Moderne Methoden der Datenanalyse
– Hypothesentests –

Roger Wolf

11. Juni 2020

Inhalt dieser Vorlesung

- Testverteilung und Irrtumswahrscheinlichkeit.
- Testverfahren.
- Fehler 1. und 2. Art.
- Operationscharakteristik.
- Beispiele:
  - Zweistichprobentest
  - Varianzanalyse

Priv. Doz. Dr. Roger Wolf
http://ekpwww.physik.uni-karlsruhe.de/~rwolf/

Hypothesentests

- Hypothesentests dienen dazu eine Hypothese mit gegebener Verteilungsfunktions einer Teststatistik auf ihre Vereinbarkeit mit dem Ergebnis einer Stichprobe (=Messung) hin zu überprüfen.

- Die zu testende Hypothese wird im allgemeinen Nullhypothese (H_0) genannt. Unwahrscheinliche Ausgänge der Stichprobe werden zur Widerlegung der Nullhypothese herangezogen, sofern sie „signifikant“ sind.

- „Signifikant“ bedeutet, dass die Abweichung der Stichprobe von der Erwartung der Hypothese im Rahmen des Fürwahrhaltens des Testenden nicht mehr durch den Zufallsfehler erklärbar ist.

- Alternativ kann die Nullhypothese gegen eine oder mehrere alternative Hypothesen H_i getestet werden.

Beispiele

- Hypothesentests sind das Hauptanwendungsgebiet der Statistik. Die Anwendungen sind sehr vielfältig und müssen ggf. von Fall zu Fall sorgfältig mathematisch formuliert werden.

- **Beispiele**:
  - Können zwei Stichproben bezüglich des betrachteten Parameters aus der selben Grundgesamtheit stammen (z.B. $H_0\colon\ \mu_1 = \mu_2$)?
  - Spricht die Energiedeposition des Teilchens in Blei eher einem Pion oder einem Elektron?
  - Gibt es ein Signal auf einem bekannten Untergrund oder nicht?
  - Ist der Spin des beobachteten Teilchens 0 oder 1?
  - Verdienen Frauen weniger als Männer bei gleicher Qualifikation?
  - Ist das Medikament wirksam oder nicht?
  - Ist der Patient krank oder nicht?

Anforderung an die Hypothese

- Die zu testende Hypothese muss so formuliert sein, dass sie überprüfbar (d.h. „falsifizierbar“) ist.

- Die Annahme der Gültigkeit der Hypothese muss die Bestimmung einer Stichprobenverteilung erlauben. Die Stichprobenverteilung wird in diesem Fall auch als Testverteilung bezeichnet.

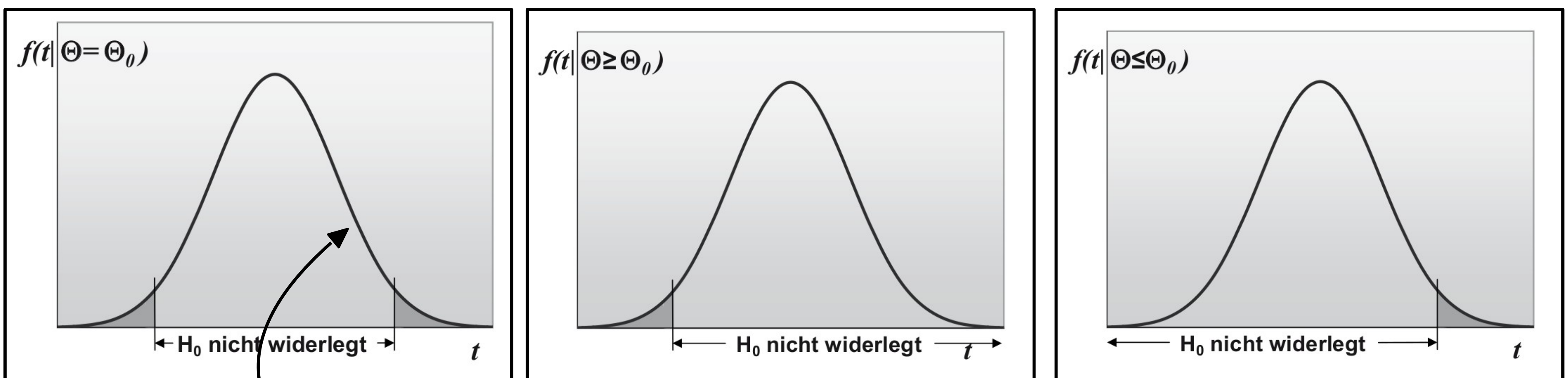
Testverteilung

- Um eine Testverteilung $f(t|\Theta=\Theta_0)$ angeben zu können, muss bei Parametertests der zu untersuchende Parameter Θ zunächst auf einen Wert $\Theta=\Theta_0$ festgelegt werden.

- Der konkrete Test kann dann erfolgen als:

Punkthypothese: $\Theta=\Theta_0$

Bereichshypothese: $\Theta\geq\Theta_0$, $\Theta\leq\Theta_0$

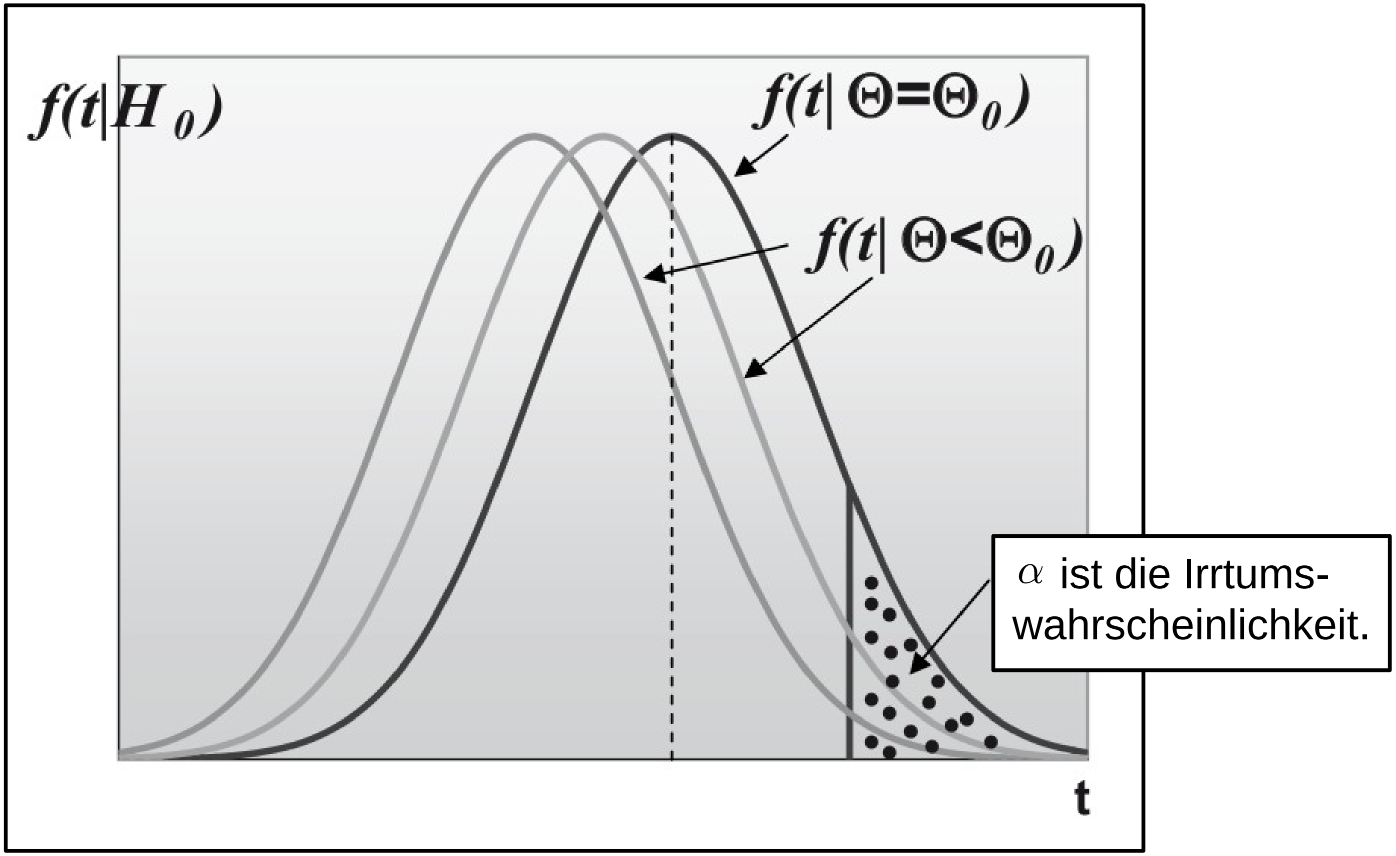


Testveteilung, die bei vorgegebenem Θ_0 abgeleitet werden kann

Der dunkelgraue Bereich kann bei Durchführung des Tests (mit vorgegebener Irrtumswahrscheinlichkeit) ausgeschlossen werden.

Irrtumswahrscheinlichkeit

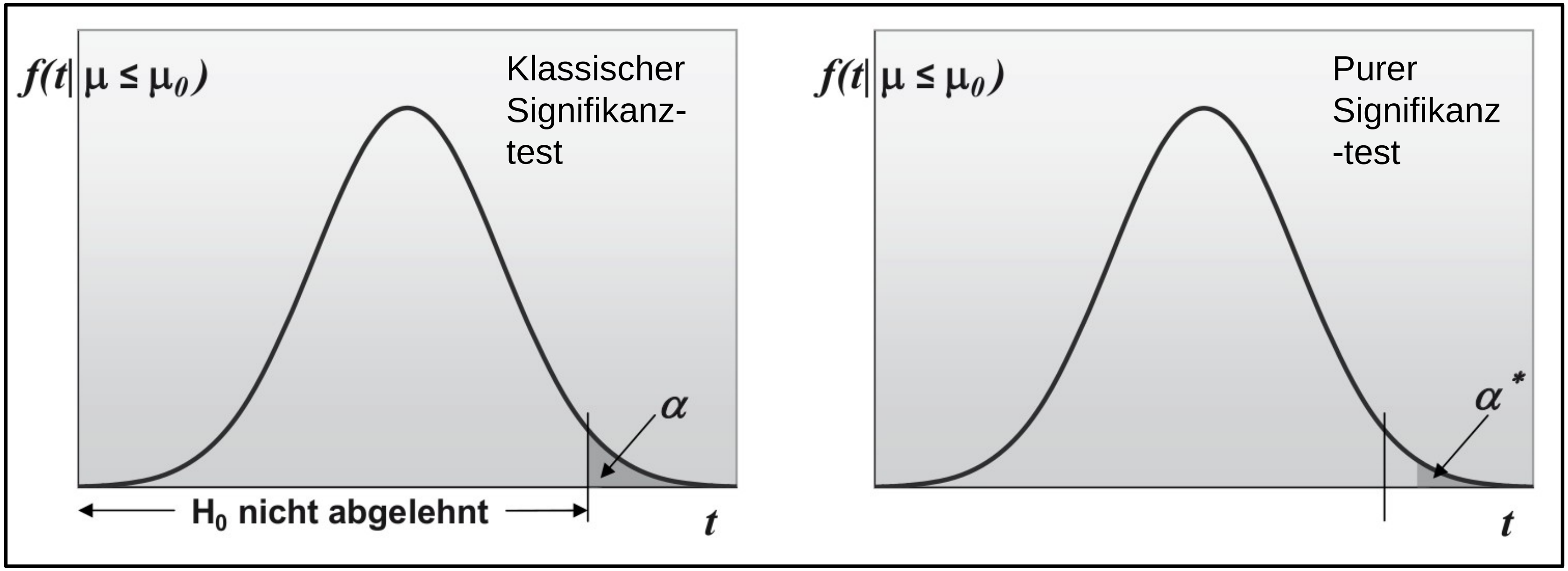
- Am Beispiel einer oberen Schranke:



- Für alle Werte $\Theta \leq \Theta_0$ ist die Irrtumswahrscheinlichkeit kleiner als für den vorgegebenen Wert Θ_0 .

Klassischer vs. purer Signifikanztest

- Beim klassischen Signifikanztest wird das Signifizanzniveau (d.h. der Parameterbereich ab dem H_0 verworfen wird) a priori festgelegt. Nach dem Test fällt die Entscheidung.
- Es gibt jedoch auch Fälle, bei denen der Signifikanzwert angegeben (und die Interpretation so dem Statistikanwender überlassen) wird:



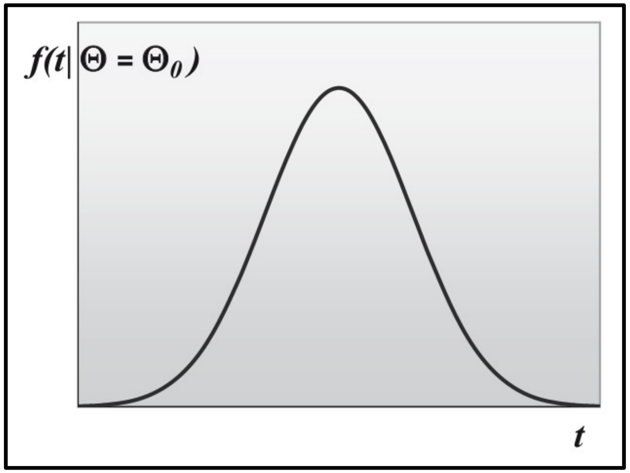
H_0 abgelehnt bei $\alpha = 0.05$

H_0 abgelehnt bei $\alpha^* < \alpha = 0.05$

Priv. Doz. Dr. Roger Wolf
http://ekpwww.physik.uni-karlsruhe.de/~rwolf/

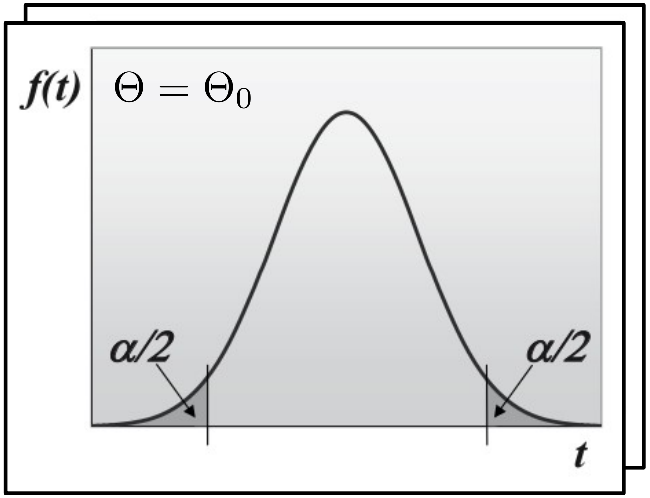
Vorgehensweise (klassischer Signifikanztest)

- Der klassische Signifikanztest besteht aus den folgenden grundlegenden Schritten:
  - Fomulierung der H_0-Hypothese mit $\Theta = \Theta_0$.
  - Wahl einer geeigneten Teststatistik t und Bestimmung der Testverteilung $f(t|\Theta = \Theta_0)$



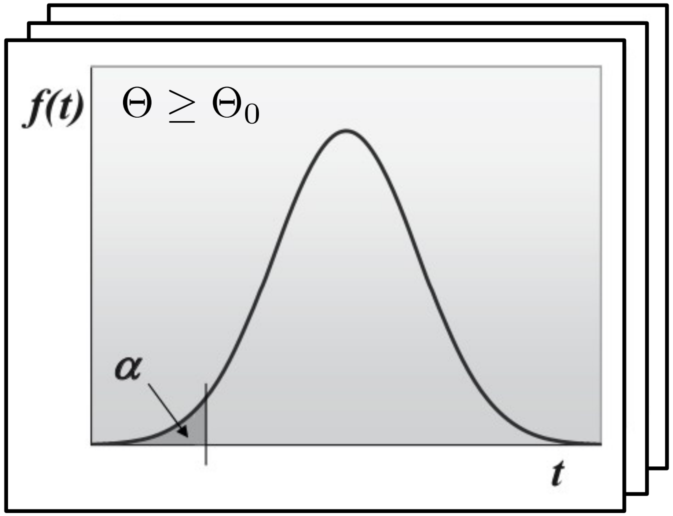
Vorgehensweise (klassischer Signifikanztest)

- Der klassische Signifikanztest besteht aus den folgenden grundlegenden Schritten:
  - Fomulierung der H_0-Hypothese mit $\Theta = \Theta_0$.
  - Wahl einer geeigneten Teststatistik t und Bestimmung der Testverteilung $f(t|\Theta = \Theta_0)$
  - Wahl der Irrtumswahscheinlichkeit α und des Ablehnungsbereichs.



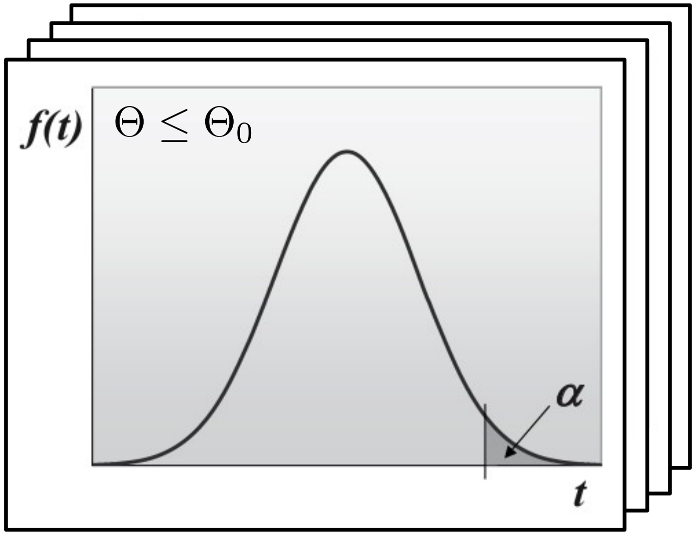
Vorgehensweise (klassischer Signifikanztest)

- Der klassische Signifikanztest besteht aus den folgenden grundlegenden Schritten:
  - Fomulierung der H_0-Hypothese mit $\Theta=\Theta_0$.
  - Wahl einer geeigneten Teststatistik t und Bestimmung der Testverteilung $f(t|\Theta=\Theta_0)$
  - Wahl der Irrtumswahscheinlichkeit α und des Ablehnungsbereichs.



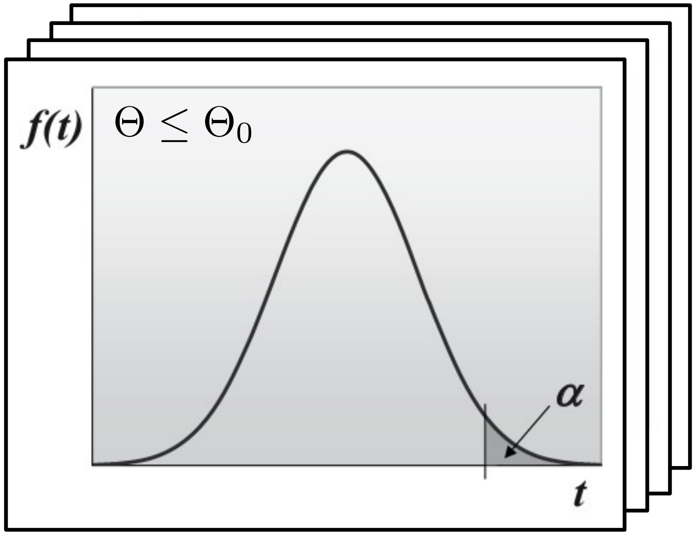
Vorgehensweise (klassischer Signifikanztest)

- Der klassische Signifikanztest besteht aus den folgenden grundlegenden Schritten:
  - Fomulierung der H_0-Hypothese mit $\Theta = \Theta_0$.
  - Wahl einer geeigneten Teststatistik t und Bestimmung der Testverteilung $f(t|\Theta = \Theta_0)$
  - Wahl der Irrtumswahscheinlichkeit α und des Ablehnungsbereichs.



Vorgehensweise (klassischer Signifikanztest)

- Der klassische Signifikanztest besteht aus den folgenden grundlegenden Schritten:
  - Fomulierung der H_0-Hypothese mit $\Theta = \Theta_0$.
  - Wahl einer geeigneten Teststatistik t und Bestimmung der Testverteilung $f(t|\Theta = \Theta_0)$
  - Wahl der Irrtumswahscheinlichkeit α und des Ablehnungsbereichs.
  - Stichprobenziehung (=Messung) und Berechnung des realisierten Wertes der Teststatistik t_{obs} .
  - Testentscheidung und ggf. Ablehnung von H_0 .



Beispiel aus der klassischen Statistik

- Hypothese: Männer in den Niederlanden sind größer sind, als der „Durchschnitts“-Europäer. Das arithmetische Mittel der Körpergrößen in Europa liegt bei $\mu_0 = 174\ \mathrm{cm}$ mit einer Standardabweichung von $\sigma_0 = 10\ \mathrm{cm}$.

  - H_0 : „Niederländer sind gleichgroß oder kleiner als Durchschnitts-Europäer„

    H_0: $\mu \leq \mu_0 = 174\,\mathrm{cm} \quad \mathrm{mit:}\ \sigma = \sigma_0$

  - Wahl einer geeigneten Teststatistik:

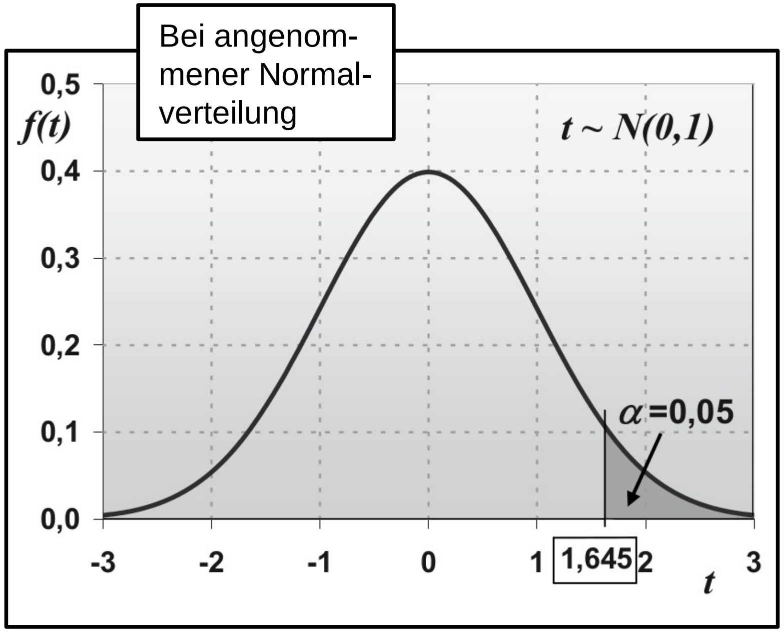
    $$t = \frac{\bar{x} - \mu_0}{\sigma_0/\sqrt{n}} \propto N(0,1)$$

    $\bar{x}$: Mittelwert der Stichprobe

    n : Länge der Stichprobe

    $N(0,1)$: Normalverteilung

  - Festlegung der Irrtumswahrscheinlichkeit:
    $\alpha = 0,05 \quad z = 1,645$ [1]



[1] bei Normalverteilung entspricht z dem Signifikanzniveau in Vielfachen der Standardabweichung (s.d.).

Beispiel aus der klassischen Statistik

- Wir machen drei unabhängige Tests mit wachsendem Umfang:
  $n_1 = 36, \quad n_2 = 100, \quad n_3 = 400$

- Auswertung und Entscheidung:

$\bar{x}_1 = 176$ cm	$\bar{x}_2 = 175$ cm	$\bar{x}_3 = 175$ cm

Beispiel aus der klassischen Statistik

- Wir machen drei unabhängige Tests mit wachsendem Umfang:
  $n_1 = 36, \quad n_2 = 100, \quad n_3 = 400$

- Auswertung und Entscheidung:

$\bar{x}_1 = 176\ \text{cm}$

$$t = \frac{2\,\text{cm}}{10\,\text{cm}/\sqrt{36}} = 1,2 < 1,654$$

$\downarrow$

H_0 kann nicht ausgeschlossen werden

$\bar{x}_2 = 175\ \text{cm}$

$$t = \frac{1\,\text{cm}}{10\,\text{cm}/\sqrt{100}} = 1 < 1,654$$

$\downarrow$

H_0 kann nicht ausgeschlossen werden

$\bar{x}_3 = 175\ \text{cm}$

$$t = \frac{1\,\text{cm}}{10\,\text{cm}/\sqrt{400}} = 2 > 1,654$$

$\downarrow$

H_0 kann ausgeschlossen werden

Beispiel aus der klassischen Statistik

- Wir machen drei unabhängige Tests mit wachsendem Umfang:
  $n_1 = 36, \quad n_2 = 100, \quad n_3 = 400$

- Auswertung und Entscheidung:

$\bar{x}_1 = 176\ \text{cm}$

$$t = \frac{2\,\text{cm}}{10\,\text{cm}/\sqrt{36}} = 1,2 < 1,654$$

$\bar{x}_2 = 175\ \text{cm}$

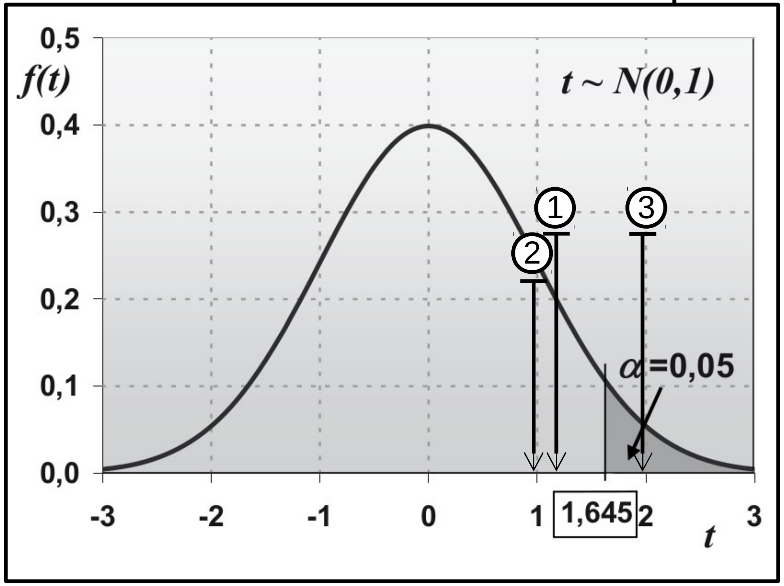
$$t = \frac{1\,\text{cm}}{10\,\text{cm}/\sqrt{100}} = 1 < 1,654$$

H_0
<u>kann nicht</u> ausge-
schlossen werden

$\bar{x}_3 = 175\ \text{cm}$

$$t = \frac{1\,\text{cm}}{10\,\text{cm}/\sqrt{400}} = 2 > 1,654$$

H_0
<u>kann</u> ausge-
schlossen werden



Anmerkungen zu Hypothesentests

- Das Beispiel veranschaulicht die wichtigsten Punkte, die bei der Verwendung von Hypothesentests zu beachten sind:

  - Beim einfachen Hypothesentest ist die Hypothese als „Gegenhypothese“ zu formulieren (im Bsp.: $H_0\colon\ \mu \leq \mu_0 = 174\,\mathrm{cm} \quad \mathrm{mit\colon}\ \sigma = \sigma_0$). Für die Hypothese selbst ist keine Testfunktion angebbar.

  - Nicht-Widerlegung ist logisch schwächer, als die Widerlegung (im Bsp. folgt auf zweimalige Nicht-Widerlegung eine Widerlegung).

  - Die Widerlegung von H_0 hängt von der Vorgabe von α ab. („Wer das Risiko auf sich nimmt mit größerer Wahrscheinlichkeit H_0 irrtümlich zu verwerfen verwirft schneller“.)

  - Die Irrtumswahrscheinlichkeit bildet nur die Wahrscheinlichkeit ab H_0 irrtümlich zu verwerfen. Über die irrtümliche Annahme von H_0 wird zunächst keine Aussage getroffen.

- Ein Hypothesentest erfordert also, eine klare Fragestellung und ein klares Bewußtsein der (meist semantischen) Schwächen und Fehlerquellen.

Fehler 1. & 2. Art

- Für die Beurteilung eines Hypothesentests sind mehrere Irrtumswahrscheinlichkeiten von Relevanz:

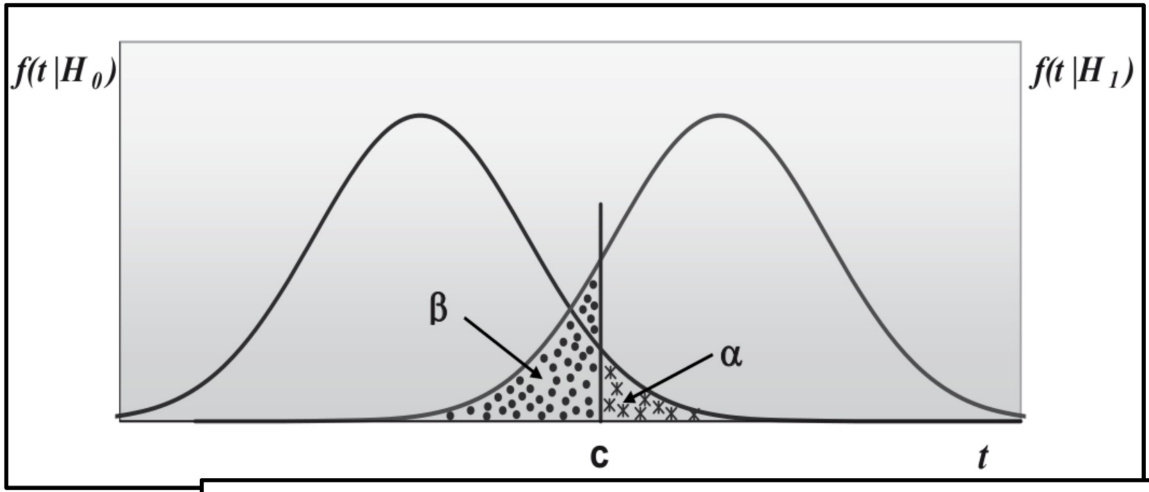
	H_0 angenommen / H_1 abgelehnt	H_1 angenommen / H_0 abgelehnt
H_0 wahr / H_1 falsch	richtige Entscheidung $(1-\alpha)$	Fehler 1. Art, α
H_1 wahr / H_0 falsch	Fehler 2. Art, β	richtige Entscheidung $(1-\beta)$

α : Fehler 1. Art

β : Fehler 2. Art

c : Kritischer Wert

Der a priori festgelegte kritische Wert entscheidet über die Annahme von H_0 oder H_1.



Beispiel: Test zweier alternativer Hypothesen H_0 & H_1 für einen einseitigen Test z.B. $H_0 : \mu \leq \mu_0$ und $H_1 : \mu > \mu_0$.

Operationscharakteristik & Machtfunktion

- Bei vorgegebenem α sind β und c festgelegt.

- Für den häufigen Fall zweier Hypothesen H_0 und H_1 mit den Parametern μ_0 und μ_1 gilt für β ceteris paribus (c.p.), d.h. wenn die jeweils anderen Variablen konstant gehalten werden:

  - $\beta(\alpha)\ (\mathrm{c.p.})$ (je kleiner α desto größer β).

  - $\beta(\mu_1 - \mu_0)\ (\mathrm{c.p.})$ (je größer die Differenz $\mu_1 - \mu_0$ desto kleiner β).

  - $\beta(n)\ (\mathrm{c.p.})$ (je größer n, desto kleiner β).

- Die Abhängigkeit $\beta(\alpha, \mu_1 - \mu_0, n)$ bezeichnet man als **Operationscharakteristik**.

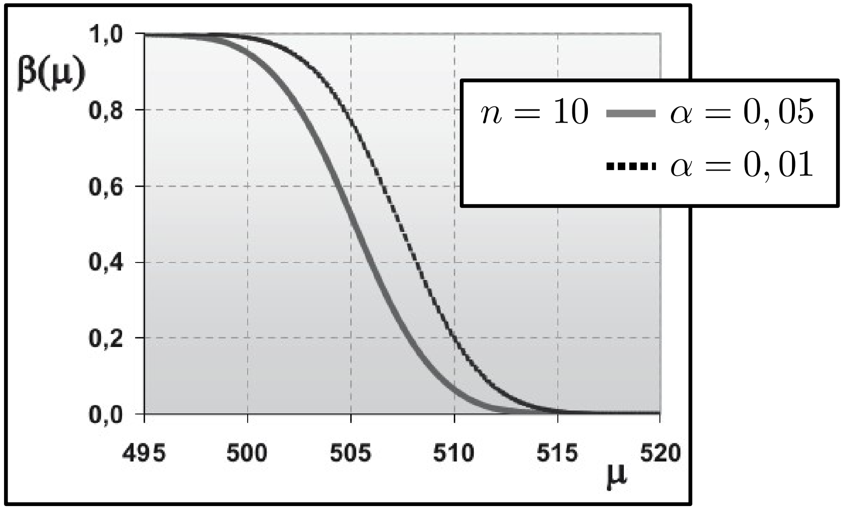
- Die Funktion $1 - \beta(\alpha, \mu_1 - \mu_0, n)$ bezeichnet man als **Ergänzung zu 1**, **Güte-**, oder **Macht-funktion**.

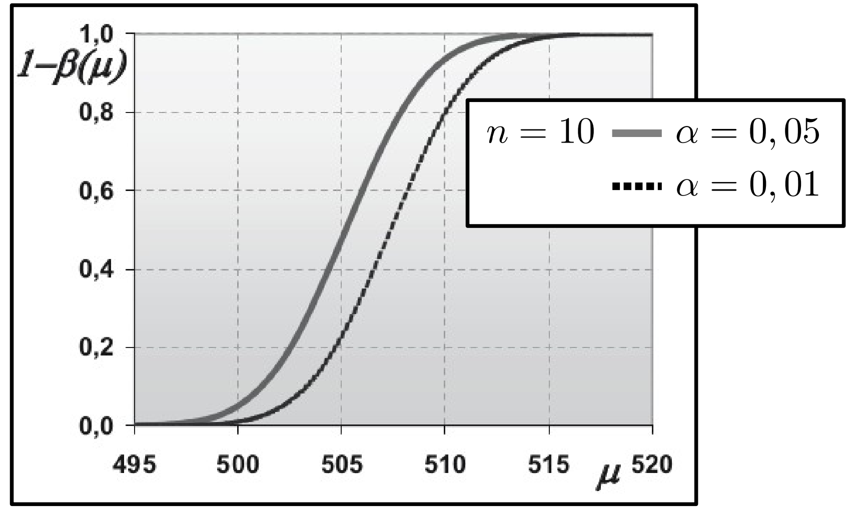
- Ist z.B. α vorgegeben, dann wird man aus allen möglichen Testverfahren dasjenige mit dem kleinsten β (d.h. der größten Macht) auswählen.

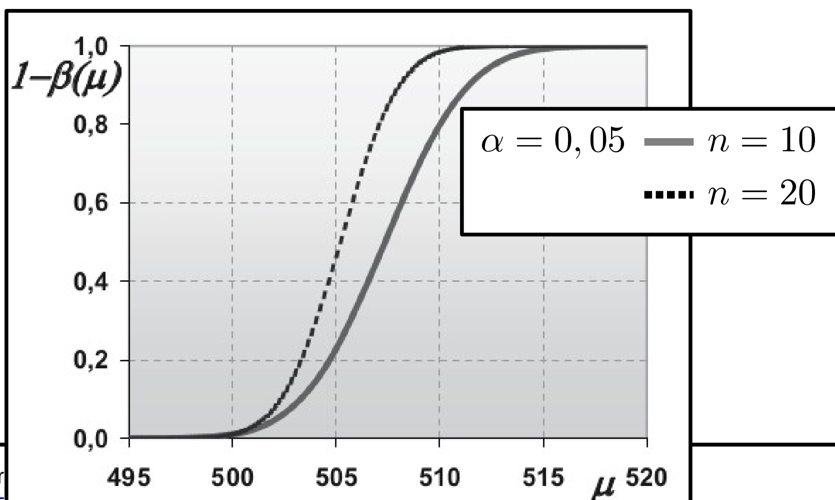
Operationscharakteristik – einseitiger Test –

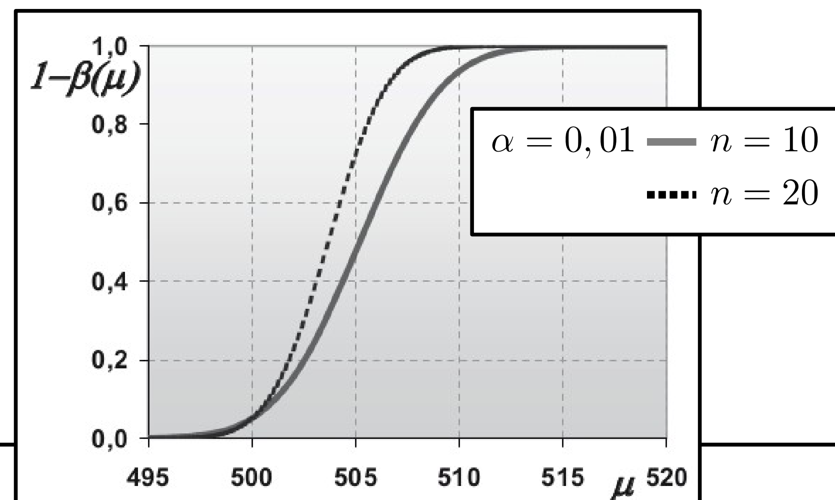
Hier für das Bsp: Füllmenge von Bierflaschen

- Operationscharakteristik und Gütefunktion für $H_0:\ \mu \le \mu_0 = 500\,\mathrm{ml}$ (Füllmenge bei Bierflaschen mit $\sigma_0 = 10\,\mathrm{ml}$):





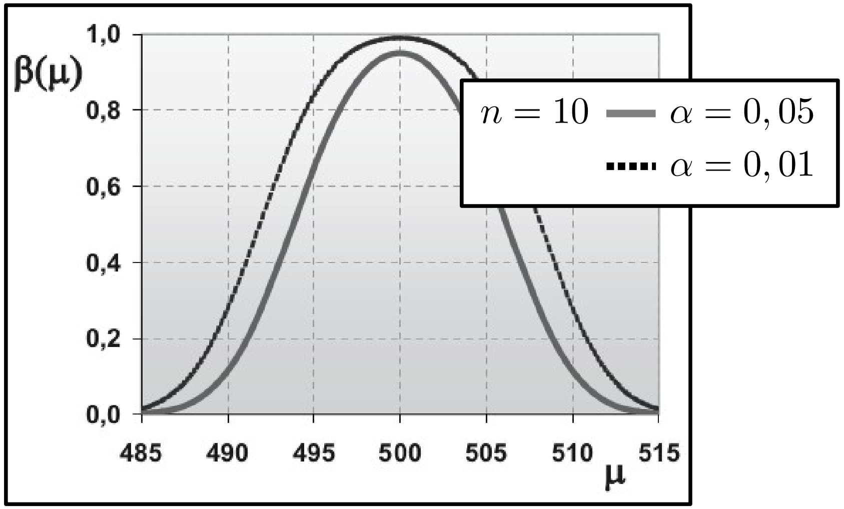


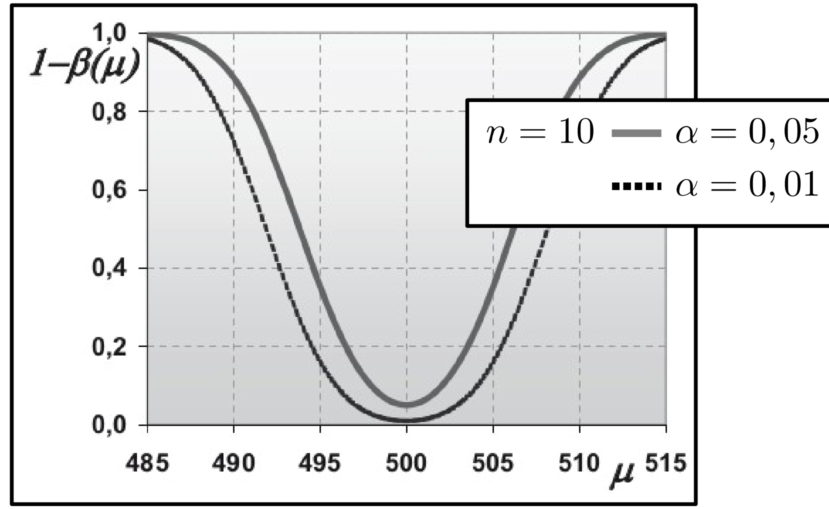


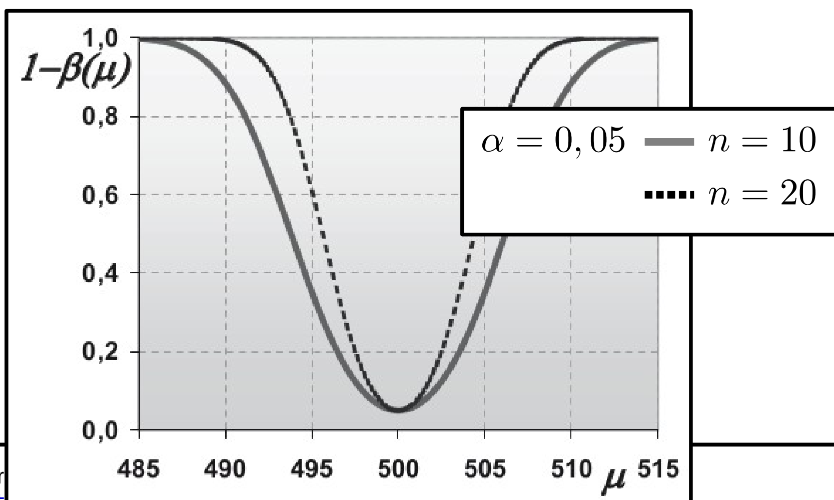
Operationscharakteristik – zweiseitiger Test –

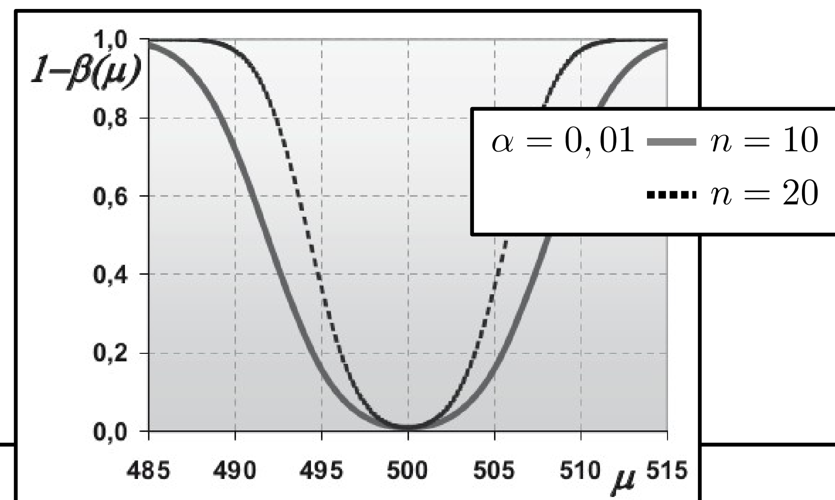
Hier für das Bsp: Füll-menge von Bierflaschen

- Operationscharakteristik und Gütefunktion für $H_0: \ \mu \leq \mu_0 = 500\,\mathrm{ml}$ (Füllmenge bei Bierflaschen mit $\sigma_0 = 10\,\mathrm{ml}$):









Beispiele

- Im folgenden werden wir einige Beispiele für Hypothesentests besprechen anhand derer Sie ein Gefühl für die Vielfalt solcher Tests bekommen können:
  - **Zweistichprobentest**: Haben zwei Stichproben den gleichen Erwartungswert?
  - **Varianzanalyse**: Sind zwei (oder mehr) Merkmalsgruppen im Rahmen ihrer Varianzen kompatibel? – (Erweiterung zu oben.)
  - **Test auf Unabhängigkeit**: Sind zwei Merkmale statistisch unabhängig? Diesen Test finden Sie im backup von Lecture 06.
  - **GoF**: Beachten Sie das alle GoF Tests, die wir in Lecture 06 diskutiert haben Hypothesentests sind. Dies gilt insbesondere für den saturierten Modelltest, den wir nochmal gesondert diskutieren werden.
  - **Maximum Likelihood**: Beachten Sie das ein Maximum Likelihood Scan mit $\Delta\mathrm{NLL}$ nominell ebenfalls einem Zwei-Hypothesentest entspricht.

Beispiel-1: Zweistichprobentest

- **Hypothese**: Hochschulabsolventinnen erhalten bei gleicher Varianz niedrigere Anfangsgehälter (μ_1) als Hochschulabsolventen (μ_2). Die Überprüfung erfolgt durch zwei Zufallsstichproben jeweils vom Umfang $n_1, n_2 = 200$.
  - H_0: „Die Gehälter der Absolventinnen sind gleich oder höher“ ($H_0\colon\ \mu_1 \geq \mu_2$).
  - Geeignete Teststatistik:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \propto N(0,1)$$

  - Irrtumswahrscheinlichkeit: $\alpha = 4,46\%\quad \text{entspricht: } (z_{1-0,0446} = 1,7)$
  - Stichprobenergebnisse: $\bar{x}_1 = 2\,200 \qquad \bar{x}_2 = 2\,300$

$$\sigma_1^2 = 700\,000 \qquad \sigma_2^2 = 920\,000$$

$$\hat{\sigma}^2 = \frac{199 \cdot 700\,000 + 199 \cdot 920\,000}{398} = 810\,000$$

$$t = \frac{|2\,200 - 2\,300|}{900 \cdot \frac{1}{10}} = \frac{100}{90} = 1,11 < 1,7$$

d.h. H_0 kann durch diese Stichprobe nicht widerlegt werden.

Beispiel-2: Varianzanalyse

- **Hypothese**: In einer Untersuchung zur Verkehrssituation in der Innenstadt geben durch Zufallsstichproben ausgewählte Befragte ein Punkturteil ab (je höher, desto positiver). Die Untersuchung trennt zwischen Innenstadtbewohnern ($\mu_1, n_1 = 36$) und Pendlern ($\mu_2, n_2 = 41$).
  - H_0 : „Beide Gruppen beurteilen die Verkehrsituation gleich“ ($H_0\colon\ \mu_1 = \mu_2$).
  - Wir diskutieren zwei Teststatistiken:

$$\#\ \text{Befragte:}\ n = n_1 + n_2 = 77 \quad \#\ \text{Gruppen:}\ r = 2$$

$$\text{a)} \quad t = \frac{\frac{1}{r-1}\sigma^2_{\text{ext}}}{\frac{1}{n-r}\sigma^2_{\text{int}}} = \frac{\frac{1}{r-1}\sum\limits_{g=1}^{r} n_g\left(\bar{x}_g - \bar{x}\right)^2}{\frac{1}{n-r}\sum\limits_{g=1}^{r}\sum\limits_{i=1}^{n_g}\left(x_{gi} - \bar{x}_g\right)^2} \propto f_{1-\alpha}(r-1, n-r)$$

$$\sigma^2_{\text{ext}}\ : \text{Varianz zwischen den Gruppen}$$

$$\sigma^2_{\text{int}}\ : \text{Varianz innerhalb der Gruppen}$$

$$\text{b)} \quad t = \frac{|\bar{x}_1 - \bar{x}_2|}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \propto N(0,1)$$

Beispiel-2: Varianzanalyse

- **Hypothese**: In einer Untersuchung zur Verkehrssituation in der Innenstadt geben durch Zufallsstichproben ausgewählte Befragte ein Punkturteil ab (je höher, desto positiver). Die Untersuchung trennt zwischen Innenstadtbewohnern ($\mu_1, n_1 = 36$) und Pendlern ($\mu_2, n_2 = 41$).

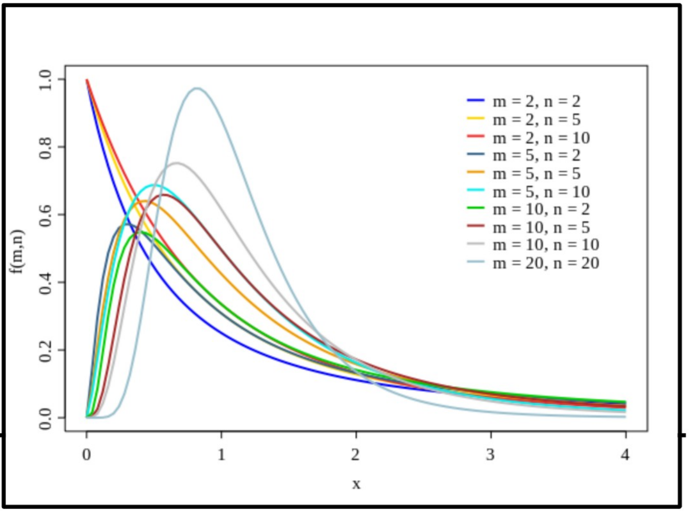
  - H_0 : „Beide Gruppen beurteilen die Verkehrsituation gleich“ ($H_0\colon\ \mu_1 = \mu_2$).

  - Wir diskutieren zwei Teststatistiken:

$$\#\ \text{Befragte:}\ n = n_1 + n_2 = 77 \quad \#\ \text{Gruppen:}\ r = 2$$

$$\text{a)}\quad t = \frac{\frac{1}{r-1}\sigma^2_{\text{ext}}}{\frac{1}{n-r}\sigma^2_{\text{int}}} = \frac{\frac{1}{r-1}\sum\limits_{g=1}^{r} n_g\left(\bar{x}_g - \bar{x}\right)^2}{\frac{1}{n-r}\sum\limits_{g=1}^{r}\sum\limits_{i=1}^{n_g}\left(x_{gi} - \bar{x}_g\right)^2} \propto \boxed{f_{1-\alpha}(r-1, n-r)}$$

Diese Teststatistik folgt der Fisher-Verteilung.



$$\sigma^2_{\text{ext}}\ :\ \text{Varianz zwischen den Gruppen}$$
$$\sigma^2_{\text{int}}\ :\ \text{Varianz innerhalb der Gruppen}$$

$$\text{b)}\quad t = \frac{|\bar{x}_1 - \bar{x}_2|}{\hat{\sigma}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \propto N(0,1)$$

Beispiel-2: Varianzanalyse

- Irrtumswahscheinlichkeit: $\alpha = 0.05 \quad f_{0.95}(1,75) = 3,968 \quad z_{1-0.0025} = 1,96$

- Stichprobenergebnisse:

$\bar{x}_1 = 50 \qquad \bar{x}_2 = 48,4 \qquad \bar{x} = 49,15$

$\sigma_1^2 \;=\; 8,4 \qquad \sigma_2^2 \;=\; 7,5$

a) $$\sum_{g=1}^{2} n_g \left(\bar{x}_g - \bar{x}\right)^2 = 36 \cdot (50 - 49,15)^2 + 41 \cdot (48,4 - 49,15)^2 = 49,0725$$

$$\sum_{g=1}^{2} \sum_{i=1}^{n_g} \left(x_{gi} - \bar{x}_g\right)^2 = \sum_{g=1}^{2} \left(n_g - 1\right) \sigma_g^2 = 35 \cdot 8,4 + 40 \cdot 7,5 = 594$$

$$t = \frac{49,0725}{\frac{1}{75} \cdot 594} = 6,196 (> 3,968)$$

- Testentscheidung: H_0 ist widerlegt ($\alpha^* = 0,015$), die Verkehrssituation wird unterschiedlich bewertet.

Beispiel-2: Varianzanalyse

- Irrtumswahscheinlichkeit: $\alpha = 0.05 \quad f_{0.95}(1,75) = 3,968 \quad z_{1-0.0025} = 1,96$

- Stichprobenergebnisse:

$\bar{x}_1 = 50 \qquad \bar{x}_2 = 48,4 \qquad \bar{x} = 49,15$

$\sigma_1^2 \;=\; 8,4 \qquad \sigma_2^2 \;=\; 7,5$

$$\text{b)} \quad \hat{\sigma}^2 = \frac{35 \cdot 8,4 + 40 \cdot 7,5}{75} = 7,92$$

$$t = \frac{50 - 48,4}{\sqrt{7,92} \cdot \left(\frac{1}{36} + \frac{1}{41}\right)} = 2,489 (> 1,96)$$

- Testentscheidung: H_0 ist widerlegt ($\alpha^* = 0,013$), die Verkehrssituation wird unterschiedlich bewertet.

Backup