Statistik II für Betriebswirte
Vorlesung 2

Dr. Andreas Wünsche

TU Bergakademie Freiberg
Institut für Stochastik

21. Oktober 2019

Test für die Wahrscheinlichkeit p eines Ereignisses

- Sei p die Wahrscheinlichkeit für das Eintreten eines Ereignisses A.
- Die Zufallsgröße X mit $X(\omega) = \begin{cases} 1\,, & \text{falls } \omega \in A\,; \\ 0\,, & \text{sonst}\,, \end{cases}$

  ist Bernoulli-verteilt mit Parameter p.
- Eine Stichprobe $X_1, \ldots, X_n$ vom Umfang n besteht aus den Ergebnissen von n unabhängigen Wiederholungen dieses Versuches, d.h. tritt beim i-ten Versuch A ein, setzt man $X_i = 1$, ansonsten $X_i = 0$.
- $Y = \sum\limits_{i=1}^{n} X_i \sim Bin(n, p)$.

Test für die Wahrscheinlichkeit p ,Binomialtest

- **Nullhypothese:** $H_0 \; : \; p = p_0 \quad \longrightarrow \quad Y = \sum_{i=1}^{n} X_i \overset{H_0}{\sim} Bin(n, p_0)$.

- Die Testentscheidung wird basierend auf den Quantilen der Binomialverteilung mit den Parametern n und p_0 herbeigeführt.

- Für einen Wert α mit $0 < \alpha < 1$ bezeichne $q_{\alpha,u}$ die **kleinste** ganze Zahl für die gilt:

  $$P_{H_0}(Y \leq q_{\alpha,u}) = P_{H_0}(Y = 0) + P_{H_0}(Y = 1) + \ldots + P_{H_0}(Y = q_{\alpha,u}) > \alpha$$

  und $q_{\alpha,o}$ die **größte** ganze Zahl mit

  $$\begin{aligned} P_{H_0}(Y \geq q_{\alpha,o}) \;\; &= \;\; P_{H_0}(Y = n) + P_{H_0}(Y = n-1) + \ldots \\ &\qquad \ldots + P_{H_0}(Y = q_{\alpha,o}) > \alpha\,. \end{aligned}$$

-
  $$P_{H_0}(Y = k) = \binom{n}{k} {p_0}^k (1 - p_0)^{n-k}\,, \quad k = 0, \ldots, n\,.$$

Binomialtest

- $T = Y = \sum_{i=1}^{n} X_i$
- Zweiseitig: $H_0: \ p = p_0$ gegen $H_A: \ p \neq p_0$
  $$K = \{t \mid t < q_{\alpha/2,u} \text{ oder } t > q_{\alpha/2,o}\}.$$
- Rechtsseitig: $H_0: \ p = p_0$ gegen $H_A: \ p > p_0$
  (oder auch so: $H_0: \ p \leq p_0$ gegen $H_A: \ p > p_0$)
  $$K = \{t \mid t > q_{\alpha,o}\}.$$
- Linksseitig: $H_0: \ p = p_0$ gegen $H_A: \ p < p_0$
  (oder auch so: $H_0: \ p \geq p_0$ gegen $H_A: \ p < p_0$)
  $$K = \{t \mid t < q_{\alpha,u}\}.$$
- Der exakte Binomialtest ist konservativ, das heißt, das Niveau α wird nicht immer ganz ausgeschöpft.
- Für große Stichprobenumfänge n ist es sinnvoll, den approximativen Binomialtest zu verwenden.

Approximativer Binomialtest für die Wahrscheinlichkeit p

- **Testgröße:** $T = \dfrac{\sum_{i=1}^{n} X_i - np_0}{\sqrt{np_0(1-p_0)}}$
- Satz von Moivre-Laplace $\Longrightarrow$ Die Verteilung von T konvergiert unter $H_0: \; p = p_0$ gegen $\mathrm{N}(0,1)$ für $n \to \infty$.
- $H_A: \; p \neq p_0 \quad \Longrightarrow \quad K = \left\{ t \in \mathbb{R}: \; |t| > z_{1-\alpha/2} \right\}.$
- $H_A: \; p > p_0 \quad \Longrightarrow \quad K = \left\{ t \in \mathbb{R}: \; t > z_{1-\alpha} \right\}.$
- $H_A: \; p < p_0 \quad \Longrightarrow \quad K = \left\{ t \in \mathbb{R}: \; t < -z_{1-\alpha} \right\}.$
- Der approximative Binomialtest ist ein asymptotischer Test und sollte nur für **große** n (Faustregel: $np_0(1-p_0) \geq 9$) verwendet werden.

Beispiel 4.2: Ausschussprozentsatz Buchsen

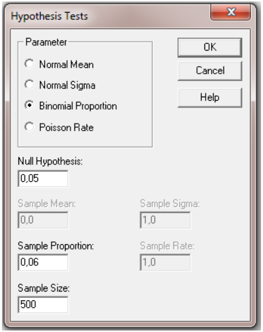
- In der Gütekontrolle einer Automatendreherei wird ein Posten Buchsen ausgeliefert.
- Es soll zum Signifikanzniveau $\alpha = 0.01$ festgestellt werden, ob der zulässige Ausschussprozentsatz von 5% überschritten wird.
- Ein Kontrolleur entnimmt dem Posten eine Stichprobe vom Umfang $n = 500$ und untersucht an den einzelnen Teilen als wichtigstes Merkmal den Buchsenansatz.
- Bei 30 Buchsen liegt der Messwert für den Buchsenansatz nicht innerhalb der vorgeschriebenen Toleranzen, sie sind Ausschuss.
- X_i ist Bernoulli-verteilt mit Parameter p, $n = 500$, $\sum_{i=1}^{n} x_i = 30$,

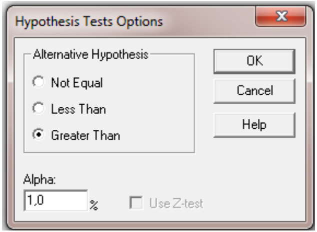
  d.h. $\hat{p} = \overline{x} = \frac{30}{500} = 0.06$.

Fortsetzung Beispiel 4.2: Ausschussprozentsatz Buchsen

- $H_0 \ : \ p \leq 0.05\,, \quad H_A \ : \ p > 0.05\,; \quad \alpha = 0.01\,.$
- $n \cdot p_0 \cdot (1 - p_0) = 500 \cdot 0.05 \cdot 0.95 = 23.75 > 9\,,$
  der approximative Test kann also verwendet werden.
- $T = \dfrac{\sum\limits_{i=1}^{n} X_i - np_0}{\sqrt{np_0(1-p_0)}}\,; \qquad t = \frac{30-500\cdot 0.05}{\sqrt{500\cdot 0.05\cdot 0.95}} = 1.026\,.$
- $K = \{t \in \mathbb{R} \ : \ t > z_{1-\alpha}\} = (z_{0.99}, \infty) = (2.326, \infty)\,.$
- $t \notin K \ \Rightarrow \ H_0$ wird nicht abgelehnt, der Ausschussanteil der Buchsen ist nicht signifikant größer als 5%.
- Mehr dazu im letzten Kapitel: Statistische Qualitätskontrolle
- In Statgraphics wird der p-Wert des exakten Binomialtests bestimmt.

Beispiel 4.2: Statgraphics





Hypothesis Tests

Sample proportion = 0,06
Sample size = 500

Approximate 99,0% lower confidence bound for p: [0,0378782]

Null Hypothesis: proportion = 0,05
Alternative: greater than
P-Value = 0,176471
Do not reject the null hypothesis for alpha = 0,01.

Vorzeichentest (auch Zeichentest)

- **Grundlage:** Für den Median $X_{0.5}$ einer **stetigen** Zufallsgröße gelten $P(X < X_{0.5}) = 0.5$ und $P(X > X_{0.5}) = 0.5$.
- **Voraussetzung:** Stichprobe $X_1, \ldots, X_n$ mit stetiger Verteilungsfunktion F_X für die Zufallsgrößen X_i.
- Der Vorzeichentest (auch Zeichentest) dient z.B. als Test für den **Median** der Verteilung.
- Eine Modifikation des Vorzeichentests, den Vorzeichen-Rang-Test (bzw. Wilcoxon-Vorzeichen-Test), werden wir erst im Abschnitt 4.3.2, Tests für zwei verbundene Stichproben, betrachten.
- M_0 sei ein hypothetischer Wert für den Median der Verteilung.
- **Hypothesen:** $H_0 : X_{0.5} = M_0$, $H_A : X_{0.5} \neq M_0$ (zweiseit. Test).
- $Y_i^+ := \begin{cases} 1, & \text{falls } X_i > M_0, \\ 0, & \text{falls } X_i < M_0. \end{cases}$

Vorzeichentest für Median

- **Test für kleine** n:
  - **Testgröße:** $T = \sum_{i=1}^{n} Y_i^+ \overset{H_0}{\sim} \mathrm{Bin}(n, 0.5)\,.$
  - **Kritischer Bereich:** $K = \{0, 1, \ldots, c\} \cup \{n-c, \ldots, n\}\,,$

    so dass $\displaystyle\sum_{i=0}^{c} \binom{n}{i} \left(\frac{1}{2}\right)^n \leq \frac{\alpha}{2}, \quad \sum_{i=0}^{c+1} \binom{n}{i} \left(\frac{1}{2}\right)^n > \frac{\alpha}{2}$

    (Binomialtest mit $p_0 = 0.5 = 1 - p_0$).
- **Test für große** n $(n \geq 20)$:
  - **Testgröße:** $T = \frac{1}{\sqrt{n}} \left(2 \sum_{i=1}^{n} Y_i^+ - n \right).$
  - **Kritischer Bereich:** $K = \{ t \in \mathbb{R} \,:\, |t| > z_{1-\alpha/2} \}\,.$

Beispiel 4.3 Vorzeichentest für Median: Mietspiegel

- Es soll überprüft werden, ob der Median der Quadratmetermiete für Wohnungen unter 50 m^2, die nach 1983 gebaut wurden, größer ist als der aus einer anderen Stadt bekannte Wert von 8 €/m^2 !
- **Daten:** (Quelle: Fahrmeier, Künstler, Pigeot, Tutz: Statistik. Der Weg zur Datenanalyse, 2007, S.444), $n = 11$:

13.22	6.81	10.22	14.03	8.04	10.16
9.43	13.07	13.63	5.05	11.63	

- **Hypothesen:** $H_0 : X_{0.5} = 8\,, \quad H_A : X_{0.5} > 8\,; \quad \alpha = 0.05\,.$
- **Realisierung der Testgröße** T : $t = 9\,.$
- **Kritischer Bereich:** $c = 2$ und damit $K = \{9, 10, 11\}$, da $\left(\frac{1}{2}\right)^{11} \cdot \left(\binom{11}{0} + \binom{11}{1} + \binom{11}{2}\right) = 0.0327$, aber $\left(\frac{1}{2}\right)^{11} \cdot \left(\binom{11}{0} + \binom{11}{1} + \binom{11}{2} + \binom{11}{3}\right) = 0.113\,.$
- H_0 wird abgelehnt, der Median ist signifikant größer als 8 €/m^2 .

Bemerkungen zum Vorzeichentest

- Theoretisch ist aufgrund der stetigen Verteilung von X ein Wert $X_i = M_0$ nicht möglich. Praktisch kann es aber, z.B. durch Runden der Werte, in der beobachteten Stichprobe Werte $x_i = M_0$ geben.
- Wie kann man Werte $x_i = M_0$ in der Stichprobe vermeiden?
  - Indem man die Messgenauigkeit erhöht.
  - Indem man die gemessenen Werte nicht rundet.
- Wie kann man vorgehen, falls es Werte $x_i = M_0$ in der Stichprobe gibt?
  - Beobachtungen mit $x_i = M_0$ werden nicht berücksichtigt ($\Rightarrow$ geringerer Stichprobenumfang).
  - Beobachtungen mit $x_i = M_0$ werden zu gleichen Teilen beiden Gruppen ($y_i^+ = 1$ bzw. $y_i^+ = 0$) zugeordnet, bei ungerader Anzahl wird eine Beobachtung nicht berücksichtigt.
  - Die Beobachtungen mit $x_i = M_0$ werden zufällig mit einer Wahrscheinlichkeit von 0.5 einer der beiden Gruppen zugeordnet.

4.3. Tests für zwei Stichproben

Man unterscheidet hier zwischen unabhängigen und verbundenen Stichproben.

- Zwei unabhängige Stichproben (auch unverbundene oder ungepaarte Stichproben genannt):

$$X_{1i}\,, \quad \text{iid.}\,, \quad i = 1, \ldots, n_1\,,$$
$$X_{2i}\,, \quad \text{iid.}\,, \quad i = 1, \ldots, n_2\,.$$

  Die Stichprobe des Merkmals X_1 ist dabei, wie der Name schon sagt, unabhängig von der Stichprobe des Merkmals X_2.
- Liegen zwei Stichproben vor, deren Werte einander paarweise zugeordnet sind, spricht man von verbundenen (auch abhängige oder gepaarte) Stichproben.
- Die mathematische Modellierung erfolgt über zwei endliche Folgen $X_1, X_2, \ldots, X_n$ und $Y_1, Y_2, \ldots, Y_n$ von jeweils n Zufallsgrößen.
- Eine verbundene (mathematische) Stichprobe wird also durch unabhängige Zufallsvektoren $(X_1, Y_1), \ldots, (X_n, Y_n)$ modelliert.

4.3.1. Tests für zwei unabhängige Stichproben

Zwei unabhängigen Stichproben

$$X_{1i}\,, \quad \text{iid.}\,, \quad i = 1, \ldots, n_1\,,$$
$$X_{2i}\,, \quad \text{iid.}\,, \quad i = 1, \ldots, n_2\,.$$

- Tests für die Lage bzw. zentrale Tendenz, Lagevergleich
  - Stichproben sind normalverteilt
    - Varianzen sind gleich $\sigma_1^2 = \sigma_2^2$ und unbekannt: Doppelter t-Test
    - Varianzen sind unbekannt: Welch-Test
  - Bei beiden Stichproben liegt eine stetige Verteilung mit gleicher Verteilungsform vor: Wilcoxon-Rangsummentest
  - Bei beiden Stichproben liegt eine stetige Verteilung vor: Permutationstest (im freien Statistik-Softwarepaket `R` mit Funktion perm.test() aus Zusatzpaket exactRankTests).
- Test für die Streuungen, Streuungsvergleich
  - Stichproben sind normalverteilt: F-Test
  - Bei beiden Stichproben liegt eine stetige Verteilung vor: Ansari-Bradley-Test (in `R` mit Funktion ansari.test()).

Doppelter $t-$Test

- **Geg.:** zwei unabh. Merkmale $X_1 \sim \mathrm{N}(\mu_1, \sigma^2)$, $X_2 \sim \mathrm{N}(\mu_2, \sigma^2)$; entsprechend zwei Stichproben vom Umfang n_1 und n_2 mit arithmetischen Mittelwerten $\overline{X}_1$ und $\overline{X}_2$ und Stichprobenvarianzen S_1^2 und S_2^2.
- **Voraussetzung:** beide Varianzen sind gleich σ^2, σ^2 ist unbekannt.
- **Hypothesen:** $H_0 : \mu_1 = \mu_2$, $H_A : \mu_1 \neq \mu_2$ (zweiseitiger Test).
- **Testgröße:** $T = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}} \sqrt{\dfrac{n_1 n_2}{n_1 + n_2}}$.

  Im Fall von $n_1 = n_2 = n$ gilt $T = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_1^2 + S_2^2}} \sqrt{n}$.
- **Kritischer Bereich:** $K = \{t \in \mathbb{R} : |t| > t_{n_1+n_2-2;1-\alpha/2}\}$.
- Die Tests sind für große Werte n_1, n_2 (Faustregel: $n_1, n_2 \geq 30$) auch ohne Normalverteilungsvoraussetzung anwendbar.

Welch-Test

- **Geg.:** Zwei unabh. Merkmale $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$; entsprechend zwei Stichproben vom Umfang n_1 und n_2 mit arithmetischen Mittelwerten $\overline{X}_1$ und $\overline{X}_2$ und Stichprobenvarianzen S_1^2 und S_2^2.
- **Voraussetzung:** σ_1^2, σ_2^2 sind unbekannt.
- **Hypothesen:** $H_0 : \mu_1 = \mu_2$, $\quad H_A : \mu_1 \neq \mu_2$ (zweiseitiger Test).
- **Testgröße:** $T = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$
- **Kritischer Bereich:** $K = \{t \in \mathbb{R} : \; |t| > t_{m;1-\alpha/2}\}$,

$$m = \left[\frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2} \right]_{\mathrm{int}}$$ (Abrundung zur ganzen Zahl)

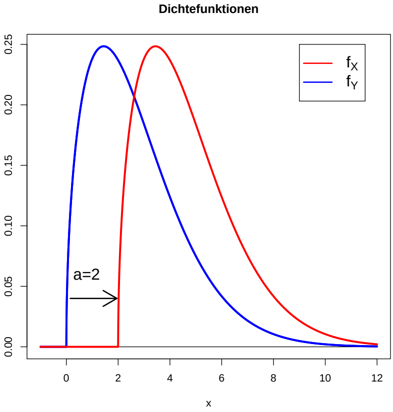
- **Bemerkung:** Dies ist ein approximativer Test.

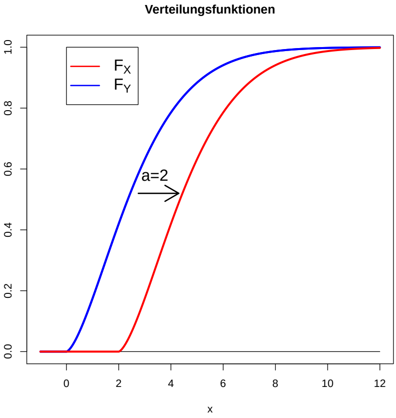
Wilcoxon-Rangsummentest

- Der Wilcoxon-Rangsummentest dient zum Vergleich zweier unabhängiger Stichproben hinsichtlich ihrer Lage.
- Er kann im Fall von nicht-normalverteilten Grundgesamtheiten an Stelle des doppelten $t-$Tests verwendet werden.
- Er wird auch als Rangtest nach Wilcoxon bezeichnet und ist äquivalent zum U-Test von Mann-Whitney.
- Er kann als zweiseitiger oder als einseitiger Test ausgeführt werden.
- **Geg.:** 2 unabhängige Stichproben
  $X_1, \ldots, X_{n_1}$ mit stetiger Verteilungsfunktion F_X und
  $Y_1, \ldots, Y_{n_2}$ mit stetiger Verteilungsfunktion F_Y, wobei
  $F_Y(t) = F_X(t+a)$ mit einer reellen Zahl a vorausgesetzt wird.

Wilcoxon-Rangsummentest

- Wenn $F_Y(t) = F_X(t+a)$ mit einer reellen Zahl a vorausgesetzt wird, besitzen beide Stichproben die gleiche Verteilungsform.
- Falls Erwartungswerte existieren, gilt damit $\mu_X = \mathbf{E}X = \mathbf{E}Y + a = \mu_Y + a$.





Wilcoxon-Rangsummentest – kleine Stichprobenumfänge

- **Hypothesen:** $H_0 : a = 0\,,\quad H_A : a \neq 0$ (zweiseitiger Test) (bzw. $H_0 : \mu_X = \mu_Y\,,\ H_A : \mu_X \neq \mu_Y$).
- In der gemeinsamen Stichprobe werden die Ränge bestimmt. Die **Testgröße** $T = R_1$ ist die Summe der Ränge zu der ersten Stichprobe.
- **Kritischer Bereich** (n_1, n_2 **klein**):

$$K = \{t > 0 \,:\, t \leq w_{n_1,n_2;\alpha/2}\} \cup \{t > 0 \,:\, t \geq w_{n_1,n_2;1-\alpha/2}\}\,;$$

  die Quantile $w_{n_1,n_2;\alpha/2}$ und $w_{n_1,n_2;1-\alpha/2}$ kann man in Tabellen finden, dabei gilt

$$w_{n_1,n_2;1-\alpha} = n_1(n_1 + n_2 + 1) - w_{n_1,n_2;\alpha}\,.$$

- **Grundüberlegung:** Beide Stichproben sollten sich unter H_0 ungefähr gleichartig in der gemeinsamen geordneten Stichprobe durchmischen.

Wilcoxon-Rangsummentest – große Stichprobenumfänge

- **Hypothesen:** $H_0 : a = 0\,, \quad H_A : a \neq 0$ (zweiseitiger Test) (bzw. $H_0 : \mu_X = \mu_Y\,,\ H_A : \mu_X \neq \mu_Y$).
- **Testgröße:** Mit der Summe R_1 der Ränge zur ersten Stichprobe in der gemeinsamen Stichprobe nutzt man

$$T = \frac{R_1 - \frac{1}{2}n_1(n_1+n_2+1)}{\sqrt{\frac{1}{12}n_1 n_2(n_1+n_2+1)}}\,.$$

- **Kritischer Bereich** (n_1, n_2 **groß**): (**Faustregel:** $n_1 \geq 4\,, n_2 \geq 4\,, n_1+n_2 \geq 20$)

$$K = \{t \in \mathbb{R} : |t| > z_{1-\alpha/2}\}$$

(da T näherungsweise standardnormalverteilt ist).

Beispiel 4.4: Flugschrauber

- Mit zwei Typen von Flugschraubern wurden jeweils 6 Flüge zwischen zwei Flughäfen durchgeführt und die totale Flugzeit in Minuten gemessen. Wird diese Strecke im Mittel gleich schnell bewältigt ?
- **Daten:** (Quelle: Aczel, Sounderpandian: Complete Business Statistics, 2006, Bsp.14-4)

Modell A	35	38	40	42	41	36
Modell B	29	27	30	33	39	37

- Ordnen und Rangvergabe (Ränge zur ersten Stichprobe in rot):

A					35	36		38		40	41	42
B	27	29	30	33			37		39			
Rang	1	2	3	4	5	6	7	8	9	10	11	12

Fortsetzung Beispiel 4.4: Flugschrauber

- **Hypothesen:** $H_0 : \mu_A = \mu_B$, $H_A : \mu_A \neq \mu_B$, $\alpha = 0.05$.
- Wert der **Testgröße** des Wilcoxon-Rangsummentests:
  $t = r_1 = 5 + 6 + 8 + 10 + 11 + 12 = 52$.
- **Kritischer Bereich:** aus Tabelle
  $w_{6,6;0.025} = 26$, $w_{6,6;0.975} = 6 \cdot 13 - 26 = 52$,
  $\Rightarrow \quad K = \{t > 0 : t \leq 26\} \cup \{t > 0 : t \geq 52\}$.
- **Testergebnis:** $t \in K$, H_0 wird abgelehnt, die Flugzeiten sind signifikant unterschiedlich.

Bindungen

- Besitzen zwei oder mehrere Beobachtungen in einer Stichprobe den gleichen Wert, so wird dies als Bindung (engl. „ties“) bezeichnet. Eine eindeutige Zuweisung der Ränge ist damit nicht mehr möglich.
- Bindungen kann man vermeiden, indem man die Messgenauigkeit erhöht und die gemessenen Werte nicht rundet (vgl. Bemerkungen zum Vorzeichentest).
- Der Testaufbau sieht Bindungen eigentlich nicht vor (die Wahrscheinlichkeit dafür ist Null), deshalb muss man die Tests geeignet modifizieren.
- In der Praxis werden beim Auftreten von Bindungen häufig Durchschnittsränge gebildet.
- Bei bestimmten Tests, wie z.B. dem Kruskal-Wallis-Test (er wird später behandelt), muss die Testgröße beim Vorliegen von Bindungen in der Stichprobe entsprechend angepasst werden (siehe Literatur).

$F-$Test

- **Geg.:** Zwei unabh. Merkmale $X_1 \sim \mathrm{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathrm{N}(\mu_2, \sigma_2^2)$; entsprechend zwei Stichproben vom Umfang n_1 und n_2 mit Stichprobenvarianzen S_1^2 und S_2^2.
- **Voraussetzung:** μ_1, μ_2 unbekannt.
- **Hypothesen:** $H_0 : \sigma_1^2 = \sigma_2^2$, $H_A : \sigma_1^2 \neq \sigma_2^2$ (zweiseitiger Test).
- **Testgröße:** $T = \dfrac{S_1^2}{S_2^2}$.
- **Kritischer Bereich:**
$$K = \{t > 0 \,:\, t < F_{n_1-1;n_2-1;\alpha/2}\} \cup \{t > 0 \,:\, t > F_{n_1-1;n_2-1;1-\alpha/2}\}$$
(Quantile der $F-$Verteilung mit $(n_1 - 1; n_2 - 1)$-Freiheitsgraden).