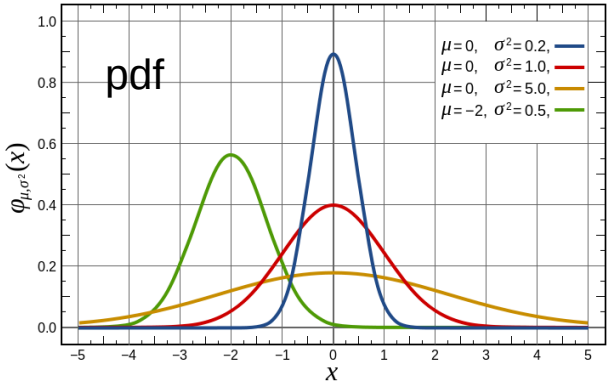
Statistik (Biol./Pharm./HST) – FS 2015

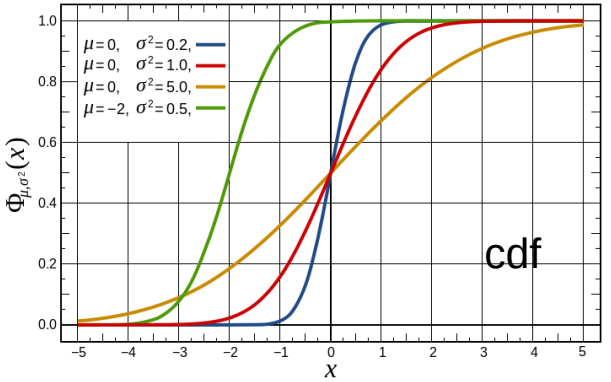
Normalverteilung

- $X \sim N(\mu, \sigma^2)$
  '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: ???





Zentraler Grenzwertsatz (ZGS)

- Ann: $X_1, \ldots, X_n \sim F \; iid; \quad E(X_i) = \mu, \;\; Var(X_i) = \sigma_X^2$

neu

$$\overline{X}_n \approx N(\mu, \underbrace{\frac{\sigma_X^2}{n}})$$

aus GGZ

oder äquivalent mit $S_n = X_1 + \cdots + X_n$:

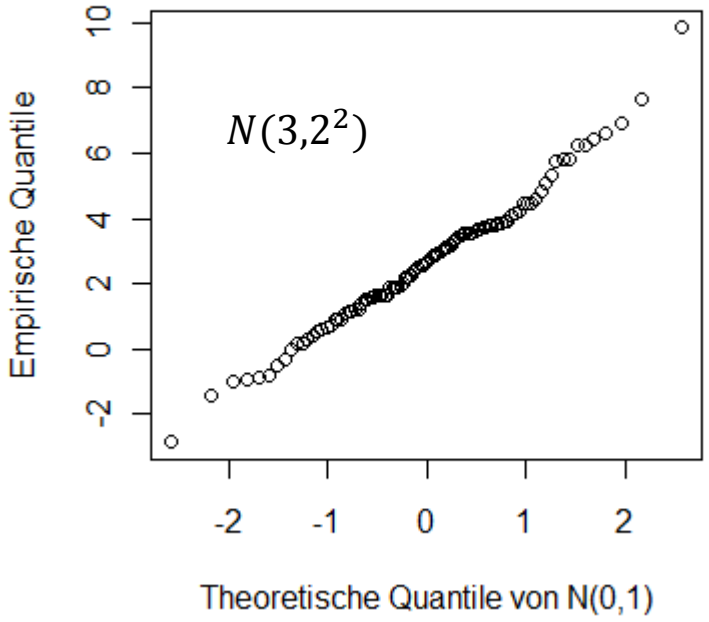
$$S_n \approx N(n\mu, n\sigma_X^2)$$

Wie prüft man, ob eine Normalverteilung vorliegt?

- Histogramm der Daten mit pdf vergleichen
  Schwierig kleine Abweichungen zu erkennen
- Einfacher: QQ-Plot – Theoretische Quantile gegen Empirische Quantile

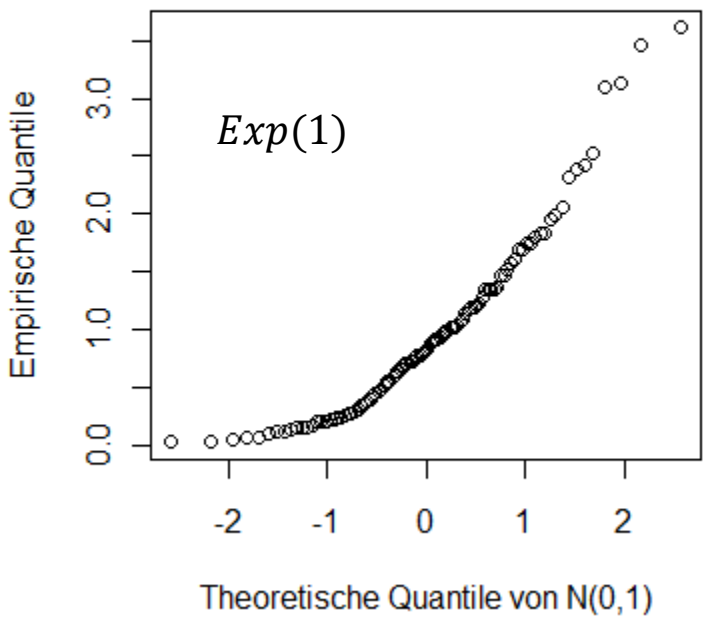
Gerade:

Normalverteilung OK



Krümmung:

Keine Normalverteilung



Reaktionszeit

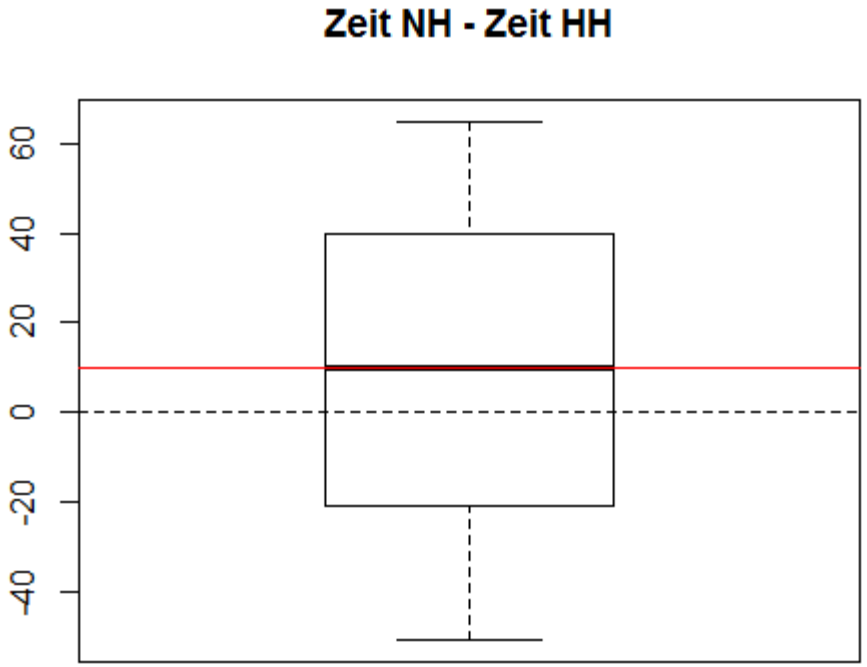
- Reagiert ein Rechtshänder mit rechts schneller als mit links ? (analog für Linkshänder)
- Experiment:
  - **Population:** Alle StudentInnen der VL
    Stichprobe: 50 **zufällige** StudentInnen angeschrieben

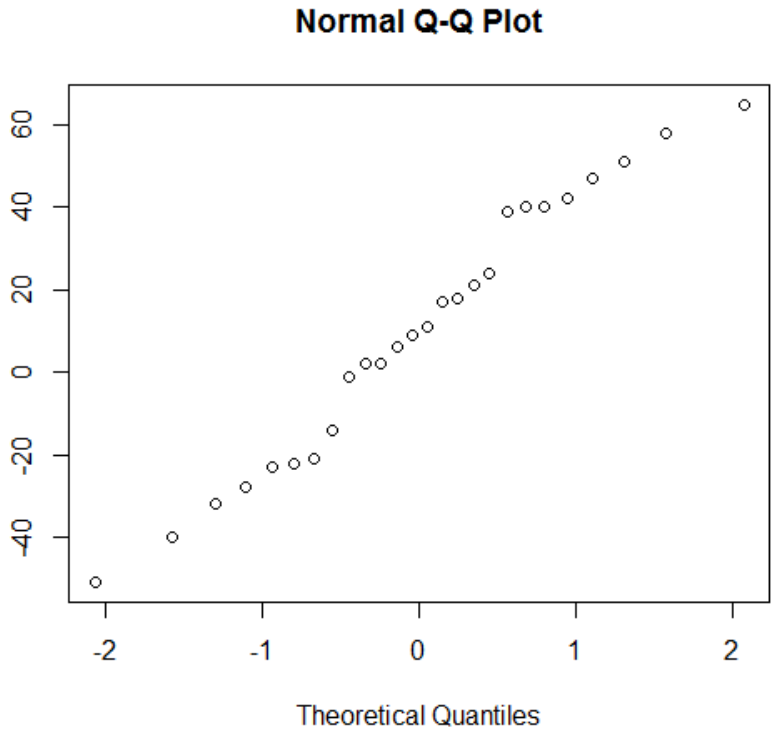
  - Reaktionszeittest online:
  http://www.bbc.co.uk/science/humanbody/sleep/sheep/
  - Mit jeder Hand einmal ausprobieren
  - Dann mit jeder Hand nocheinmal
  - **Reihenfolge randomisiert** (Geburtstag)
  - **Robust:** Bei beiden Messungen das beste und schlechteste Resultat streichen, dann mitteln und an mich senden
  - Berechne “Nebenhand – Haupthand”
  - Z.B.: Haupthand – 227 ms, Nebenhand – 248 ms
  Differenz = 248 ms – 227 ms = 21 ms

  - **Anreiz:** Verlose Gutschein unter Teilnehmern: Rücklauf erhöhen

Ergebnis

- 50 StudentInnen angeschrieben
- Rücklauf: 26
- **Haupthand ist im Mittel 10 ms schneller**





Stichprobe vs. Population

- In der Stichprobe war die Haupthand im Mittel um 10 ms schneller

- Bedeutet das, dass die Nebenhand **auch in der ganzen Bevölkerung** im Mittel schneller ist?

- Antwort darauf:
  z-Test, t-Test, Wilcoxon-Test, Vorzeichen-Test

z-Test: σ_X bekannt

1. Modell: X_i ist eine kontinuierliche Messgrösse
   $X_1, \dots, X_n$ i.i.d. $N(\mu, \sigma_X^2)$, σ_X bekannt
2. Nullhypothese: $H_0: \mu = \mu_0$
   Alternative: $H_A: \mu \neq \mu_0$ (oder < oder >)
3. Teststatistik:
   $$Z = \frac{\overline{X}_n - \mu_0}{\sigma_{\overline{X}_n}} = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{\sigma_X} = \frac{beobachtet - erwartet}{Standardfehler}$$
   Verteilung der Teststatistik unter $H_0: Z \sim N(0,1)$
4. Signifikanzniveau: α
5. Verwerfungsbereich für die Teststatistik:
   $$K = \left(-\infty, -\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right] \cup \left[-\Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \infty\right) \; bei \; H_A: \mu \neq \mu_0$$
   $$K = (-\infty, -\Phi^{-1}(1-\alpha)] \quad bei \quad H_A: \mu < \mu_0$$
   $$K = [\Phi^{-1}(1-\alpha), \infty) \quad bei \quad H_A: \mu > \mu_0$$
6. Testentscheid: Überprüfe, ob der beobachtete Wert der Teststatistik im Verwerfungsbereich liegt.

Problem in Praxis: σ_X ist unbekannt !

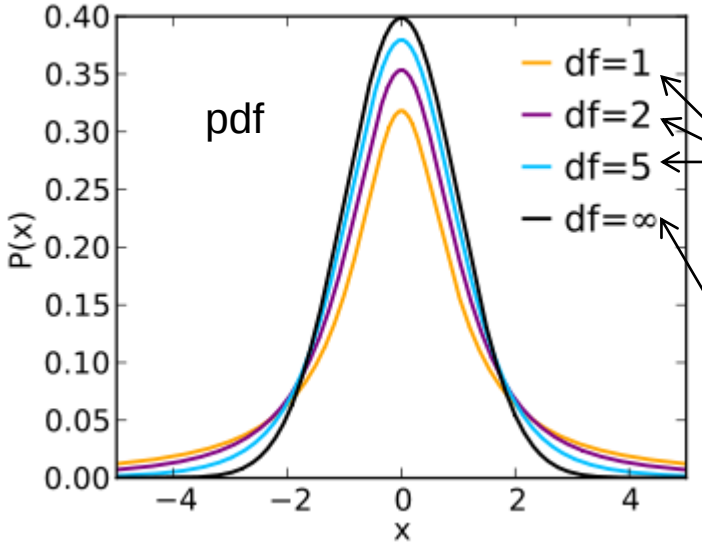
- Schätze Varianz: $\hat{\sigma}_X^2 = \frac{1}{n-1}\sum(X_i - \overline{X}_n)$

- Neue Teststatistik: $T = \frac{\overline{X}_n - \mu_0}{\frac{\hat{\sigma}_X}{\sqrt{n}}}$

- Verteilung von T, falls H_0 stimmt: $T \sim t_{n-1}$

“Student’s” t-Verteilung – Zoo Teil 3

- Annahme: $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma_X^2)$ und unabhängig
- $\hat{\sigma}_X{}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ ist geschätzte Varianz
- $T = (\bar{X}_n - \mu)/(\frac{\hat{\sigma}_X}{\sqrt{n}})$ folgt einer ‘t-Verteilung mit n Freiheitsgraden’, $T \sim t_n$
- Werte sind tabelliert oder im Computer verfügbar
- Falls $n = \infty$: $t_n = N(0,1)$



Mehr Streuung = Unsicherheit

N(0,1)

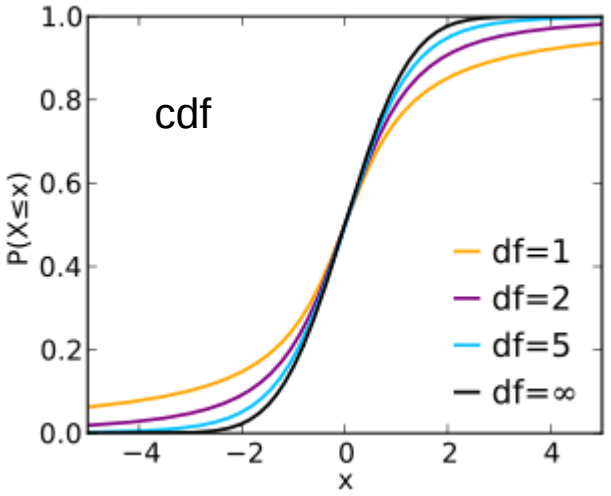
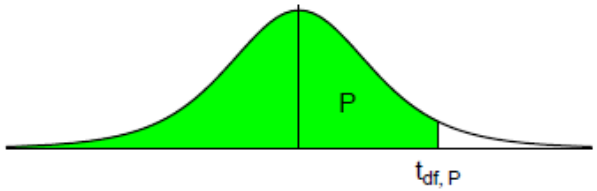


Tabelle: t-Verteilung

Perzentile der t-Verteilung



Bsp.: $t_{9;\ 0.975} = 2.262$

df	$t_{0.60}$	$t_{0.70}$	$t_{0.80}$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787

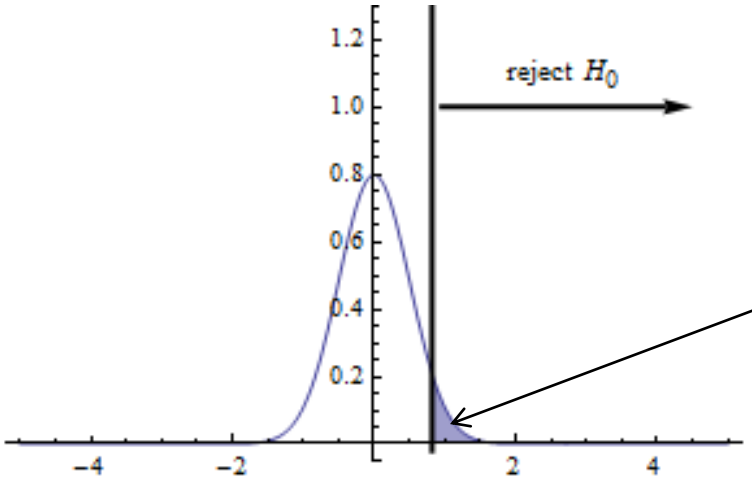
t-Test: σ_X unbekannt

R: t.test
power.t.test

1. Modell: X_i ist eine kontinuierliche Messgrösse
   $X_1, \dots, X_n$ i.i.d. $N(\mu, \sigma_X^2)$, σ_X durch $\widehat{\sigma_X}$ geschätzt
2. Nullhypothese: $H_0: \mu = \mu_0$
   Alternative: $H_A: \mu \neq \mu_0$ (oder < oder >)
3. Teststatistik:
   $$\mathrm{T} = \frac{\overline{X}_n - \mu_0}{\hat{\sigma}_{\overline{X}_n}} = \frac{\sqrt{n}\left(\overline{X}_n - \mu_0\right)}{\hat{\sigma}_X} = \frac{beobachtet - erwartet}{Standardfehler}$$
   Verteilung der Teststatistik unter $H_0: T \sim t_{n-1}$
4. Signifikanzniveau: α
5. Verwerfungsbereich für die Teststatistik:
   $$K = \left(-\infty, -t_{n-1;1-\frac{\alpha}{2}}\right] \cup \left[t_{n-1;1-\frac{\alpha}{2}}, \infty\right) \quad bei \quad H_A: \mu \neq \mu_0$$
   $$K = \left(-\infty, -t_{n-1;1-\alpha}\right] \quad bei \quad H_A: \mu < \mu_0$$
   $$K = \left[t_{n-1;1-\alpha}, \infty\right) \quad bei \quad H_A: \mu > \mu_0$$
6. Testentscheid: Überprüfe, ob der beobachtete Wert der Teststatistik im Verwerfungsbereich liegt.

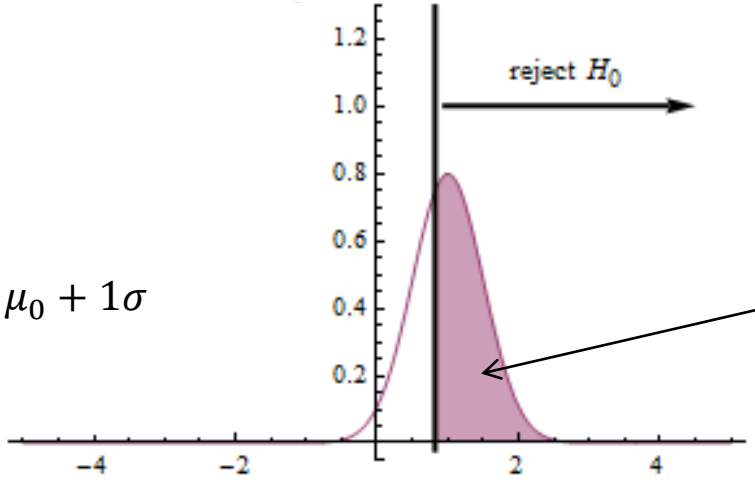
Wdh: Fehler 1. Art & Macht

H_0 wahr



Fehler 1. Art

Angenommen: $\mu = \mu_0 + 1\sigma$



Macht für
$\mu = \mu_0 + 1\sigma$

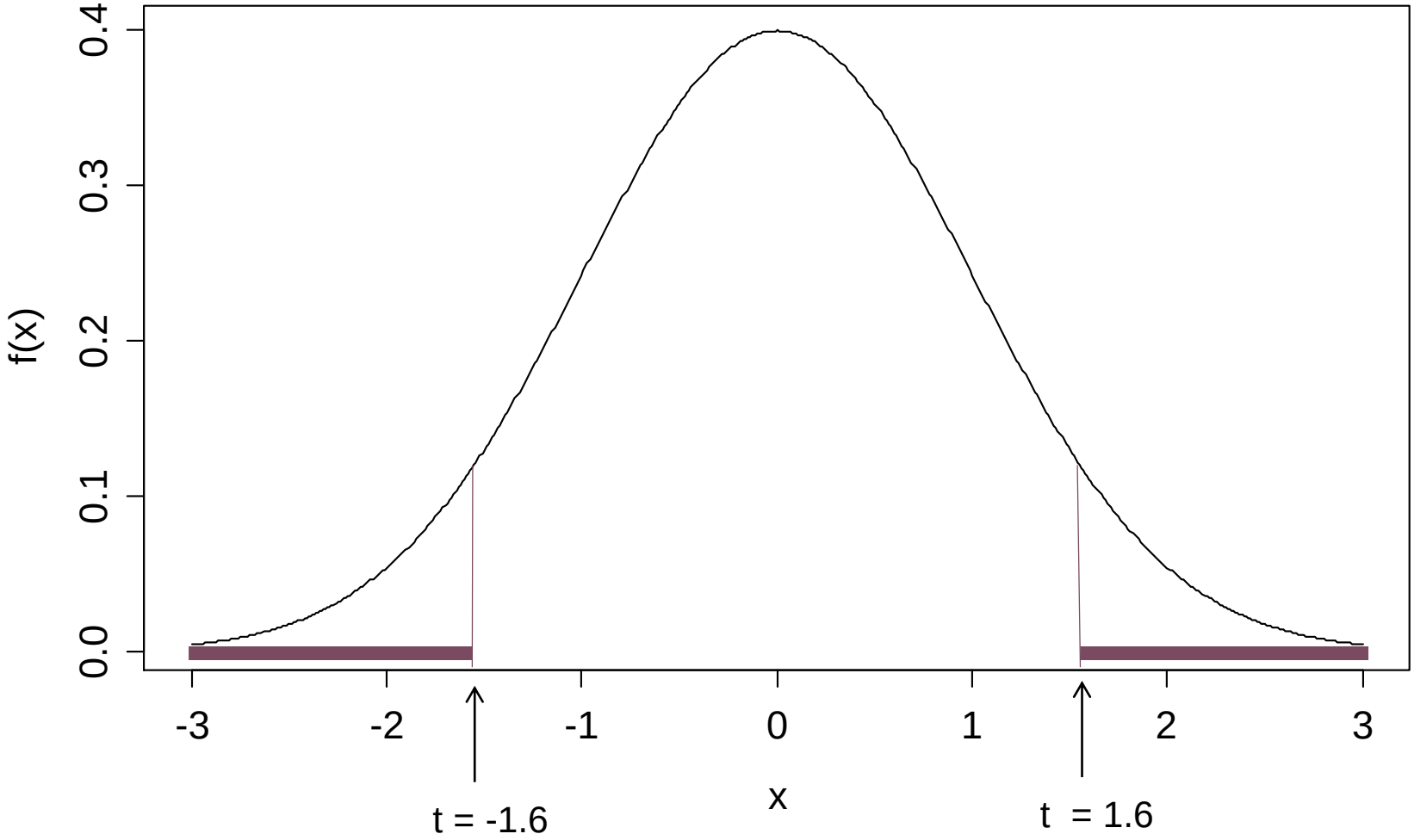
P-Wert (zweiseitiger Test)

Aus Tabelle grob abschätzen:

$p \approx 0.1$

Genauer mit Computer:

$p = 0.125$



(1-α)-Vertrauensintervall für μ

- Äquivalente Definitionen:
  - enthält wahren Wert mit Wa. $1-\alpha$
  - enthält alle Werte μ_0, bei denen $H_0: \mu = \mu_0$ vs. $H_A: \mu \neq \mu_0$ mit Sig.niveau α ***nicht*** verwirft
- Im t-Test, Schritt 5: Nicht verwerfen, falls $\left|\frac{\overline{X}_n - \mu}{\frac{\hat{\sigma}_X}{\sqrt{n}}}\right| < t_{n-1;1-\frac{\alpha}{2}}$; nach μ auflösen

- VI: $\left[\overline{x}_n - t_{n-1;1-\frac{\alpha}{2}} * \frac{\hat{\sigma}_x}{\sqrt{n}}; \overline{x}_n + t_{n-1;1-\frac{\alpha}{2}} * \frac{\hat{\sigma}_x}{\sqrt{n}}\right]$
- Im Bsp: $\left[10 - 2.06 * \frac{32}{\sqrt{26}}; 10 + 2.06 * \frac{32}{\sqrt{26}}\right] =$
  $= [-3; 23]$ ms

Annahmen und Ausblick

- Stichprobenvarianz bekannt (z-Test)
  → gelöst im t-Test
- Daten normalverteilt (z-Test, t-Test)
  → abgeschwächt im Wilcoxon Test
  → gelöst im Vorzeichentest
- Beobachtungen unabhängig und gleich verteilt
  → keine einfache Lösung

Vorzeichentest = Binomialtest

1. Modell: $X_1, \ldots, X_n$ iid
   wobei X_1 eine beliebige Verteilung hat.
2. Nullhypothese: $H_0: \mu = \mu_0$ (μ ist der Median)
   Alternative: $H_A: \mu \neq \mu_0$ (oder einseitige Variante)
3. Teststatistik: V: Anzahl X_is mit $(X_i > \mu_0)$
   Verteilung der Teststatistik unter $H_0: V \sim Bin(n, \pi_0)\ mit\ \pi_0 = 0.5$
4. Signifikanzniveau: α
5. Verwerfungsbereich für die Teststatistik:
   $K = [0, c_u] \cup [c_o, n]$ falls $H_A: \mu \neq \mu_0$
   Die Grenzen c_u und c_o müssen mit der Binomialverteilung oder der Normalapproximation berechnet werden.
6. Testentscheid: Entscheide, ob der beobachtete Wert der Teststatistik im Verwerfungsbereich der Teststatistik liegt.

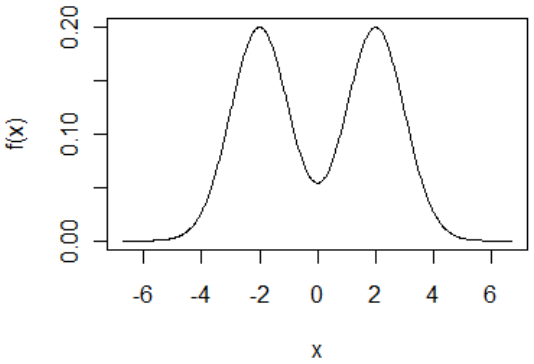
Bsp: Vorzeichentest

- Angenommen: $\mathrm{H}_0: \mu = \mu_0 = 10$, $H_A: \mu \neq 10$
- 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
  $H_0: \pi = 0.5$, $H_A: \pi \neq 0.5$, n=5, x=3 (Anzahl '+')

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

- Vorteil: Keine Annahme an Verteilung
- Nachteil: Kleinere Macht

Wilcoxon-Test: Intuition



- Kompromiss zw. Vorzeichen- und t-Test
- Annahme: $X_i \sim F \;\; iid$, F ist symmetrisch
- Teste Median μ: $H_0: \mu = \mu_0$
  (einseitig oder zweiseitig)

Bsp: Wilcoxon-Test

- Bsp: H_0: $\mu_0 = 0$
- Beobachte -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 posititven Gruppe: 1+3+4=8
  Minimale Rangsumme: 0
  Maximale Rangsumme: 1+2+3+4+5 = 15

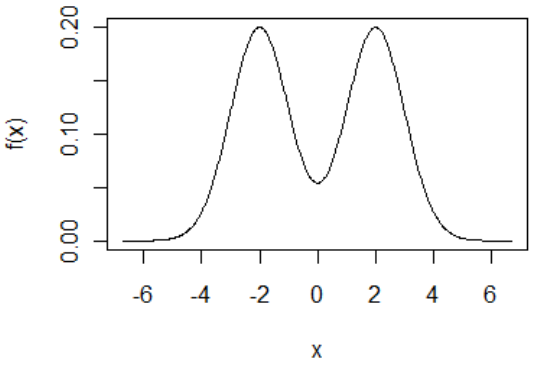
- Mit dem Computer:

```
> wilcox.test(c(-1.9, 0.2, 2.9, -4.1, 3.9), mu=0)

 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: Intuition



- Kompromiss zw. Vorzeichen- und t-Test
- Annahme: $X_i \sim F \;\; iid$, F ist symmetrisch
- Teste Median μ: $H_0: \mu = \mu_0$
  (einseitig oder zweiseitig)

- Intuition der Teststatistik
  - Rangiere $|\mathrm{x}_i - \mu_0| \rightarrow \mathrm{r_i}$
  - Gib Rängen ursprüngliches Vorzeichen von $(x_i - \mu_0)$ (“signed ranks”)
  - Teststatistik T: Summe aller Ränge, bei denen $(x_i - \mu_0)$ positiv ist
- Falls H_0 stimmt, sollte diese Rangsumme nicht zu gross und nicht zu klein sein

Übersicht der Tests

Test	Annahmen				n_{min} bei $\alpha = 0.05$	Macht für ein Beispiel (1)
	σ_X bekannt	$X_i \sim N$	Symm. Verteilung	iid		
z	x	x	x	x	1	89 %
t		x	x	x	2	79 %
Wilcoxon			x	x	6	79 %
VZ				x	5	48 %

(1): $\mathrm{X_i} \sim \mathrm{N}(\mu, \sigma^2), n = 10;\ H_0: \mu = 0; H_A: \mu \neq 0; \alpha = 0.05$

Macht berechnet für konkrete Alternative: $X_i \sim N(1,1)$