Rechnernutzung in der Physik

Karlsruher Institut für Technologie
Wintersemester 2020/21

Teil – 1 “Computerbasierte Methoden der Statistik”

Roger Wolf (roger.wolf@cern.ch), Ralf Ulrich (ralf.ulrich@kit.edu)

Inhalt

Roger Wolf

Ralf Ulrich

Literaturempfehlungen

- **Einführende Literatur zu Statistik und Numerik:**
  - G. Cowan, *Statistical data analysis*, Oxford (1997) (KIT-Bibliothek).
  - G. Bohm, G. Zech, *Einführung in Statistik und Messwertanalyse für Physiker*, DESY (2006) (eBook deutsch, eBook english).
  - V. Blobel, E. Lormann, *Statistische und numerische Methoden der Datenanalyse*, DESY (2012) (Webseite).
  - R. J. Barlow, *Statistics: A Guide to the use of statistical methods in the physical sciences*, Wiley (1989) (KIT-Bibliothek).
  - W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical recipes*, Cambridge Univ. Press (2007) (Webseite).

- Skriptensammlung von Prof. G. Quast (Link):

6 Hypothesentests in der modernen Physik

6.1 Grundlagen und Begriffe

Wir klären zunächst grundsätzlich, worum es bei einem Hypothesentest geht und fassen alle wesentlichen Aspekte an einem Beispiel zusammen.

Hypothesentests

- Was eine statistische Hypothese haben wir bereits in VL-03 Folie 12 geklärt.

- Hypothesentests dienen dazu, eine Hypothese mit gegebener Stichprobenverteilung einer Teststatistik $t(\vec{x})$ auf ihre Vereinbarkeit mit dem Ergebnis einer Stichprobe (=Messung) hin zu überprüfen.

- Die zu testende Hypothese wird im Allgemeinen als Nullhypothese (H_0) bezeichnet. 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 (d.h. die Varianz von $t(\vec{x})$) erklärt werden kann.

- **NB**: 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. $H_0\colon\ \mu_1 = \mu_2$)?

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. $H_0\colon\ \mu_1 = \mu_2$)?

  - Entspricht die Energiedeposition des Teilchens in Blei eher der Erwartung für ein Pion oder für ein Elektron?

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. $H_0\colon\ \mu_1 = \mu_2$)?

  - Entspricht die Energiedeposition des Teilchens in Blei eher der Erwartung für ein Pion oder für ein Elektron?

  - Gibt es ein neues Signal auf einem bekannten Untergrund?

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. $H_0\colon\ \mu_1 = \mu_2$)?
  - Entspricht die Energiedeposition des Teilchens in Blei eher der Erwartung für ein Pion oder für ein Elektron?
  - Gibt es ein neues Signal auf einem bekannten Untergrund?
  - Ist der Spin des beobachteten Teilchens 0 oder 1?

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. $H_0\colon\ \mu_1 = \mu_2$)?

  - Entspricht die Energiedeposition des Teilchens in Blei eher der Erwartung für ein Pion oder für ein Elektron?

  - Gibt es ein neues Signal auf einem bekannten Untergrund?

  - Ist der Spin des beobachteten Teilchens 0 oder 1?

  - Verdienen Frauen weniger als Männer bei gleicher Qualifikation?

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. $H_0\colon\ \mu_1 = \mu_2$)?

  - Entspricht die Energiedeposition des Teilchens in Blei eher der Erwartung für ein Pion oder für ein Elektron?

  - Gibt es ein neues Signal auf einem bekannten Untergrund?

  - Ist der Spin des beobachteten Teilchens 0 oder 1?

  - Verdienen Frauen weniger als Männer bei gleicher Qualifikation?

  - Ist das Medikament wirksam oder nicht?

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. $H_0\colon\ \mu_1 = \mu_2$)?
  - Entspricht die Energiedeposition des Teilchens in Blei eher der Erwartung für ein Pion oder für ein Elektron?
  - Gibt es ein neues Signal auf einem bekannten Untergrund?
  - 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.

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.

- Die Testverteilung wird unter Annahme von H_0 entweder analytisch oder mit Hilfe eines Ensemble Tests ermittelt.

Zurück zu Folie 7
Zurück zu Folie 29

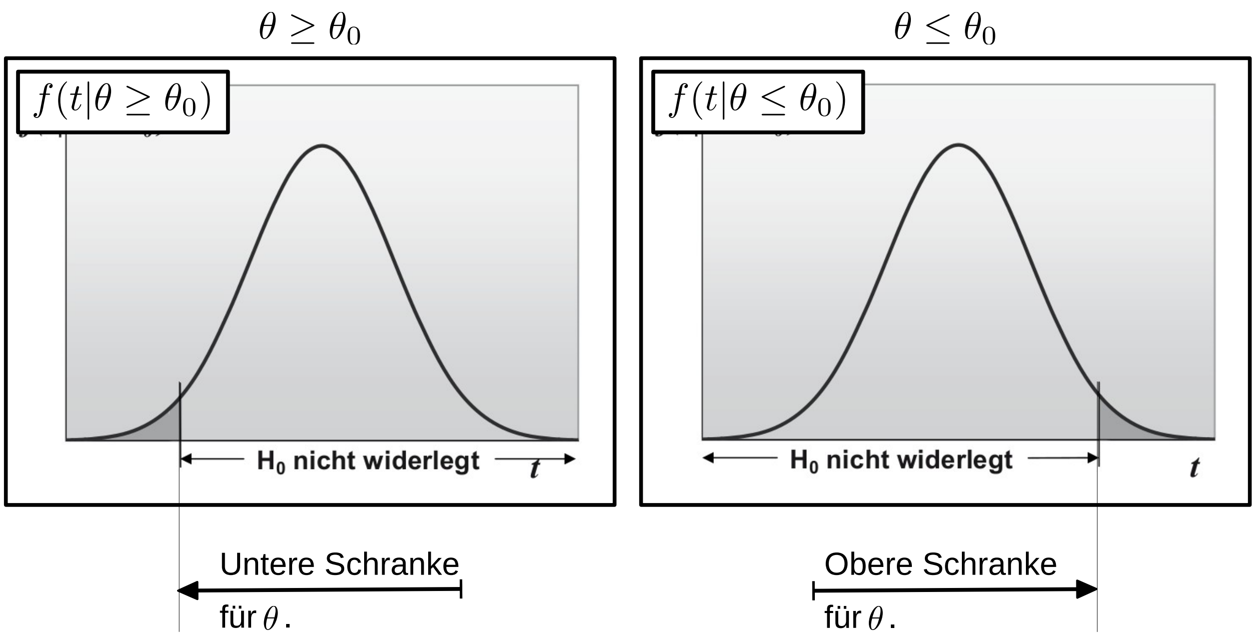
Testverteilung

- 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.

- Die Testverteilung wird unter Annahme von H_0 entweder analytisch oder mit Hilfe eines Ensemble Tests ermittelt.

- Der konkrete Test erfolgt dann als (als einmalige Auswertung von $t(\vec{x})$) auf einer Stichprobe, entweder als:

Bereichshypothese:



Testverteilung

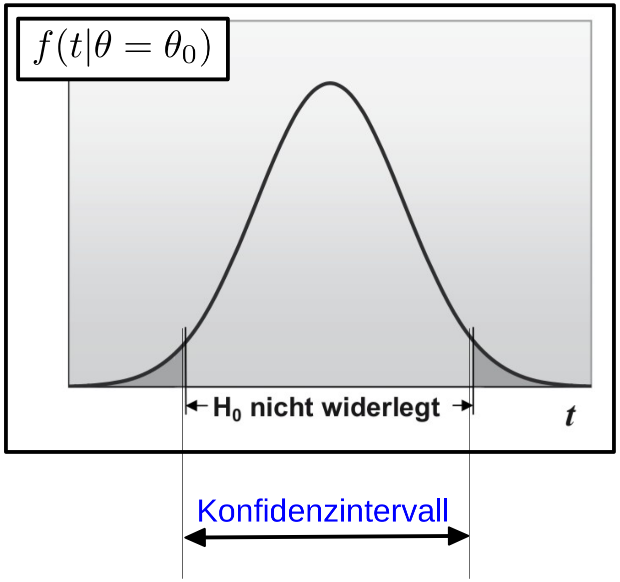
- Um eine Testverteilung $f(t|\theta=\theta_0)$ angeben zu können, muss bei Parametertests der zu untersuchende Parameter θ zunächst auf einen Wert $\theta=\theta_0$ festgelegt werden.

- Die Testverteilung wird unter Annahme von H_0 entweder analytisch oder mit Hilfe eines Ensemble Tests ermittelt.

- Der konkrete Test erfolgt dann als (als einmalige Auswertung von $t(\vec{x})$) auf einer Stichprobe, oder als:

Punkthypothese:

$$\theta=\theta_0$$



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.

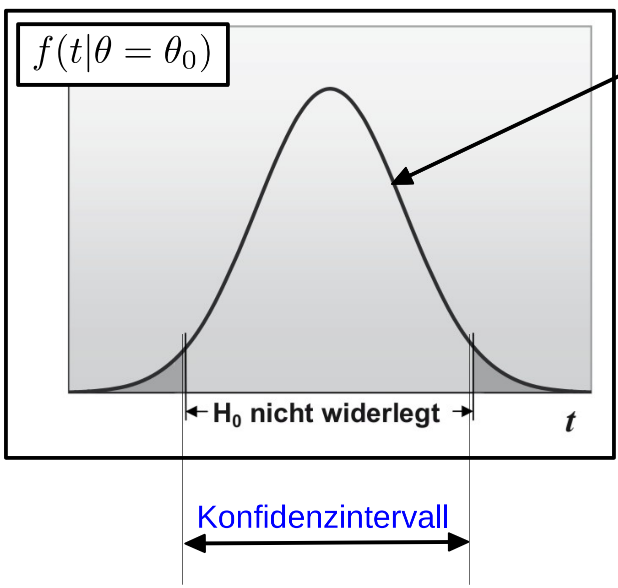
- Die Testverteilung wird unter Annahme von H_0 entweder analytisch oder mit Hilfe eines Ensemble Tests ermittelt.

- Der konkrete Test erfolgt dann als (als einmalige Auswertung von $t(\vec{x})$) auf einer Stichprobe, oder als:

Punkthypothese:

$\theta=\theta_0$

Testverteilung, die bei vorgegebenem θ_0 abgeleitet werden kann



Testverteilung

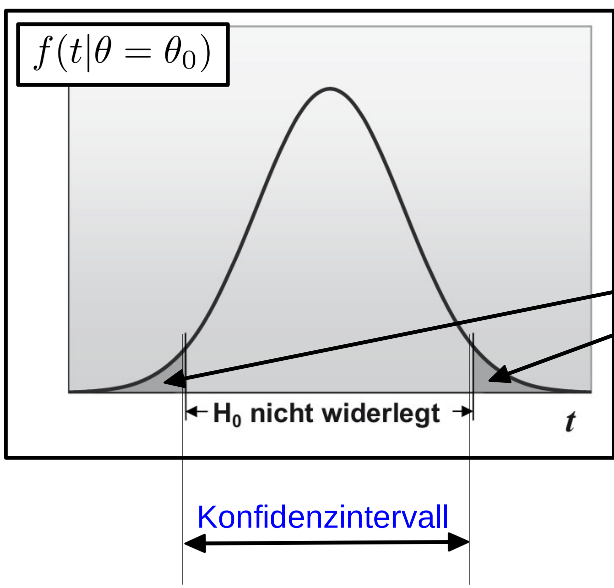
- Um eine Testverteilung $f(t|\theta=\theta_0)$ angeben zu können, muss bei Parametertests der zu untersuchende Parameter θ zunächst auf einen Wert $\theta=\theta_0$ festgelegt werden.

- Die Testverteilung wird unter Annahme von H_0 entweder analytisch oder mit Hilfe eines Ensemble Tests ermittelt.

- Der konkrete Test erfolgt dann als (als einmalige Auswertung von $t(\vec{x})$) auf einer Stichprobe, oder als:

Punkthypothese:

$$\theta = \theta_0$$



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

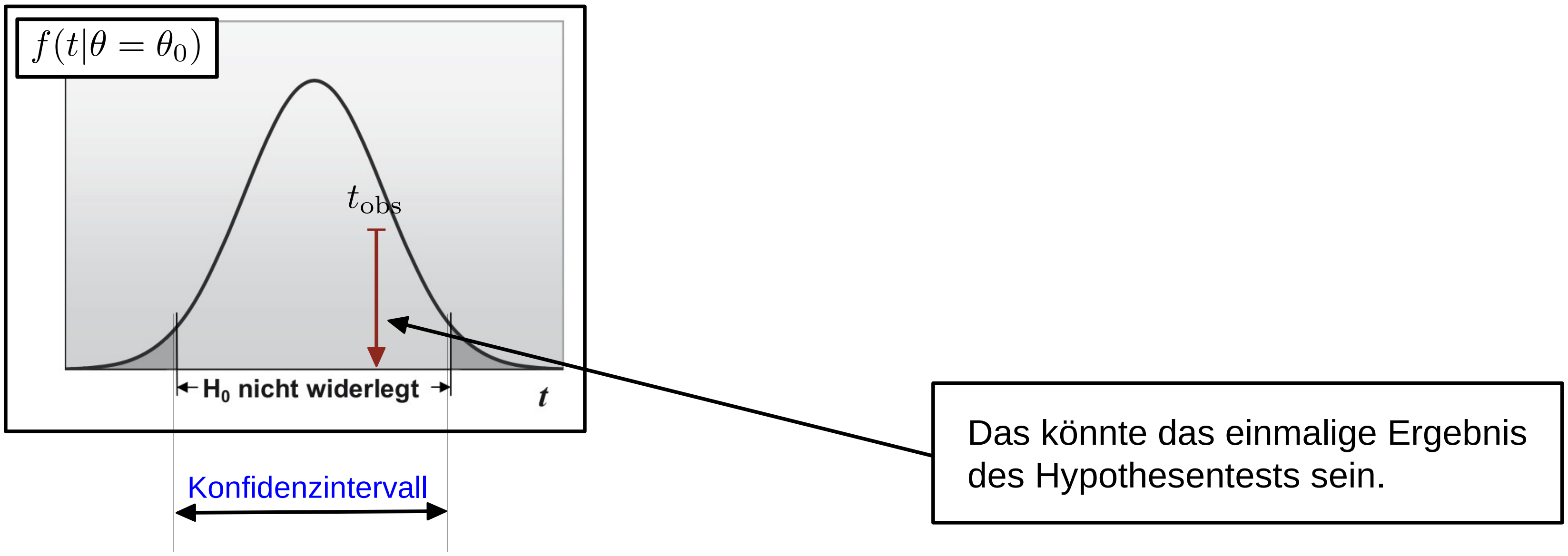
Zurück zu Folie 24

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.
- Die Testverteilung wird unter Annahme von H_0 entweder analytisch oder mit Hilfe eines Ensemble Tests ermittelt.
- Der konkrete Test erfolgt dann als (als einmalige Auswertung von $t(\vec{x})$) auf einer Stichprobe, oder als:

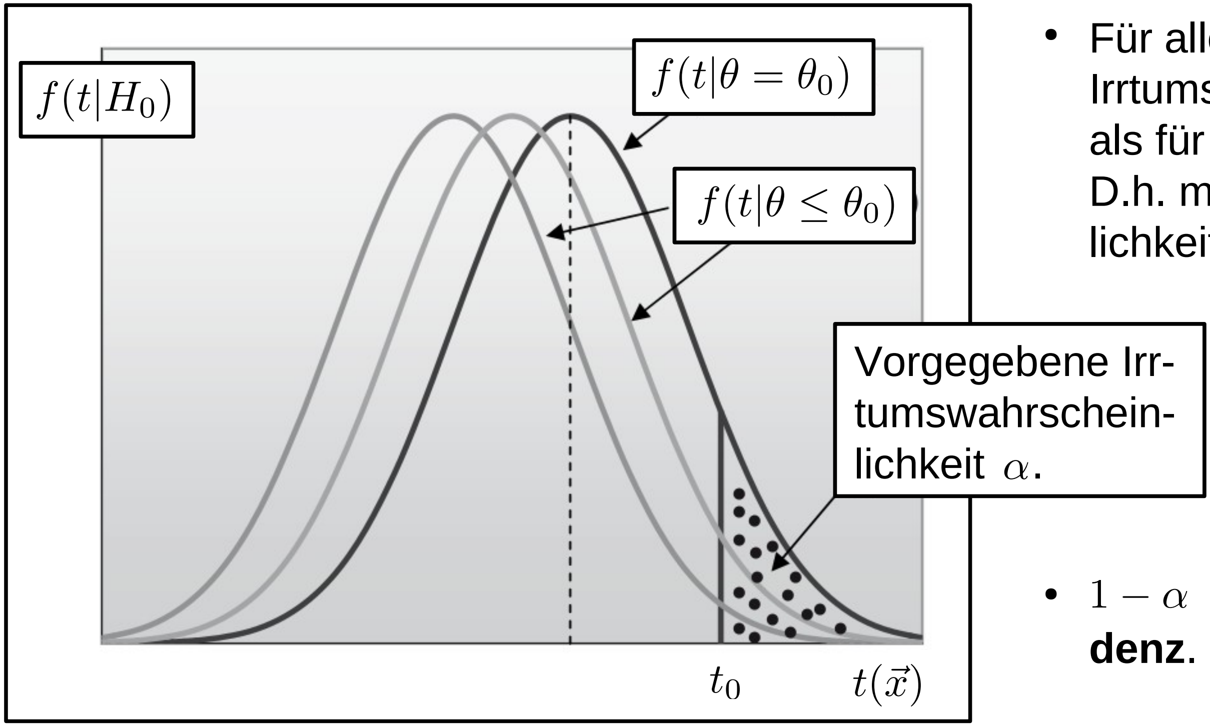
Punkthypothese:

$$\theta=\theta_0$$



Irrtumswahrscheinlichkeit und Konfidenz

- Die Irrtumswahrscheinlichkeit wird *vor* der Anwendung des Tests festgelegt und definiert ab welchem Wert von $t(\vec{x})$ der Test verworfen wird. Hier am Beispiel einer oberen Schranke:



- Für alle Werte $\theta\leq\theta_0$ ist die Irrtumswahrscheinlichkeit kleiner als für den vorgegebenen Wert θ_0. D.h. mit einer Irrtumswahrscheinlichkeit $\leq\alpha$ ist $\theta\leq\theta_0$.

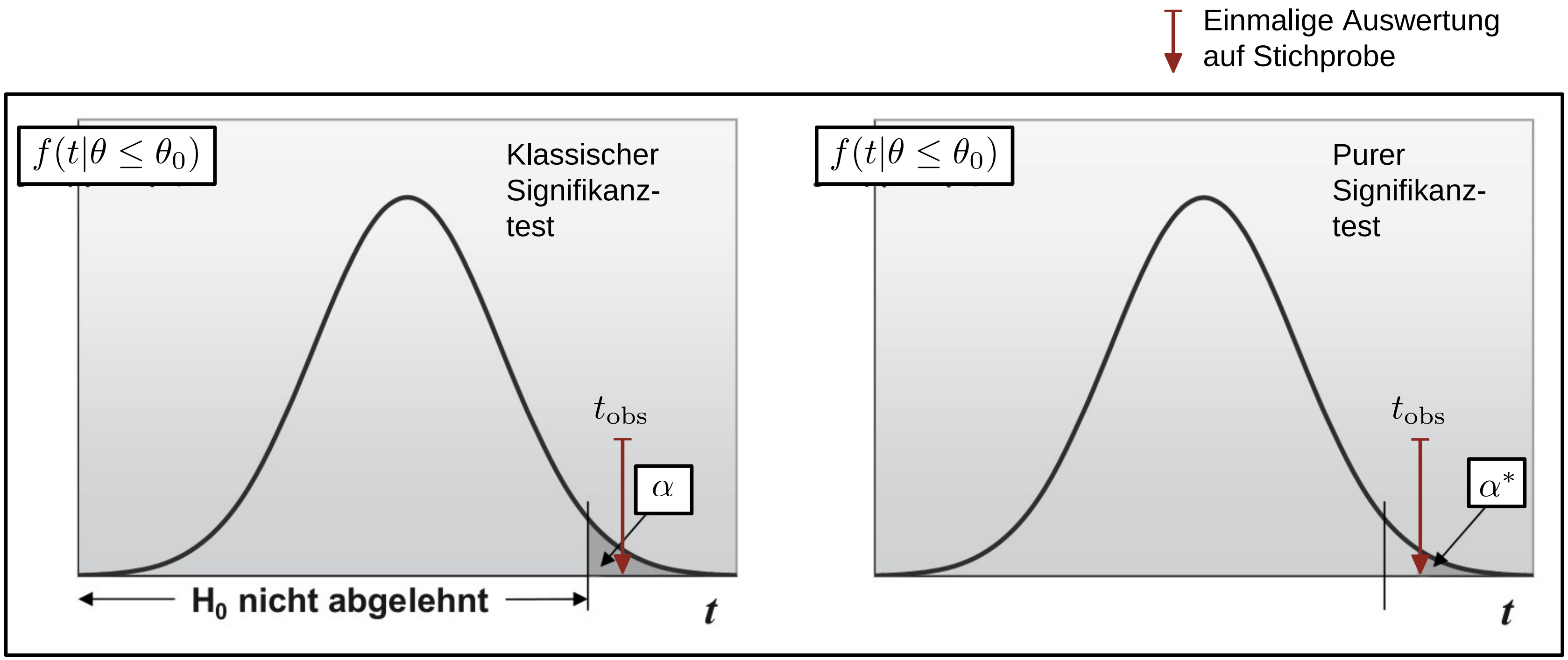
- $1-\alpha$ bezeichnet man als **Konfidenz**.

- In der frequentistischen Wahrscheinlichkeitsinterpretation entspricht ein solcher Scan von θ der Neymankonstruktion.

Klassischer vs. purer Signifikanztest

- Beim klassischen Signifikanztest wird das Signifikanzniveau (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 (zusätzlich) angegeben wird und die Interpretation so dem Statistikanwender überlassen bleibt:



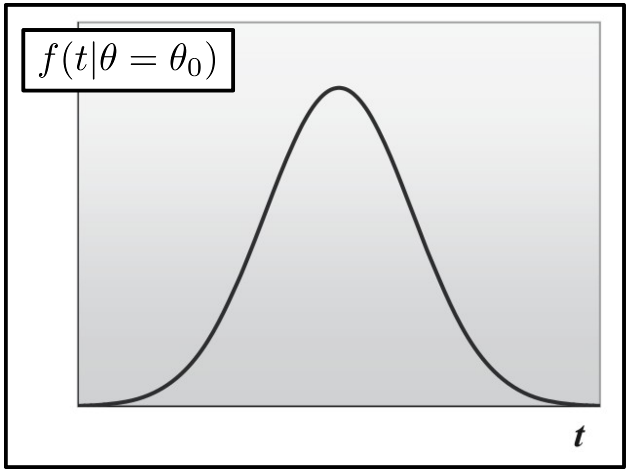
H_0 abgelehnt bei $\alpha = 0.05$ $\quad$ H_0 abgelehnt bei $\alpha^* < \alpha = 0.05$

Zusammenfassung: Klassischer Signifikanztest

- Der klassische Signifikanztest besteht aus den folgenden grundlegenden Schritten:
  - Fomulierung der H_0-Hypothese mit $\theta = \theta_0$.

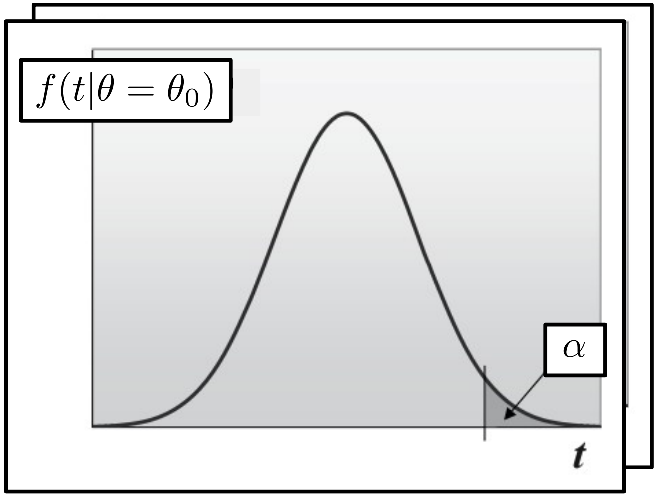
Zusammenfassung: Klassischer Signifikanztest

- 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 $f(t|\theta = \theta_0)$.



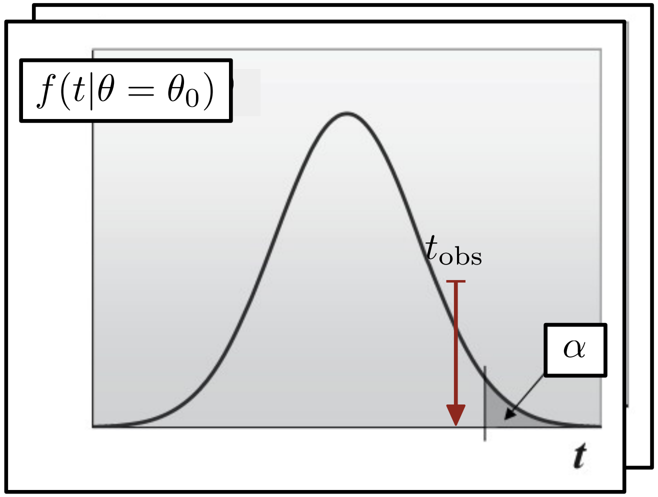
Zusammenfassung: Klassischer Signifikanztest

- 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 (hier am Bsp. der oberen Schranke).



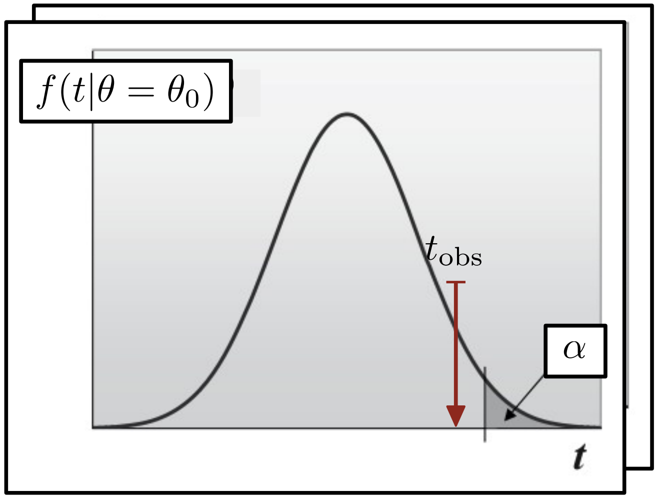
Zusammenfassung: Klassischer Signifikanztest

- 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 (hier am Bsp. der oberen Schranke).
  - Stichprobenziehung (=Messung) und Berechnung des realisierten Wertes der Teststatistik t_{obs} .



Zusammenfassung: 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 (hier am Bsp. der oberen Schranke).
  - Stichprobenziehung (=Messung) und Berechnung des realisierten Wertes der Teststatistik t_{obs} .
  - Testentscheidung und ggf. Ablehnung von H_0 .



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

Beispiel aus der klassischen Statistik

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

Beispiel aus der klassischen Statistik

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

  - $H_0\colon\ \mu \leq \mu_0 = 174\,\mathrm{cm} \quad \mathrm{mit:}\ \sigma = \sigma_0$ (d.h. „Niederländer sind gleichgroß oder kleiner als Durchschnitts-Europäer")

Beispiel aus der klassischen Statistik

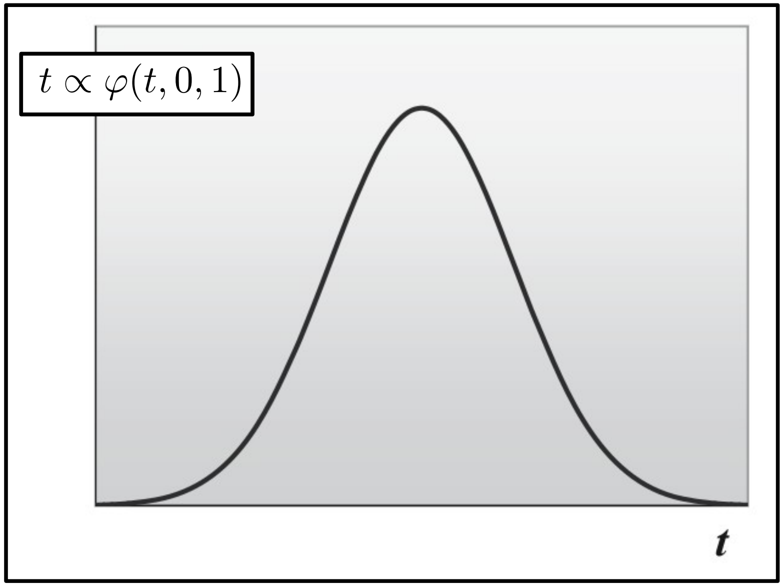
- **Hypothese**: Männer in den Niederlanden sind größer als der „Durchschnitts“-Europäer. Das arithmetische Mittel der Körpergrößen in Europa liegt bei $\mu_0 = 174\ \mathrm{cm}$ mit einer Standardabweichung von $\sigma_0 = \pm 10\ \mathrm{cm}$.

  - H_0: $\mu \leq \mu_0 = 174\,\mathrm{cm} \quad \mathrm{mit:}\ \sigma = \sigma_0$ (d.h. „Niederländer sind gleichgroß oder kleiner als Durchschnitts-Europäer“)

  - Teststatistik:

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

$$\begin{array}{ll} \mu = \bar{x} & : \mathrm{Mittelwert\ der\ Stichprobe} \\ n & : \mathrm{Länge\ der\ Stichprobe} \\ \varphi(\mu, 0, 1) & : \mathrm{Normalverteilung} \end{array}$$



Beispiel aus der klassischen Statistik

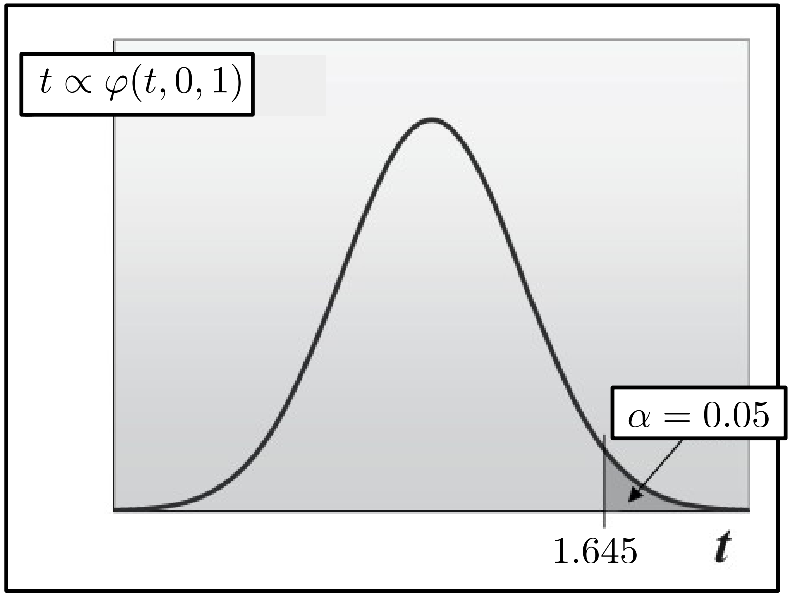
- **Hypothese**: Männer in den Niederlanden sind größer als der „Durchschnitts"-Europäer. Das arithmetische Mittel der Körpergrößen in Europa liegt bei $\mu_0 = 174\ \mathrm{cm}$ mit einer Standardabweichung von $\sigma_0 = \pm 10\ \mathrm{cm}$.

  - H_0: $\mu \leq \mu_0 = 174\,\mathrm{cm} \quad \mathrm{mit:}\ \sigma = \sigma_0$ (d.h. „Niederländer sind gleichgroß oder kleiner als Durchschnitts-Europäer")

  - Teststatistik:

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

$$\begin{aligned} &\mu = \bar{x} && : \text{Mittelwert der Stichprobe} \\ &n && : \text{Länge der Stichprobe} \\ &\varphi(\mu, 0, 1) && : \text{Normalverteilung} \end{aligned}$$



  - Irrtumswahrscheinlichkeit:
    $\alpha = 0,05 \quad t = 1.645$ [1]

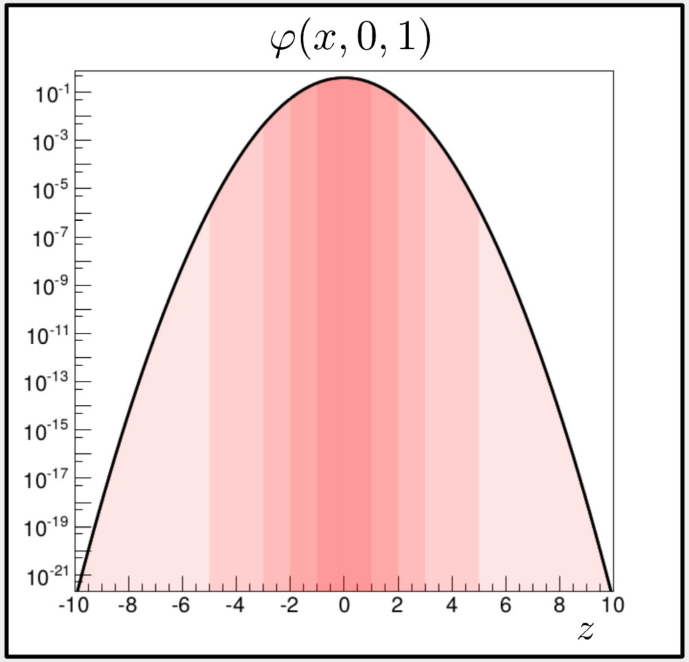
[1] bei Standardnormalverteilung entspricht $t = 1.645$ dem Signifikanz-niveau für $\alpha = 0.05$.

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

Irrtumswahrscheinlichkeiten & Signifikanzniveaus

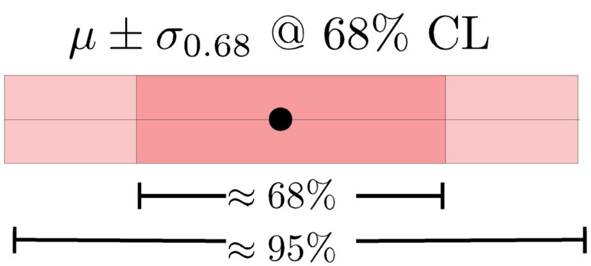
Zurück zu Folie 21
Zurück zu Folie 29
Zurück zu Folie 34
Zurück zu Folie 56

- Als Referenz für die weitere Vorlesung finden Sie hier eine Zuordnung regelmäßig verwendeter Irrtumswahrscheinlichkeiten α zu Signifikanzniveaus der Standardnormalverteilung $\varphi(x, 0, 1)$.



α	Signifikanzniveau z Einseitig	Zweiseitig
0.317	0.782	1
0.1	1.285	1.645
0.05	1.645	1.960
0.045	1.695	2
0.010	2.325	2.580
0.003	–	3
5.7×10^{-7}	–	5

NB: Die $\pm 1\sigma$ und $\pm 2\sigma$ Fehlerbalken, die Sie kennen entsprechen den zweiseitigen Signifikanzniveaus von $z = 1, 2$.



Beispiel aus der klassischen Statistik

- Es wurden drei unabhängige Tests mit wachsendem Umfang durchgeführt:

- Auswertung und Entscheidung:

$\bar{x}_1 = 176\ \mathrm{cm}\,;\quad \mathrm{n}_1 = 36$ $\bar{x}_2 = 175\ \mathrm{cm}\,;\quad \mathrm{n}_2 = 100$ $\bar{x}_3 = 175\ \mathrm{cm}\,;\quad \mathrm{n}_3 = 400$

Beispiel aus der klassischen Statistik

- Es wurden drei unabhängige Tests mit wachsendem Umfang durchgeführt:

- Auswertung und Entscheidung:

$\bar{x}_1 = 176 \text{ cm}\,; \quad \mathrm{n}_1 = 36$

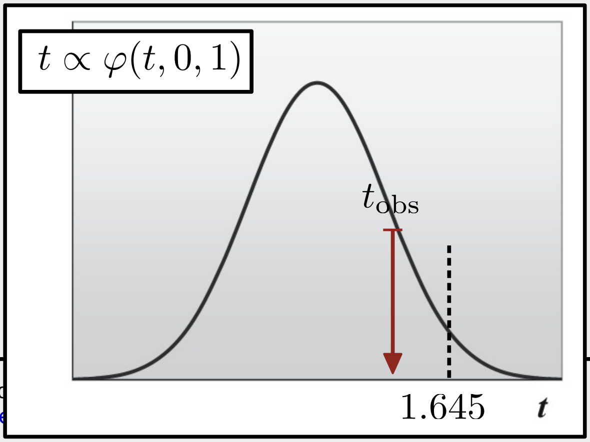
$\bar{x}_2 = 175 \text{ cm}\,; \quad \mathrm{n}_2 = 100$

$\bar{x}_3 = 175 \text{ cm}\,; \quad \mathrm{n}_3 = 400$

$$t = \frac{2\,\mathrm{cm}}{10\,\mathrm{cm}/\sqrt{36}} = 1.2 < 1.654$$

H_0

kann *nicht* ausge-
schlossen werden



Beispiel aus der klassischen Statistik

- Es wurden drei unabhängige Tests mit wachsendem Umfang durchgeführt:

- Auswertung und Entscheidung:

$\bar{x}_1 = 176\ \text{cm}\,; \quad \mathrm{n}_1 = 36$

$$t = \frac{2\,\text{cm}}{10\,\text{cm}/\sqrt{36}} = 1.2 < 1.654$$

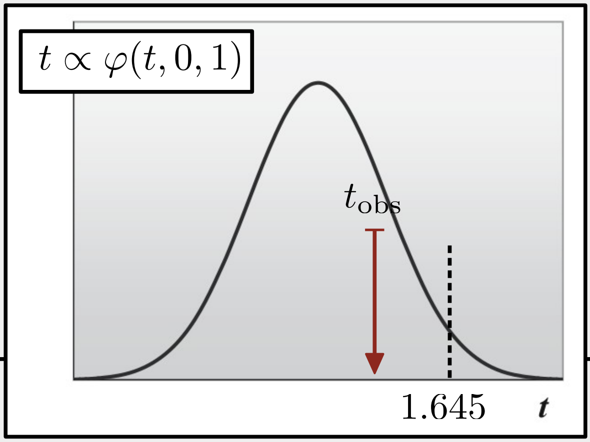
H_0 kann *nicht* ausge-schlossen werden

$\bar{x}_2 = 175\ \text{cm}\,; \quad \mathrm{n}_2 = 100$

$$t = \frac{1\,\text{cm}}{10\,\text{cm}/\sqrt{100}} = 1 < 1.654$$

H_0 kann *nicht* ausge-schlossen werden

$\bar{x}_3 = 175\ \text{cm}\,; \quad \mathrm{n}_3 = 400$



Beispiel aus der klassischen Statistik

- Es wurden drei unabhängige Tests mit wachsendem Umfang durchgeführt:

- Auswertung und Entscheidung:

$\bar{x}_1 = 176\ \text{cm}\,; \quad \mathrm{n}_1 = 36$

$$t = \frac{2\,\text{cm}}{10\,\text{cm}/\sqrt{36}} = 1.2 < 1.654$$

$\downarrow$

H_0 kann *nicht* ausgeschlossen werden

$\bar{x}_2 = 175\ \text{cm}\,; \quad \mathrm{n}_2 = 100$

$$t = \frac{1\,\text{cm}}{10\,\text{cm}/\sqrt{100}} = 1 < 1.654$$

$\downarrow$

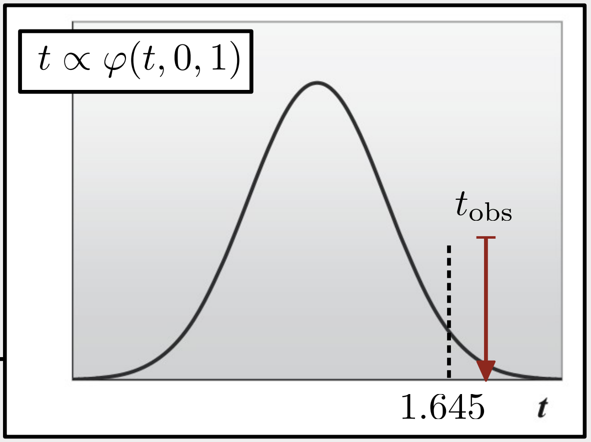
H_0 kann *nicht* ausgeschlossen werden

$\bar{x}_3 = 175\ \text{cm}\,; \quad \mathrm{n}_3 = 400$

$$t = \frac{1\,\text{cm}}{10\,\text{cm}/\sqrt{400}} = 2 > 1.654$$

$\downarrow$

H_0 *kann* ausgeschlossen werden



Beispiel aus der klassischen Statistik

- Es wurden drei unabhängige Tests mit wachsendem Umfang durchgeführt:

- Auswertung und Entscheidung:

$\bar{x}_1 = 176\ \mathrm{cm}\,;\quad \mathrm{n}_1 = 36$ → H_0 kann *nicht* ausge-schlossen werden

$\bar{x}_2 = 175\ \mathrm{cm}\,;\quad \mathrm{n}_2 = 100$ → H_0 kann *nicht* ausge-schlossen werden

$\bar{x}_3 = 175\ \mathrm{cm}\,;\quad \mathrm{n}_3 = 400$ → H_0 *kann* 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.: $H_0\colon\ \mu \leq \mu_0 = 174\,\mathrm{cm}\ \mathrm{mit\colon}\ \sigma = \sigma_0$). Für die Hypothese selbst ist keine Testfunktion angebbar.
  - Nicht-Widerlegung ist (statistisch) schwächer, als 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“.)
  - Dabei ist zu beachten, dass die Irrtumswahrscheinlichkeit nur die Wahrscheinlichkeit abbildet 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.

6 Hypothesentests in der modernen Physik

6.2 Fehler 1. und 2. Art

Für die folgende Diksussion legen wir die Befüllung von Bierflaschen zugrunde. Die Flaschen sollen mit einer Irrtumswahrscheinlichkeit von $\alpha = 0.05$ nicht mit mehr als 500 ml Bier befüllt sein (obere Schranke mit 95% Konfidenz). Die Befüllung *einer* Flasche erfolgt mit einer Varianz von $\sigma_0 = \pm 10\,\mathrm{ml}$. Regelmäßige Stichproben der Länge n sollen die Qualität der Befüllung sicherstellen.

Fehler 1. & 2. Art

- Für die Beurteilung eines Hypothesentests ist mehr als eine Irrtumswahrscheinlichkeit von Relevanz:

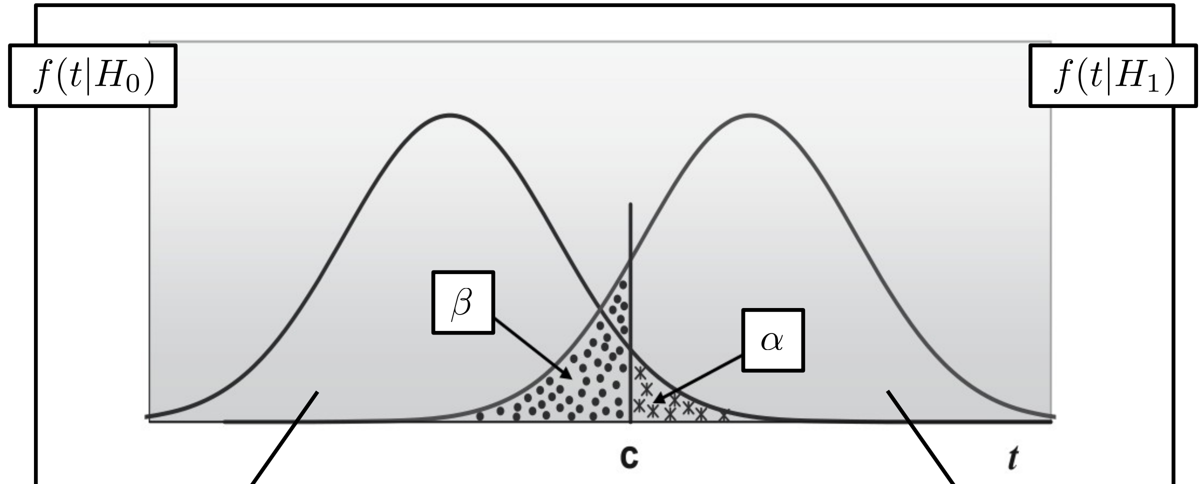
Wir formulieren das Beispiel als „Zwei Hypothesen Test“: $H_0 : \ \mu \leq c \quad H_1 : \ \mu > c$

$\alpha :$ Fehler 1. Art

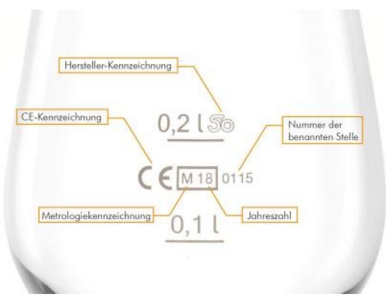
$\beta :$ Fehler 2. Art

$c \ :$ Kritischer Wert

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



Füllmenge unter Maximalwert (H_0)



Füllmenge über Maximalwert (H_1)

Fehler 1. & 2. Art

- Für die Beurteilung eines Hypothesentests ist mehr als eine Irrtumswahrscheinlichkeit von Relevanz:

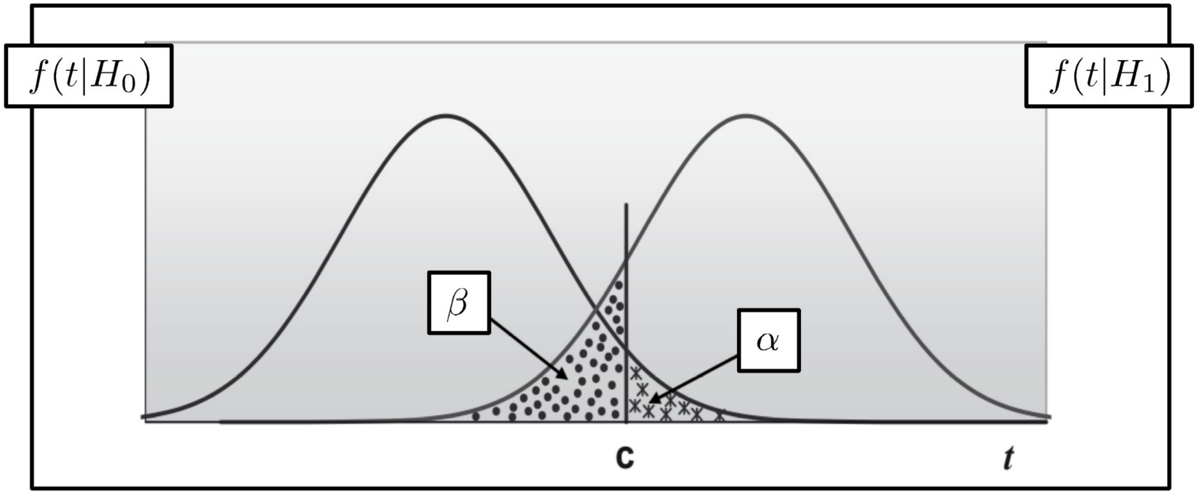
Wir formulieren das Beispiel als „Zwei Hypothesen Test“: $H_0 : \mu \leq c \quad H_1 : \mu > c$

$\alpha :$ Fehler 1. Art

$\beta :$ Fehler 2. Art

$c \;:$ Kritischer Wert

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



	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 $(1-\beta)$

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

Operationscharakteristik & Trennschärfe

- 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** oder **Gütefunktion**. Der Funktionswert bei vorgegebenem μ_1 heißt **Trennschärfe**.
- Ist z.B. α vorgegeben, dann wird man aus allen möglichen Testverfahren dasjenige mit dem kleinsten β (d.h. der größten Trennschärfe) auswählen.

Operationscharakteristik – am Beispiel –

- In unserem Beispiel testen wir die Füllhöhe durch, ziehen einer Stichprobe der Länge $n = 10$ und anschließende Mittelwertbildung:

$$\mu = \bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i$$

- Wir können davon ausgehen, dass die gewählte Teststatistik μ normalverteilt ist. Für die Wurzel der Varianz von μ gilt:

$$\sigma_\mu = \frac{\sigma_0}{\sqrt{10}} = \frac{10\,\mathrm{ml}}{\sqrt{10}} \approx 3.16\,\mathrm{ml}$$

- Die Füllhöhe wird mit einer Irrtumswahrscheinlichkeit von $\alpha = 0.05$ beanstandet falls

$$\mu > c \equiv 500\,\mathrm{ml} + 1.645\,\sigma_\mu = 500\,\mathrm{ml} + (1.645 \cdot 3.16)\,\mathrm{ml} \approx 505\,\mathrm{ml}$$

(siehe Folie 11)

Operationscharakteristik – am Beispiel –

- In unserem Beispiel testen wir die Füllhöhe durch, ziehen einer Stichprobe der Länge $n = 10$ und anschließende Mittelwertbildung:

$$\mu = \bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i$$

- Wir können davon ausgehen, dass die gewählte Teststatistik μ normalverteilt ist. Für die Wurzel der Varianz von μ gilt:

$$\sigma_\mu = \frac{\sigma_0}{\sqrt{10}} = \frac{10\,\mathrm{ml}}{\sqrt{10}} \approx 3.16\,\mathrm{ml}$$

- Die Füllhöhe wird mit einer Irrtumswahrscheinlichkeit von $\alpha = 0.05$ beanstandet falls

$$\mu > c \equiv 500\,\mathrm{ml} + 1.645\,\sigma_\mu = 500\,\mathrm{ml} + (1.645 \cdot 3.16)\,\mathrm{ml} \approx 505\,\mathrm{ml}$$

(siehe Folie 11)

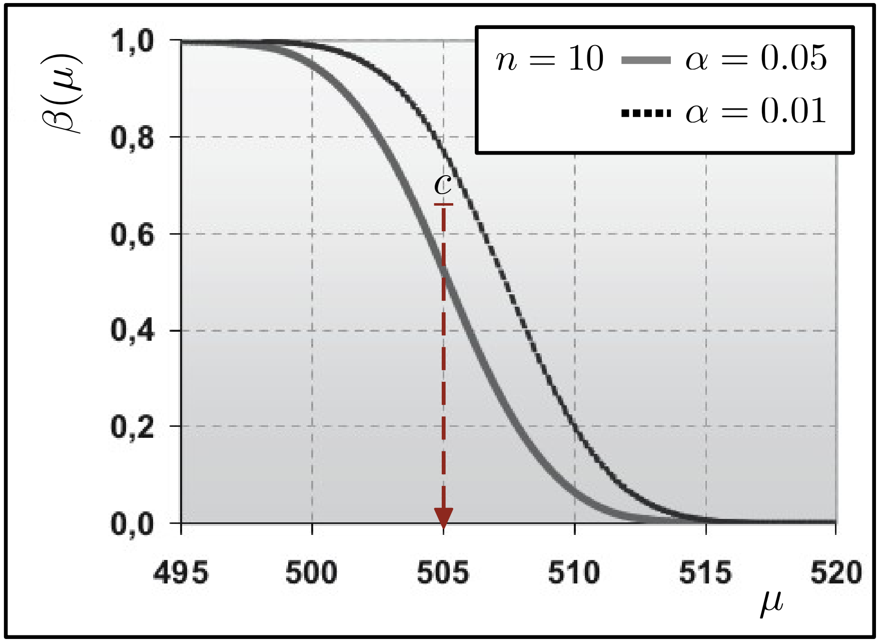
- Falls die Flaschen wirklich mit $\mu_0 = 500\,\mathrm{ml}$ befüllt sind, wird der Füllstand mit einer Wahrscheinlichkeit von $\alpha = 0.05$ irrtümlich beanstandet. Wie groß ist die Wahrscheinlichkeit für Nicht-Beanstandung, falls die Füllhöhe tatsächlich höher oder niedriger ist?

Operationscharakteristik – am Beispiel –

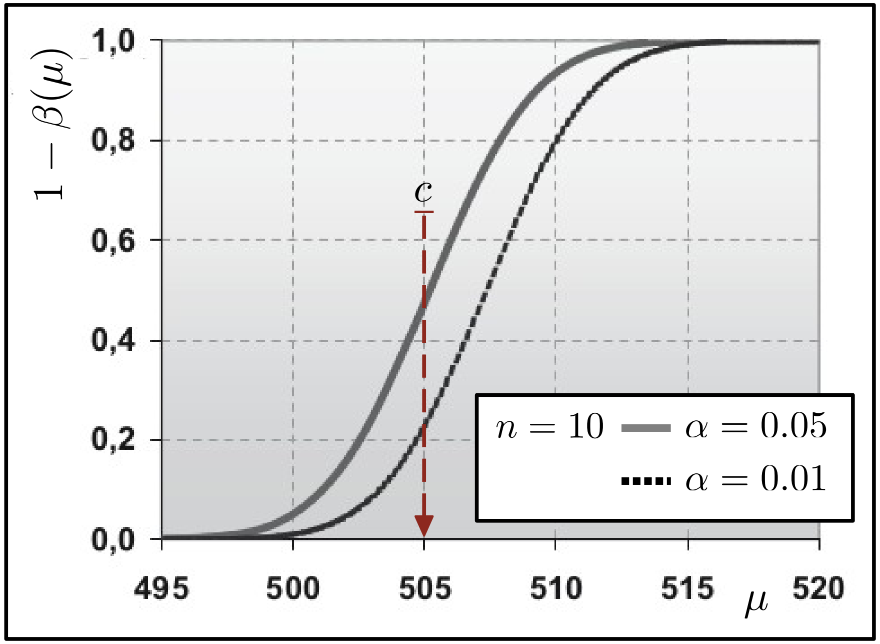
- Unter Variation der Hypothese H_1 erhält man die Operationscharakteristik aus:

$$\beta(\mu|H_1) = \int_{-\infty}^{c(\alpha)} \varphi(x, \mu, \sigma_\mu)\,\mathrm{d}x$$

Operationscharakteristik:



Gütefunktion:

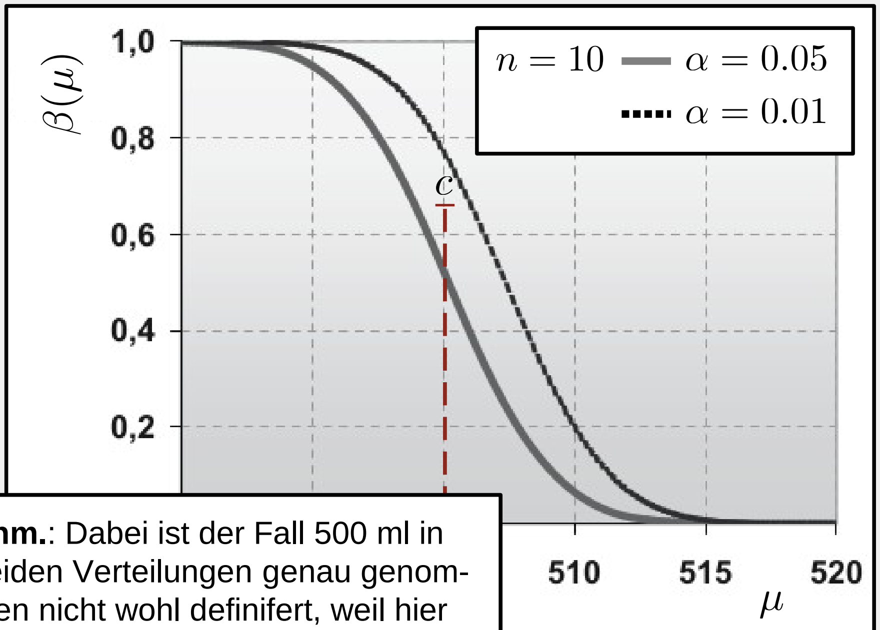


Operationscharakteristik – am Beispiel –

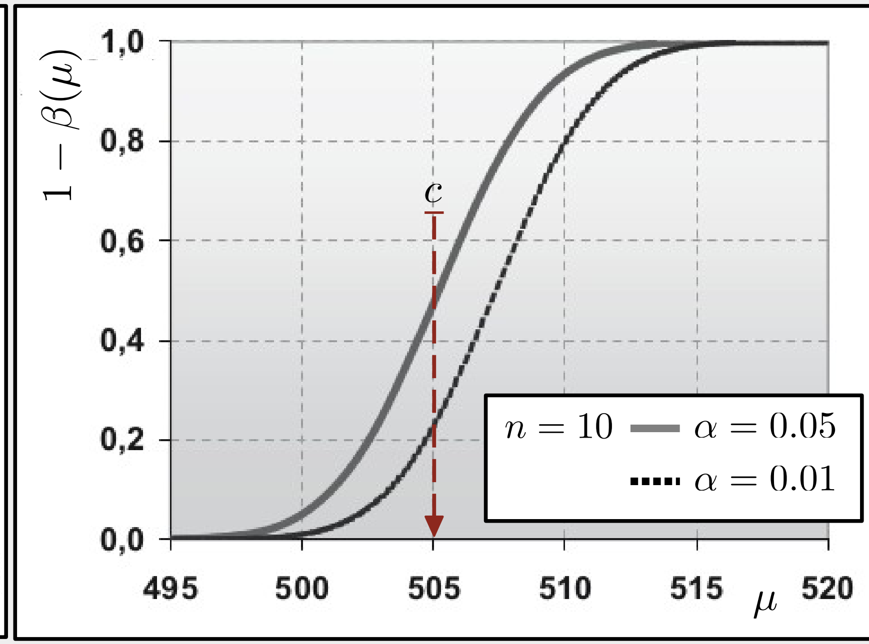
- Unter Variation der Hypothese H_1 erhält man die Operationscharakteristik aus:

$$\beta(\mu|H_1) = \int_{-\infty}^{c(\alpha)} \varphi(x, \mu, \sigma_\mu)\,\mathrm{d}x$$

Operationscharakteristik:



Gütefunktion:



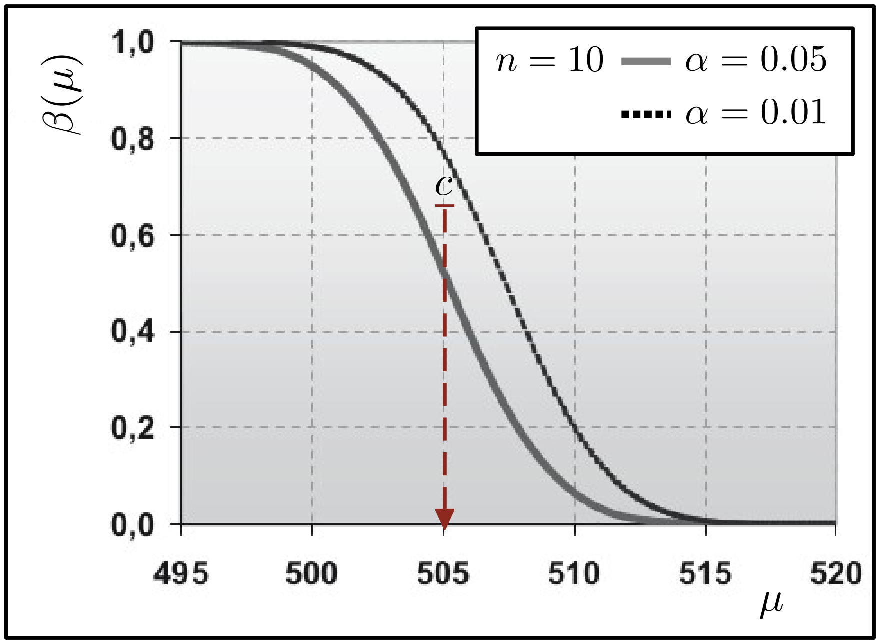
Anm.: Dabei ist der Fall 500 ml in beiden Verteilungen genau genommen nicht wohl definifert, weil hier kein Fehler in der Befüllung vorliegt.

Operationscharakteristik – am Beispiel –

- Unter Variation der Hypothese H_1 erhält man die Operationscharakteristik aus:

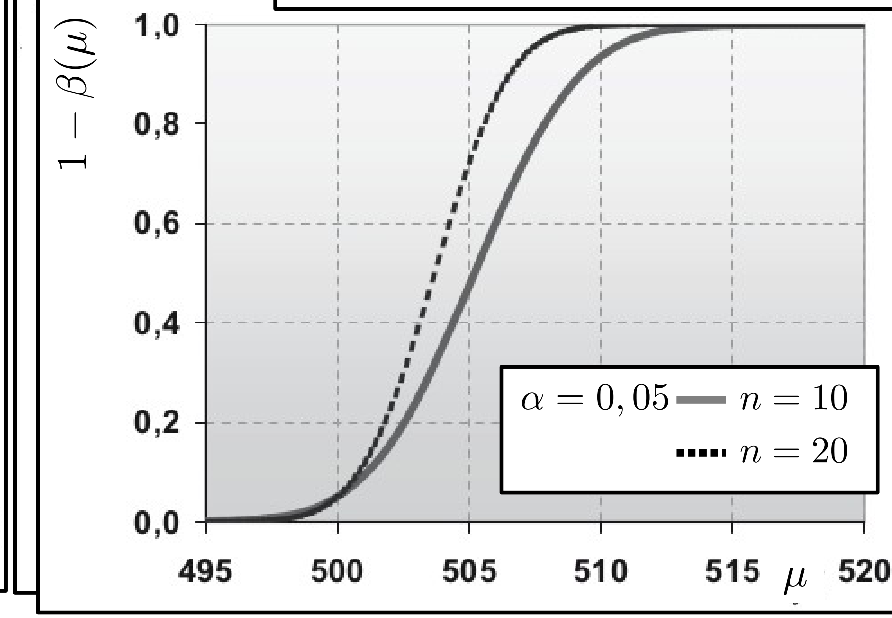
$$\beta(\mu|H_1) = \int_{-\infty}^{c(\alpha)} \varphi(x, \mu, \sigma_\mu)\,\mathrm{d}x$$

Operationscharakteristik:



Gütefunktion:

Sie sehen hier zwei Tests unterschiedlicher Stichprobenlänge für $\alpha = 0.05$. Welchen Test würden Sie wählen?



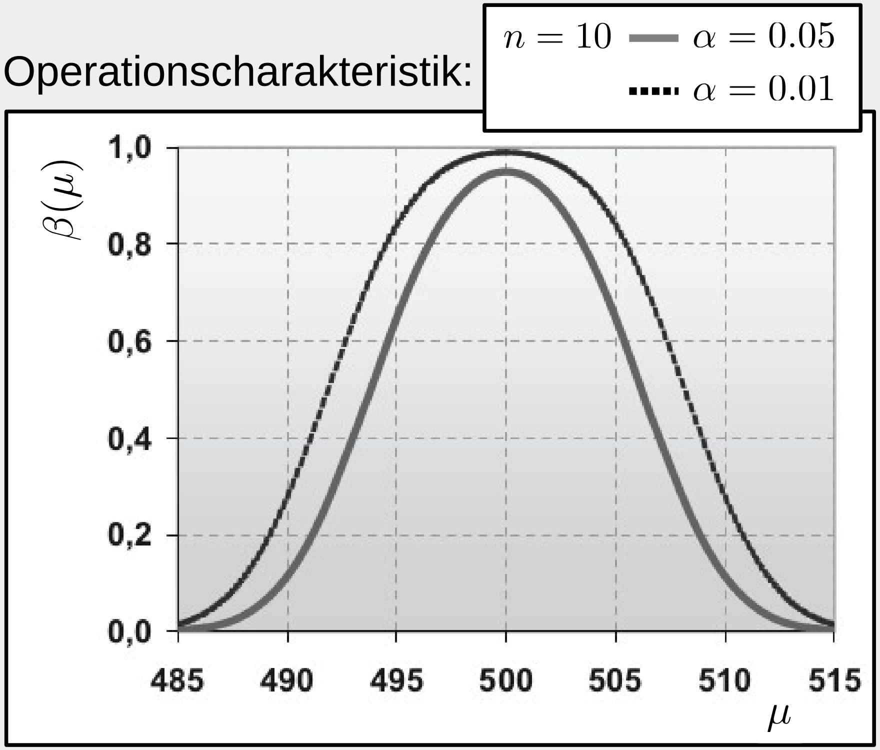
Operationscharakteristik – am Beispiel –

Anm.: Hier sehen Sie das Beispiel eines zweiseitigen Hypothesentests (vgl. Folie 6).

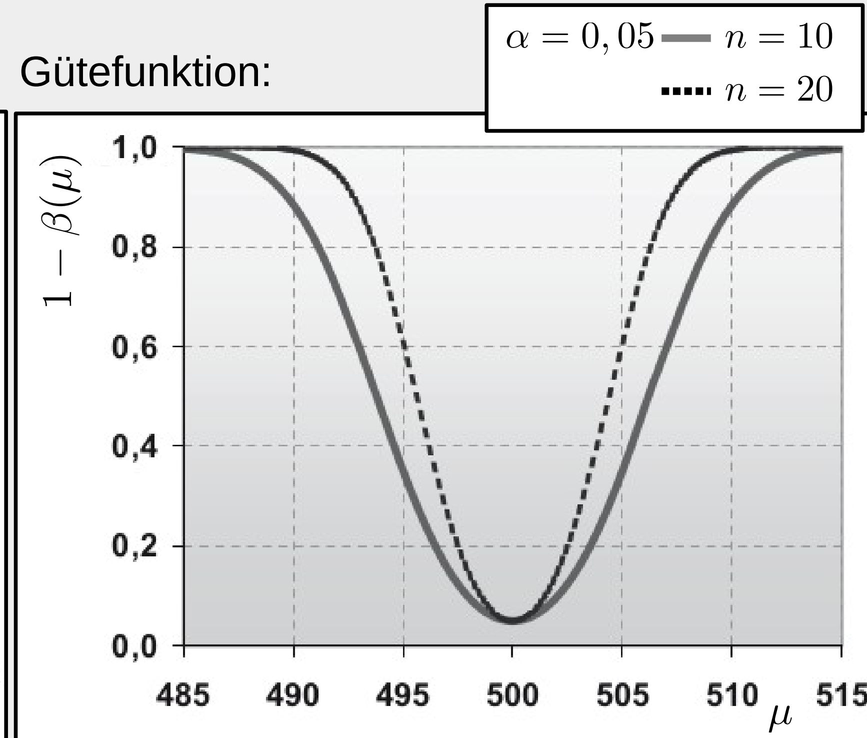
- Unter Variation der Hypothese H_1 erhält man die Operationscharakteristik aus:

$$\beta(\mu|H_1) = \int_{-\infty}^{c(\alpha)} \varphi(x, \mu, \sigma_\mu)\,\mathrm{d}x$$

Operationscharakteristik:



Gütefunktion:



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

Effizienz und Reinheit

- Bei Klassifikationsproblemen werden Ereignisse oftmals entweder einem Typ-0 oder einem Typ-1 zugeordnet.

- In diesem einfachen Fall besteht die folgende Relation zwischen dem Fehler 1. und 2. Art und den häufig verwendeten Begriffen **Effizienz** und **Reinheit**:

$$\epsilon = \frac{\text{Richtig erkannte Ereignisse vom Typ-0}}{\text{Alle Ereignisse vom Typ-0}} = 1 - \alpha \qquad \text{(Effizienz)}$$

$$\rho = \frac{(1-\alpha)\,N_0}{(1-\alpha)\,N_0 + \beta\,N_1} \qquad \text{(Reinheit)}$$

α : Fehler 1. Art

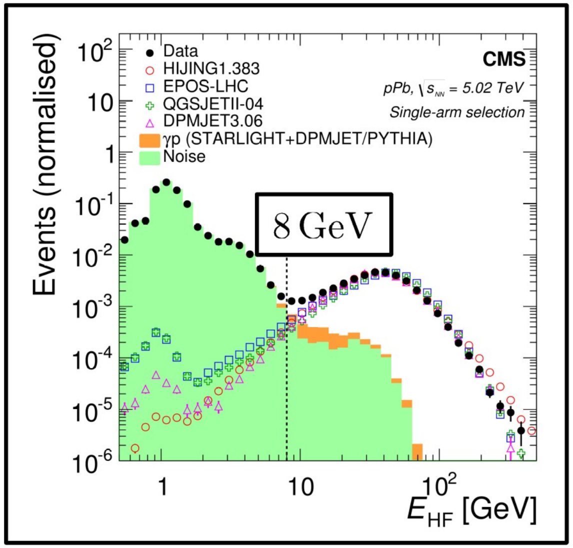
β : Fehler 2. Art

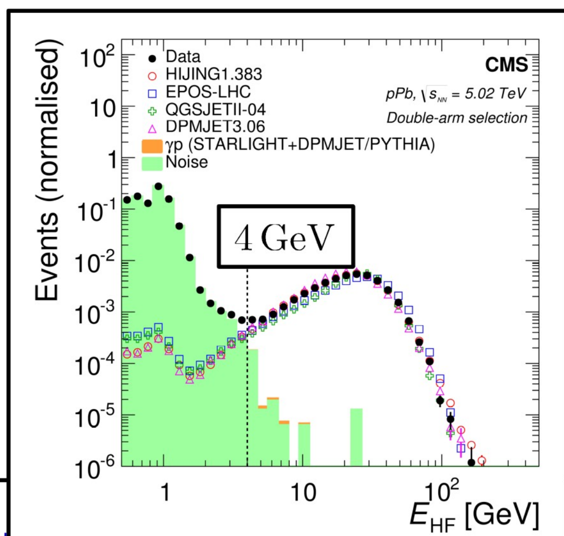
N_0 : Wahre Anzahl der Ereignisse vom Typ-0

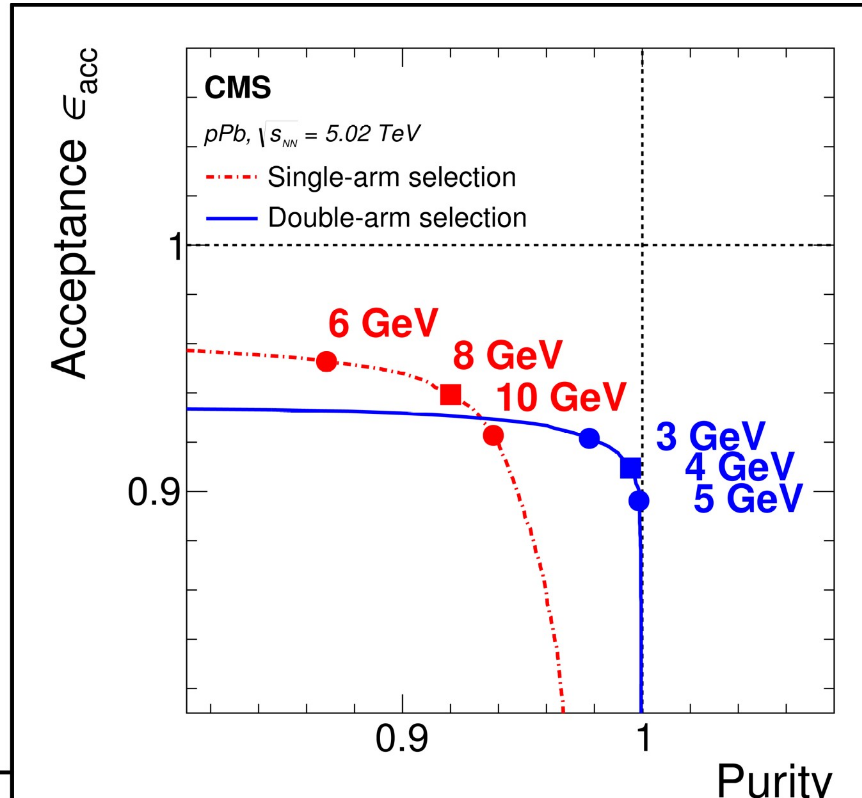
N_1 : Wahre Anzahl der Ereignisse vom Typ-1

ROC Kurve

- In solchen Fällen, in denen sich Effizienz und Reinheit definieren lassen, wird die Gütefunktion eines Tests auch durch die Kurve der *Reciever Operating Characteristic* (**ROC**) angegeben.



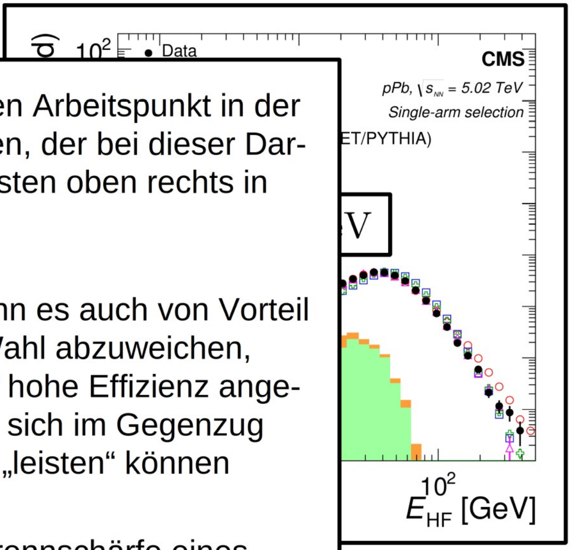




PLB 759 (2016) 641

ROC Kurve

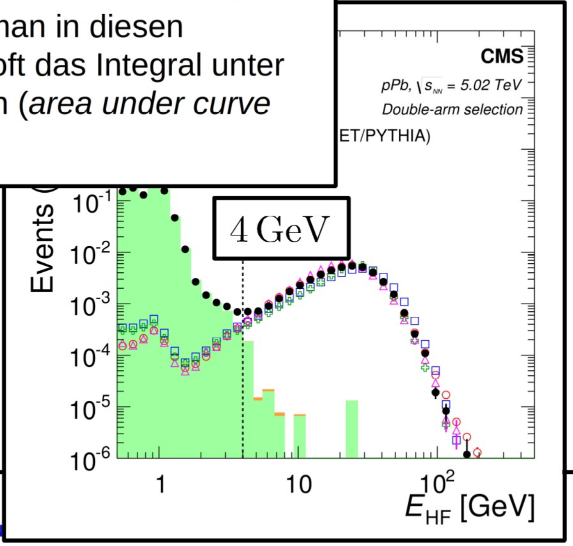
- In solchen Fällen, in denen sich Effizienz und Reinheit definieren lassen, wird die Gütefunktion eines Tests auch durch die Kurve der *Reciever Operating Characteristic* (**ROC**) angegeben.

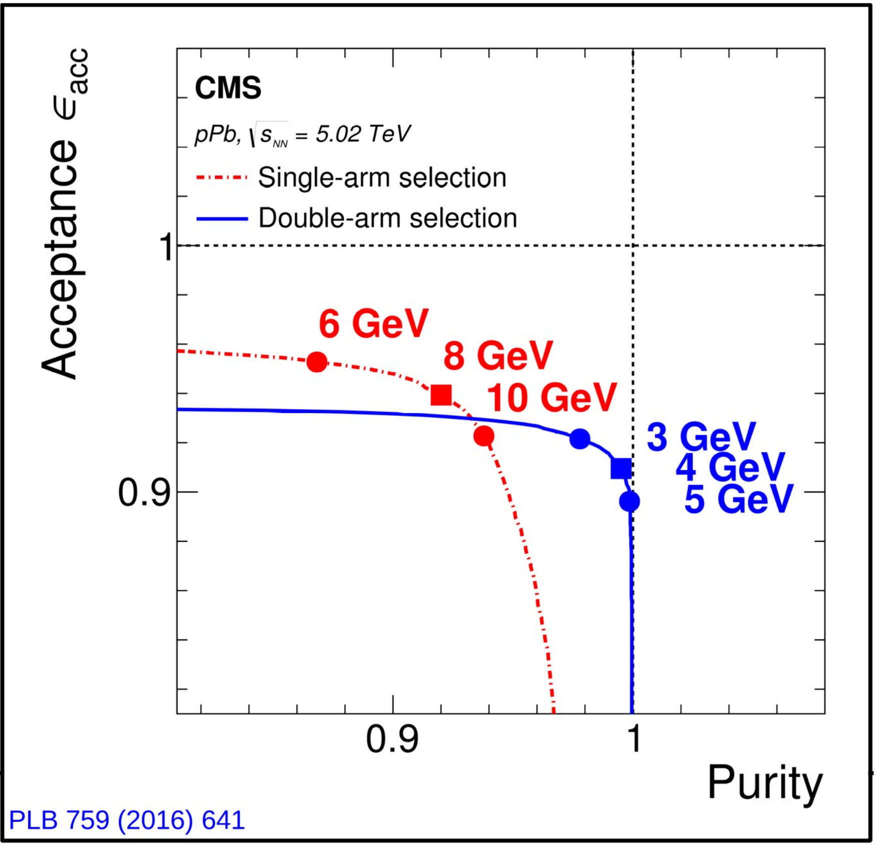


- Sie würden i.a. den Arbeitspunkt in der ROC Kurve wählen, der bei dieser Darstellung am weitesten oben rechts in der Kurve liegt.

- In Einzelfällen kann es auch von Vorteil sein von dieser Wahl abzuweichen, wenn Sie z.B. auf hohe Effizienz angewiesen sind, oder sich im Gegenzug niedrige Effizienz „leisten" können

- Als Maß für die Trennschärfe eines Algorithmus gibt man in diesen einfachen Fällen oft das Integral unter der ROC Kurve an (*area under curve* **AUC**).





PLB 759 (2016) 641

6.3 Klassische Beispiele und Anwendungen

Im Folgenden werden wir einige Beispiele für Hypothesentests explizit vorrechnen, anhand derer Sie ein Gefühl für die Vielfalt solcher Tests entwickeln können.

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 $n_1, n_2 = 200$.

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“ ($H_0\colon\ \mu_1 \geq \mu_2$).

  - Teststatistik (Anm.: Streng formuliert für die Bestimmung einer unteren Schanke):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \propto \varphi(x, 0, 1)$$

  - Irrtumswahrscheinlichkeit: $\alpha = 0.05 \quad \text{entspricht:}\ (z = -1.645)$ (siehe Folie 11)

Zurück zu Folie 33

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" ($H_0\colon\ \mu_1 \geq \mu_2$).

  - Teststatistik (Anm.: Streng formuliert für die Bestimmung einer unteren Schanke):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \propto \varphi(x, 0, 1)$$

  - Irrtumswahrscheinlichkeit: $\alpha = 0.05 \quad \text{entspricht:}\ (z = -1.645)$ (siehe Folie 11)

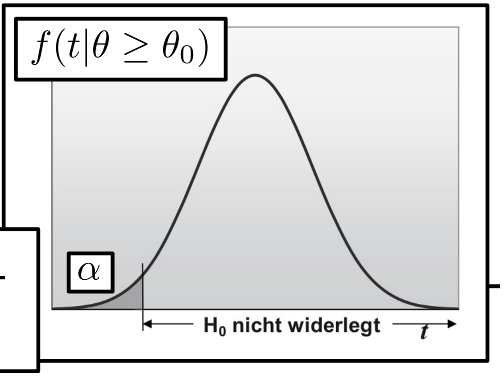
  - Stichprobenergebnisse:

$$\bar{x}_1 = 2\,200 \qquad \bar{x}_2 = 2\,300$$
$$s_1^2 = 700\,000 \qquad s_2^2 = 920\,000$$

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

$$t = \frac{2\,200 - 2\,300}{900 \cdot \frac{1}{10}} = -\frac{100}{90} = -1.11 > -1.645$$

(siehe Folie 5)



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

Beispiel-2: Varianzanalyse

- **Hypothese**: In einer Untersuchung zur Verkehrssituation in der Innenstadt geben zufällig Befragte ein Punkturteil ab (je höher, desto positiver). Die Untersuchung wird getrennt für Innenstadtbewohner ($\mu_1, n_1 = 36$) und Pendler ($\mu_2, n_2 = 41$) vorgenommen. Ist die Bewertung innerhalb beider Personengruppen gleich?
  - H_0 : „Beide Gruppen beurteilen die Verkehrsituation gleich“ ($H_0\colon\ \mu_1 = \mu_2$).
  - Wir führen diesen Test zuächst als F-Test durch:

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

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

s_{ext}^2 : Varianz zwischen den Gruppen

s_{int}^2 : Varianz innerhalb der Gruppen

Beispiel-2: Varianzanalyse

- **Hypothese**: In einer Untersuchung zur Verkehrssituation in der Innenstadt geben zufällig Befragte ein Punkturteil ab (je höher, desto positiver). Die Untersuchung wird getrennt für Innenstadtbewohner ($\mu_1, n_1 = 36$) und Pendler ($\mu_2, n_2 = 41$) vorgenommen. Ist die Bewertung innerhalb beider Personengruppen gleich?
  - H_0 : „Beide Gruppen beurteilen die Verkehrsituation gleich“ ($H_0\!:\ \mu_1 = \mu_2$).
  - Wir führen diesen Test zuächst als F-Test durch:

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

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

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

Diese Teststatistik folgt der Fisher-Verteilung, für den Quotienten zweier χ^2-verteilter Zufallsgrößen:

- $r-1$ entspricht den Freiheitsgraden der Verteilung im Zähler.
- $n-r$ entspricht den Freiheitsgraden der Verteilung im Nenner.
- Es handelt sich hierbei um einen **F-Test**.
- Die einfachste Art die CDF zu $f\,(F)$ auszuwerten ist mit Hilfe von scipy. Alternativ können Sie die Werte aus Tabellen auslesen.

NB: Dies ist ein Test auf $s^2_{\text{ext}} = s^2_{\text{int}}$.

Beispiel-2: Varianzanalyse

Z.B. bestimmt mit: scipy.stats.f.cdf(3.968, 1, 75).

- Irrtumswahscheinlichkeit: $\alpha = 0.05 \quad 1 - F(3.968|1, 75) = 0.05$
- Stichprobenergebnisse:

$$\bar{x}_1 = 50 \qquad \bar{x}_2 = 48.4 \qquad \bar{x} = 49.15$$

$$s_1^2 = \ 8.4 \qquad s_2^2 = \ 7.5$$

$$\text{a)} \quad \sum_{g=1}^{2} n_g \left(\bar{x}_g - \bar{x}\right)^2 = 36 \cdot (50 - 49.15)^2 + 41 \cdot (48.4 - 49.15)^2 = 49.0725$$

$$\sum_{g=1}^{2} \sum_{i=1}^{n_g} \left(x_{gi} - \bar{x}_g\right)^2 = \sum_{g=1}^{2} \left(n_g - 1\right) \sigma_g^2 = 35 \cdot 8.4 + 40 \cdot 7.5 = 594$$

$$t = \frac{49.0725}{\frac{1}{75} \cdot 594} = 6.196 (> 3.968)$$

- Testentscheidung: H_0 ist widerlegt ($\alpha^* = 0.015$), die Verkehrssituation wird unterschiedlich bewertet.

Beispiel-2 als Zweistichprobentest

- **Hypothese**: In einer Untersuchung zur Verkehrssituation in der Innenstadt geben zufällig Befragte ein Punkturteil ab (je höher, desto positiver). Die Untersuchung wird getrennt für Innenstadtbewohner ($\mu_1, n_1 = 36$) und Pendler ($\mu_2, n_2 = 41$) vorgenommen. Ist die Bewertung innerhalb beider Personengruppen gleich?

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

  - Sie können den Test auch auf der Teststatistik von Folie 29 durchführen:

$$\text{b)} \quad t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \propto \varphi(x, 0, 1)$$

Anm.: Achtung im Gegensatz zu Folie 29 steht hier ein Betragsstrich im Zähler. Wir führen nämlich einen zweiseitigen Test durch.

Beispiel-2 als Zweistichprobentest

- Irrtumswahscheinlichkeit: $\alpha = 0.05 \quad z = 1.96$ (diesesmal zweiseitig! Siehe Folie 11)

- Stichprobenergebnisse:

$$\bar{x}_1 = 50 \qquad \bar{x}_2 = 48.4 \qquad \bar{x} = 49.15$$
$$s_1^2 = \; 8.4 \qquad s_2^2 = \; 7.5$$

Beispiel-2 als Zweistichprobentest

- Irrtumswahscheinlichkeit: $\alpha = 0.05 \quad z = 1.96$ (diesesmal zweiseitig! Siehe Folie 11)

- Stichprobenergebnisse:

$$\bar{x}_1 = 50 \qquad \bar{x}_2 = 48.4 \qquad \bar{x} = 49.15$$

$$s_1^2 = 8.4 \qquad s_2^2 = 7.5$$

$$\text{b)} \quad s^2 = \frac{35 \cdot 8.4 + 40 \cdot 7.5}{75} = 7.92$$

$$t = \frac{50 - 48.4}{\sqrt{7.92} \cdot \left(\frac{1}{36} + \frac{1}{41}\right)} = 2.489 (> 1.96)$$

- Testentscheidung: H_0 ist widerlegt ($\alpha^* = 0.013$), die Verkehrssituation wird unterschiedlich bewertet.

Beispiel-2 als Zweistichprobentest

- Irrtumswahscheinlichkeit: $\alpha = 0.05 \quad z = 1.96$ (diesesmal zweiseitig! Siehe Folie 11)

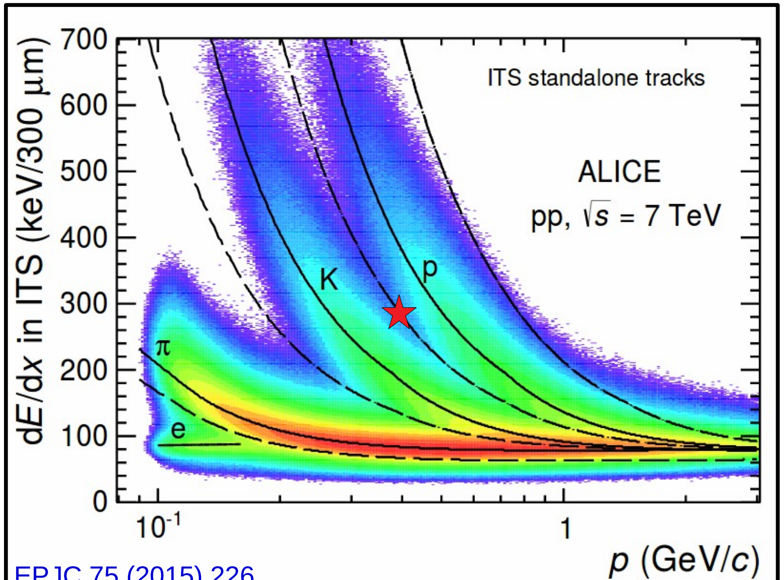
- Stichprobenergebnisse:

$$\bar{x}_1 = 50 \qquad \bar{x}_2 = 48.4 \qquad \bar{x} = 49.15$$
$$s_1^2 = \ 8.4 \qquad s_2^2 = \ 7.5$$

- Testentscheidung: H_0 ist widerlegt ($\alpha^* = 0.013$), die Verkehrssituation wird unterschiedlich bewertet.

Wichtigste Hypothesentests in der (Teilchen-)physik

- Ist die Energiedeposition eines Teilchens mit vorgegebenem Impuls in Materie kompatibel mit der Teilchenhypothese eines Elektrons, Pions, Kaons oder Protons?

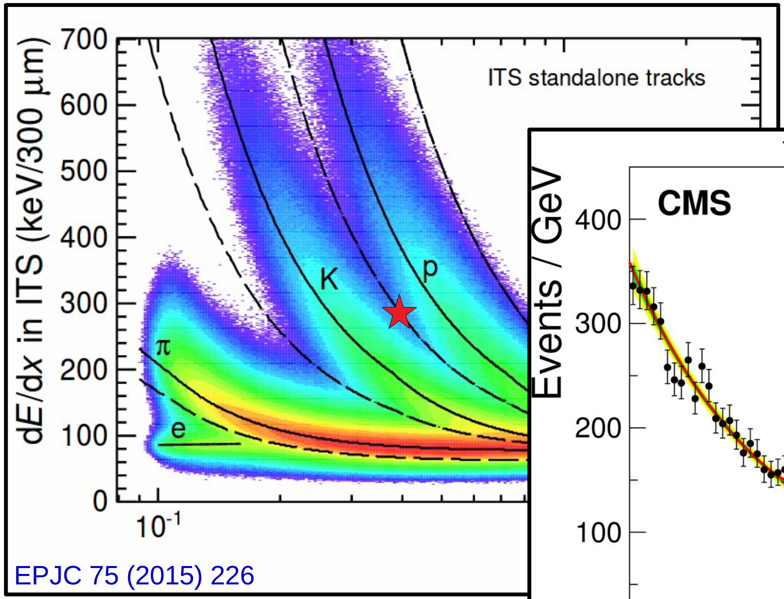


EPJC 75 (2015) 226

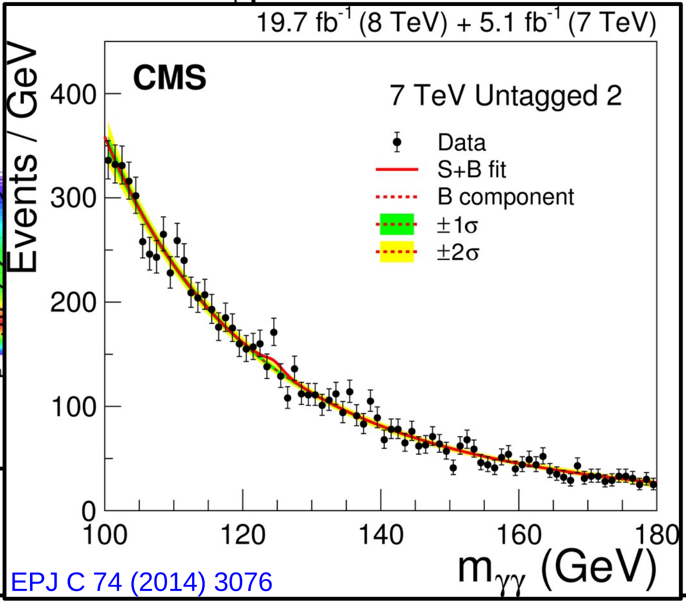
Wichtigste Hypothesentests in der (Teilchen-)physik

- Ist die Energiedeposition eines Teilchens mit vorgegebenem Impuls in Materie kompatibel mit der Teilchenhypothese eines Elektrons, Pions, Kaons oder Protons?

- Weisen die Daten auf die Existenz eines neues Teilchens hin?



EPJC 75 (2015) 226



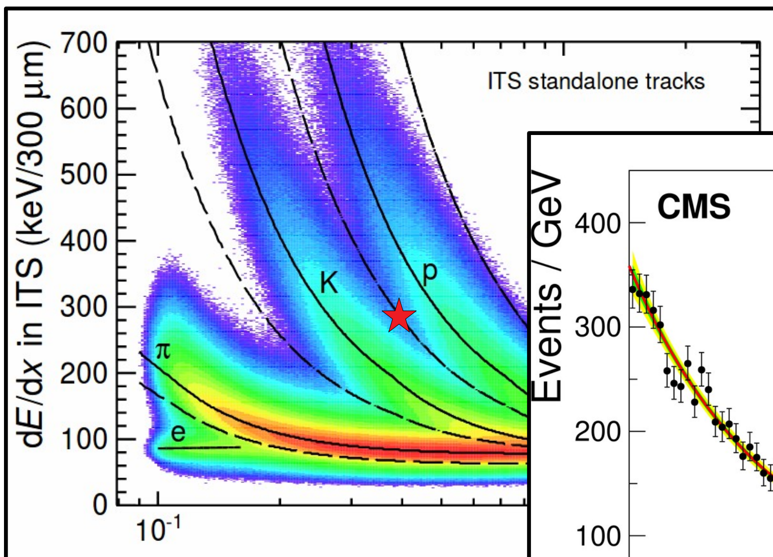
EPJ C 74 (2014) 3076

Wichtigste Hypothesentests in der (Teilchen-)physik

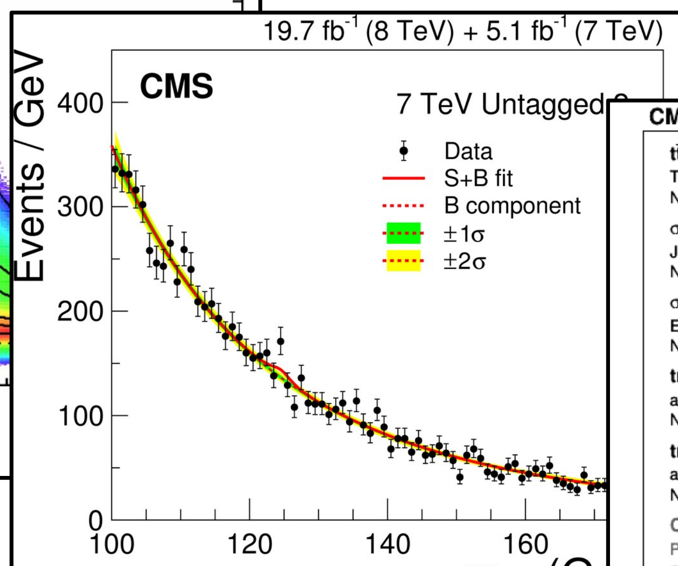
- Ist die Energiedeposition eines Teilchens mit vorgegebenem Impuls in Materie kompatibel mit der Teilchenhypothese eines Elektrons, Pions, Kaons oder Protons?

- Weisen die Daten auf die Existenz eines neues Teilchens hin?

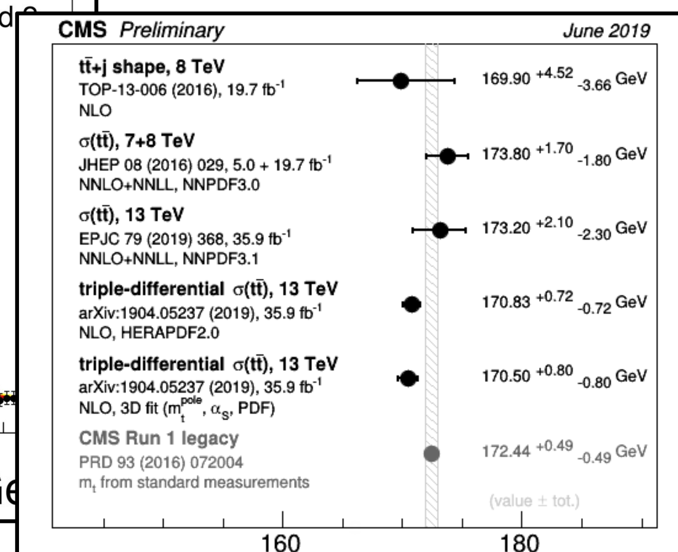
- Sind zwei oder mehrere Messungen der vermeintlich gleichen physikalischen Größe miteinander kompatibel?



EPJC 75 (2015) 226



EPJ C 74 (2014) 3076



CMS Summary (May 2019)

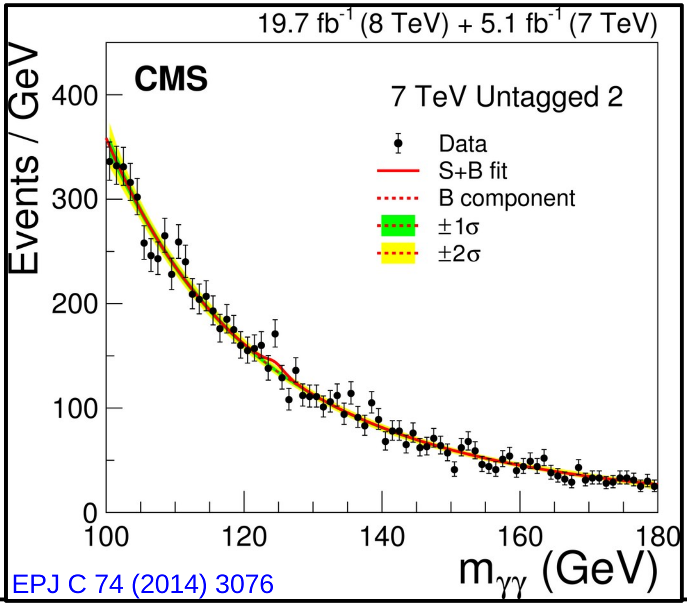
Zweihypothesen Poisson Test

- Ein wichtiger Test in der Teilchenphysik, den wir im Folgenden eingehender diskutieren werden, ist der Zweihypothesen Test basierend auf der Poisson Zählstatistik, z.B.:

- Weisen die Daten auf die Existenz eines neues Teilchens hin?

- In diesem Fall haben wir zwei Hypothesen, z.B.:

$$H_0 : \text{Etablierte Hypothese (nur Untergrund)}$$
$$H_1 : \text{Alternative Hypothese (zusätzliches Signal)}$$



Lemma von Neyman-Pearson (einfache Hypothesen)

Für den Test zweier einfacher Hypothesen H_0 und H_1 ist für den Likelihood-Quotienten

$$\lambda(\vec{x}) = \frac{\mathcal{L}(\vec{x}|H_0)}{\mathcal{L}(\vec{x}|H_1)}$$

$1-\beta$ bei vorgegebenem α maximal, d.h. $\lambda(\vec{x})$ ist die Teststatistik mit der größten Trennschärfe. Der Likelihood-Quotient $\lambda(\vec{x})$ ist äquivalent zur Differenz der Log-Likelihood:

$$q(\vec{x}) = -2\ln\left(\frac{\mathcal{L}(\vec{x}|H_0)}{\mathcal{L}(\vec{x}|H_1)}\right) = -2\Big(\ln\left(\mathcal{L}(\vec{x}|H_0)\right) - \ln\left(\mathcal{L}(\vec{x}|H_1)\right)\Big)$$

- Das Lemma von Neyman-Pearson gilt nur für <u>einfache Hypothesen</u>, es lässt sich jedoch auf zusammengesetzte Hypothesen verallgemeinern (siehe Folie 38).

Zurück zu Folie 57

Lemma von Neyman-Pearson (zusammengesetzte Hyp.)

Für eine Verallgemeinerung des Lemmas von Neyman-Pearson bilden Sie den Likelihood-Quotienten aus

$$\lambda(\vec{x}, \{\theta_i\}) = \frac{\max\limits_{\theta_i} \left(\mathcal{L}(\vec{x}, \{\theta_i\}|H_1)\right)}{\max\limits_{\theta_i' \in \Omega} \left(\mathcal{L}(\vec{x}, \{\theta_i'\}|H_0)\right)} ,$$

wobei $\theta_i' \in \Omega$ bedeutet, dass die Parameter $\{\theta_i'\}$ mit H_0 kompatibel sein müssen. Die Hypothese H_0 wird mit der Irrtumswahrscheinlichkeit

$$\alpha = P(\lambda > \lambda_\alpha|H_0) = \int\limits_{\lambda_\alpha}^{\infty} g(\lambda|H_0)\,\mathrm{d}\lambda$$

verworfen.

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

Beispiel CP-Eigenzustand des beobachteten Higgs Bosons

- Goldener Zerfallskanal für solche Messungen: $H \to ZZ^* \to 4\ell$

$$P(Y_L^m(\theta,\varphi)) = (-1)^L \cdot Y_L^m(\theta,\varphi)$$

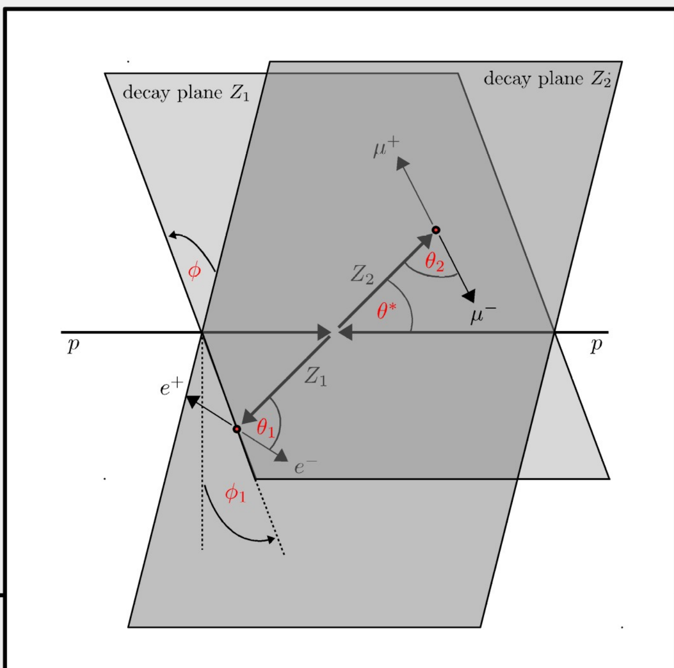
$$P(4\ell) = (-1)^L(-1)^2(+1)^2 = (-1)^L$$

Paritätsoperator

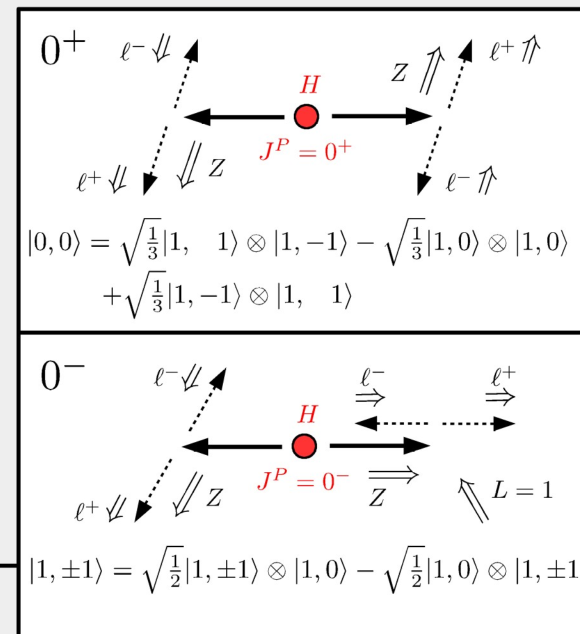
Anteil der Ortswellenfunktion

Anteil der Spins im Endzustand

Wie Drehimpuls- & Spinkonfiguration Rückschlüsse auf den CP-Eigenzustand des Higgs Bosons zulassen:

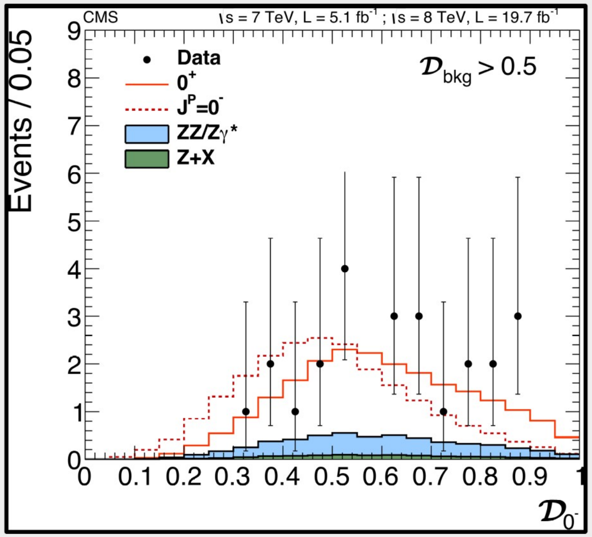


Charakterisierung der Zerfallsebenen des $H \to ZZ^* \to 4\ell$ Zerfalls.

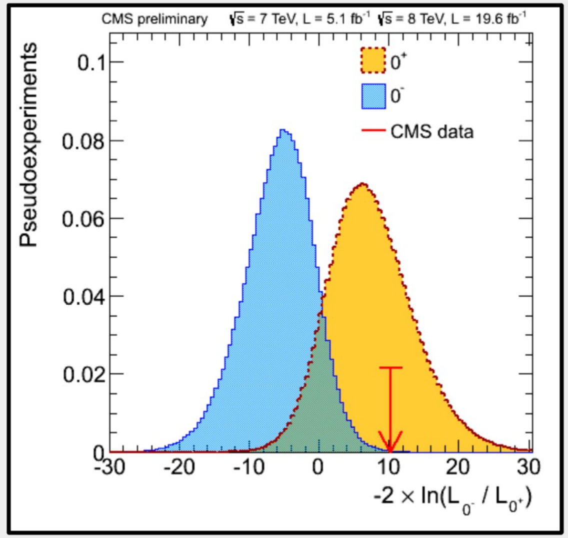


Beispiel CP-Eigenzustand des beobachteten Higgs Bosons

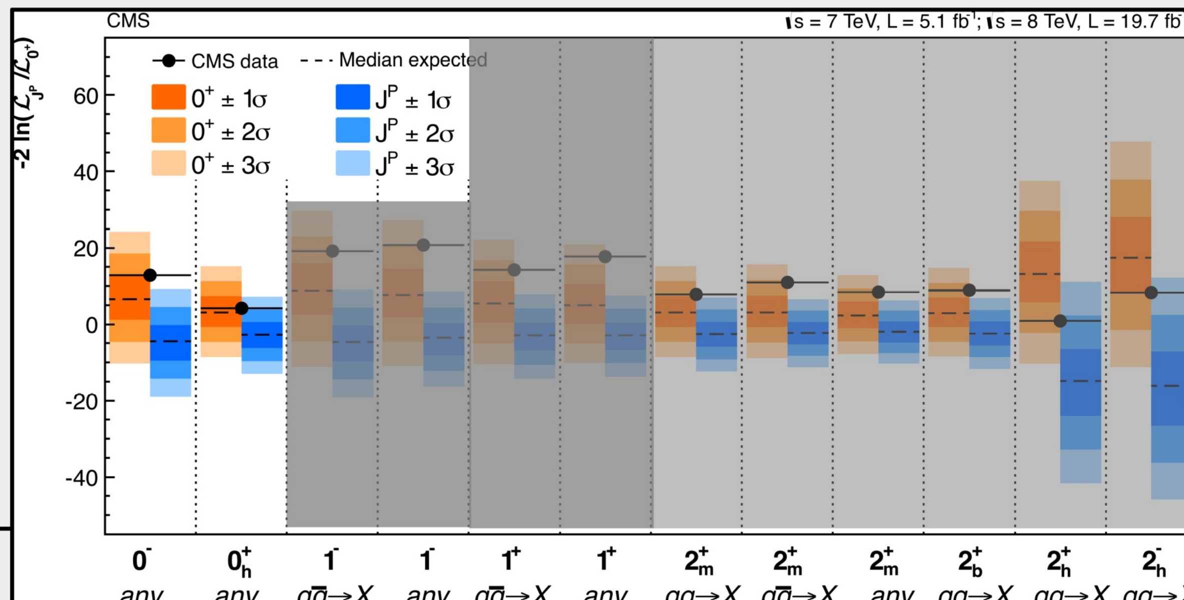
- Goldener Zerfallskanal für solche Messungen: $H \to ZZ^* \to 4\ell$



PRD 89 (2014) 092007



Test reiner Spin-Paritäts Hypothesen basierend auf $\mathcal{O}(50)$ Ereignissen:



Theorem von Wilks

Für Stichproben $\{x_i\}$, $i = 0, 1, \dots n$ einer Zufallsvariablen x der Länge $n \to \infty$ gilt: Wird die Grundgesamtheit der Zufallsvariablen x durch eine Wahrscheinlichkeitsdichte $p(x, \{\theta_j\})$ beschrieben, und legt die Hypothese H_0 $r = m_1 - m_0$ der Parameter $\{\theta_j\}$ fest, so folgt die Testfunktion

$$q(\vec{x}, \vec{\theta}) = -2 \ln \Big(\lambda(\vec{x}, \vec{\theta}) \Big)$$

einer χ^2-Verteilung mit r Freiheitsgraden. Dabei entspricht m_i der Anzahl der freien Parameter der Hypothese H_i.

Annals Math. Statist. 9 (1938) 1

- **Beispiel**: Zählexperiment mit einem Signal (s) zusätzlich zu bekanntem Untergrund (b)

$$q(\vec{x}, \theta) = -2 \ln \left(\frac{\mathcal{L}(\vec{x}|\theta s + b)}{\mathcal{L}(\vec{x}|b)} \right) \propto \chi^2(x', 1) \longleftarrow$$

χ^2-Verteilung mit <u>einem</u> Freiheitsgrad

- D.h. wenn Sie sich sicher sind, dass Ihre Stichprobe den Anforderungen von Wilks‘ Theorem entspricht, können Sie p-Werte und Irrtumswahrscheinlichkeiten <u>analytisch</u>, d.h. ohne die Auswertung von Pseudo-Experimenten bestimmen.

Beispiel-1: Zählexperiment mit Poisson-Statistik

- **Hypothesentest**: Kleines Signal s auf bekanntem Untergrund $b \gg s$.

$$\mathcal{P}_0(n,b) = \frac{b^n}{n!}e^{-b} \quad ; \quad \mathcal{P}_1(n,s+b) = \frac{(s+b)^n}{n!}e^{-(s+b)} \qquad n \equiv s+b: \ \text{beob.} \quad \text{Ergeignisse}$$

Anm.:Wir bestimmen s für H_1 a posteriori aus $s \equiv n - b$.

$$\lambda = \frac{P_0(n,b)}{P_1(n,s+b)}$$

$$q = -2\ln\left(\left(\frac{b}{s+b}\right)^n e^{-b+s+b}\right) = -2\left(n\,\ln\left(\frac{b}{s+b}\right)+s\right) = 2\left(\underbrace{(s+b)}_{n\equiv s+b}\ln\left(\frac{s}{b}+1\right)-s\right)$$

mit: $s \ll b$

$$\ln(1+x) = \sum_{\nu=1}^{\infty}\frac{(-1)^{\nu+1}}{\nu}x^\nu = \frac{s}{b} - \frac{1}{2}\frac{s^2}{b^2} + \mathcal{O}\left((s/b)^3\right) \quad \text{für } x = s/b \in (-1\,;\,1]$$

$$q \approx 2\left(\left(s+b\right)\left(\frac{s}{b}-\frac{1}{2}\frac{s^2}{b^2}\right)-s\right) = 2\left(\frac{s^2}{b}+s-\frac{1}{2}\frac{s^3}{b^2}-\frac{1}{2}\frac{s^2}{b}-s\right) = \frac{s^2}{b}-\frac{1}{2}\frac{s^3}{b^2}$$

Beispiel-1: Zählexperiment mit Poisson-Statistik

- Nach dem Theorem von Willks gilt:

(Siehe Folie 41) (Siehe VL-01 Folie 35)

$$q = \chi^2(x, r=1)\Big|_n = \frac{(n-\mu_0)^2}{\sigma_0^2}$$

$$\sqrt{\frac{(n-\mu_0)^2}{\sigma_0^2}} = \boxed{\sqrt{q} \approx \frac{s}{\sqrt{b}}} \qquad \text{(Signifikanz der Beobachtung)}$$

(Siehe Folie 42)

- **NB**: Man erkennt aus dieser Rechnung, dass die altbekannte Formel für die Berechnung der Signifikanz bei einem einfachen Zählexperiment

$$z = \frac{n-b}{\sqrt{b}}$$

nur näherungsweise gilt. Worin besteht diese Näherung?

Beispiel-1: Zählexperiment mit Poisson-Statistik

- Nach dem Theorem von Willks gilt:

(Siehe Folie 41) (Siehe VL-01 Folie 35)

$$q = \chi^2(x, r=1)\Big|_n = \frac{(n-\mu_0)^2}{\sigma_0^2}$$

$$\sqrt{\frac{(n-\mu_0)^2}{\sigma_0^2}} = \boxed{\sqrt{q} \approx \frac{s}{\sqrt{b}}} \qquad \text{(Signifikanz der Beobachtung)}$$

(Siehe Folie 42)

- **NB**: Man erkennt aus dieser Rechnung, dass die altbekannte Formel für die Berechnung der Signifikanz bei einem einfachen Zählexperiment

$$z = \frac{n-b}{\sqrt{b}}$$

nur näherungsweise gilt. Worin besteht diese Näherung? – In der Annahme $\sigma_0 = \sqrt{b}$.

Beispiel-2: Mittelwert normalverteilter Zufallsgrößen

- Wir berechnen den Likelihood-Quotienten aus dem Mittelwert $\overline{x}$ einer Stichprobe normalverteilter $\{x_i\}$ der Länge n und dem Erwartungswert μ_0 auf der Grundgesamtheit:

- Likelihood:

$$\mathcal{L}(\{x_i\},\mu,\sigma) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

- Likelihood-Quotient:

$$\lambda(\{x_i\},\overline{x},\mu_0,\sigma) = \frac{\mathcal{L}(\{x_i\},\mu_0,\sigma)}{\mathcal{L}(\{x_i\},\overline{x},\sigma)} = \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\overline{x})^2\right)\right)$$

$$q(\{x_i\},\overline{x},\mu_0,\sigma) = -2\ln(\lambda) = \frac{n}{\sigma^2}(\mu_0-\overline{x})^2$$

Beispiel-2: Mittelwert normalverteilter Zufallsgrößen

- Wir berechnen den Likelihood-Quotienten aus dem Mittelwert $\overline{x}$ einer Stichprobe normalverteilter $\{x_i\}$ der Länge n und dem Erwartungswert μ_0 auf der Grundgesamtheit:

- Likelihood:

$$\mathcal{L}(\{x_i\},\mu,\sigma) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

- Likelihood-Quotient:

$$\lambda(\{x_i\},\overline{x},\mu_0,\sigma) = \frac{\mathcal{L}(\{x_i\},\mu_0,\sigma)}{\mathcal{L}(\{x_i\},\overline{x},\sigma)} = \exp\left(-\frac{1}{2\sigma^2}\underbrace{\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\overline{x})^2\right)}\right)$$

(Nebenrechnung)

$$q(\{x_i\},\overline{x},\mu_0,\sigma) = -2\ln(\lambda) = \frac{n}{\sigma^2}(\mu_0-\overline{x})^2$$

$$\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\overline{x})^2 = \left(\sum_{i,j=1}^{n}x_ix_j - 2\underbrace{\sum_{i=1}^{n}x_i\mu_0}_{\equiv 2n\overline{x}\mu_0} + n\mu_0^2\right) - \left(\sum_{i,j=1}^{n}x_ix_j - 2\underbrace{\sum_{i=1}^{n}x_i\overline{x}}_{\equiv 2n\overline{x}^2} + n\overline{x}^2\right)$$

$$= n\mu_0^2 + 2n\overline{x}\mu_0 - n\overline{x}^2 = n(\mu_0-\overline{x})^2$$

Beispiel-2: Mittelwert normalverteilter Zufallsgrößen

- Wir berechnen den Likelihood-Quotienten aus dem Mittelwert $\overline{x}$ einer Stichprobe normalverteilter $\{x_i\}$ der Länge n und dem Erwartungswert μ_0 auf der Grundgesamtheit:

- Likelihood:

$$\mathcal{L}(\{x_i\},\mu,\sigma) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

- Likelihood-Quotient:

$$\lambda(\{x_i\},\overline{x},\mu_0,\sigma) = \frac{\mathcal{L}(\{x_i\},\mu_0,\sigma)}{\mathcal{L}(\{x_i\},\overline{x},\sigma)} = \exp\left(-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\mu_0)^2 - \sum_{i=1}^{n}(x_i-\overline{x})^2\right)\right)$$

$$q(\{x_i\},\overline{x},\mu_0,\sigma) = -2\ln(\lambda) = \frac{n}{\sigma^2}(\mu_0-\overline{x})^2$$

- Die Stichprobenverteilung für $\overline{x}$ ist bekannt: $g(\overline{x}) = \varphi(\overline{x},\mu_0,\sigma/\sqrt{n})$. Wir interessieren uns für die Stichprobenverteilung $g(q)$, die wir im folgenden durch Variablentransformation $\overline{x}\to q(\overline{x})$ ermitteln.

Beispiel-2: Mittelwert normalverteilter Zufallsgrößen

- **NB**: Der Wertebereich von $g(q)$ wird <u>zweimal</u> überdeckt, einmal für $\overline{x} > \mu_0$ und einmal für $\overline{x} < \mu_0$. Es gilt also:

$$g(q) = 2\left|\frac{\mathrm{d}\overline{x}}{\mathrm{d}q}\right| g(\overline{x}) \qquad ; \qquad g(\overline{x}) = \varphi(\overline{x}, \mu_0, \sigma/\sqrt{n}) = \sqrt{\frac{n}{2\pi\sigma^2}}\exp\left(-\frac{n}{2\sigma^2}\left(\mu_0 - \overline{x}\right)^2\right)$$

$$q(\overline{x}) = \frac{n}{\sigma^2}\left(\mu_0 - \overline{x}\right)^2 \qquad ; \qquad \overline{x}(q) = \mu_0 - \sqrt{\frac{\sigma^2}{n}}\, q^{\frac{1}{2}} \qquad ; \qquad \left|\frac{\mathrm{d}\overline{x}}{\mathrm{d}q}\right| = \frac{1}{2}\sqrt{\frac{\sigma^2}{n}}\, q^{-\frac{1}{2}}$$

$$g(q) = 2\frac{1}{2}\sqrt{\frac{\sigma^2}{n}}\, q^{-\frac{1}{2}} \cdot \sqrt{\frac{n}{2\pi\sigma^2}}\exp\left(-\frac{n}{2\sigma^2}\left(\overline{x} - \mu_0\right)^2\right)\Bigg|_{\overline{x} = \mu_0 - \sqrt{\frac{\sigma^2}{n}}\, q^{\frac{1}{2}}} = \boxed{\frac{1}{\sqrt{2\pi}} q^{-\frac{1}{2}} \exp\left(-\frac{q}{2}\right)}$$

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

Beispiel-2: Mittelwert normalverteilter Zufallsgrößen

- **NB**: Der Wertebereich von $g(q)$ wird <u>zweimal</u> überdeckt, einmal für $\overline{x} > \mu_0$ und einmal für $\overline{x} < \mu_0$. Es gilt also:

$$g(q) = 2\left|\frac{\mathrm{d}\overline{x}}{\mathrm{d}q}\right| g(\overline{x}) \qquad ; \qquad g(\overline{x}) = \varphi(\overline{x}, \mu_0, \sigma/\sqrt{n}) = \sqrt{\frac{n}{2\pi\sigma^2}}\exp\left(-\frac{n}{2\sigma^2}\left(\mu_0 - \overline{x}\right)^2\right)$$

$$q(\overline{x}) = \frac{n}{\sigma^2}\left(\mu_0 - \overline{x}\right)^2 \qquad ; \qquad \overline{x}(q) = \mu_0 - \sqrt{\frac{\sigma^2}{n}}\,q^{\frac{1}{2}} \qquad ; \qquad \left|\frac{\mathrm{d}\overline{x}}{\mathrm{d}q}\right| = \frac{1}{2}\sqrt{\frac{\sigma^2}{n}}\,q^{-\frac{1}{2}}$$

$$g(q) = 2\frac{1}{2}\sqrt{\frac{\sigma^2}{n}}\,q^{-\frac{1}{2}} \cdot \sqrt{\frac{n}{2\pi\sigma^2}}\exp\left(-\frac{n}{2\sigma^2}\left(\overline{x} - \mu_0\right)^2\right)\Bigg|_{\overline{x}=\mu_0-\sqrt{\frac{\sigma^2}{n}}\,q^{\frac{1}{2}}} = \boxed{\frac{1}{\sqrt{2\pi}}q^{-\frac{1}{2}}\exp\left(-\frac{q}{2}\right)}$$

Zu vergleichen mit: $\chi^2(x, r=1) = \frac{1}{2^{r/2}\Gamma(r/2)}x^{r/2-1}e^{-x/2}\Big|_{n=1}$

$$= \boxed{\frac{1}{\sqrt{2\pi}}q^{-\frac{1}{2}}\exp\left(-\frac{q}{2}\right)}$$

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

Beispiel-2: Mittelwert normalverteilter Zufallsgrößen

- **NB**: Der Wertebereich von $g(q)$ wird <u>zweimal</u> überdeckt, einmal für $\overline{x} > \mu_0$ und einmal für $\overline{x} < \mu_0$. Es gilt also:

$$g(q) = 2\left|\frac{\mathrm{d}\overline{x}}{\mathrm{d}q}\right| g(\overline{x}) \qquad ; \qquad g(\overline{x}) = \varphi(\overline{x}, \mu_0, \sigma/\sqrt{n}) = \sqrt{\frac{n}{2\pi\sigma^2}} \exp\left(-\frac{n}{2\sigma^2}\left(\mu_0 - \overline{x}\right)^2\right)$$

$$q(\overline{x}) = \frac{n}{\sigma^2}\left(\mu_0 - \overline{x}\right)^2 \qquad ; \qquad \overline{x}(q) = \mu_0 - \sqrt{\frac{\sigma^2}{n}}\, q^{\frac{1}{2}} \qquad ; \qquad \left|\frac{\mathrm{d}\overline{x}}{\mathrm{d}q}\right| = \frac{1}{2}\sqrt{\frac{\sigma^2}{n}}\, q^{-\frac{1}{2}}$$

$$g(q) = 2\frac{1}{2}\sqrt{\frac{\sigma^2}{n}}\, q^{-\frac{1}{2}} \cdot \sqrt{\frac{n}{2\pi\sigma^2}} \exp\left(-\frac{n}{2\sigma^2}\left(\overline{x} - \mu_0\right)^2\right)\Bigg|_{\overline{x}=\mu_0-\sqrt{\frac{\sigma^2}{n}}\, q^{\frac{1}{2}}} = \boxed{\frac{1}{\sqrt{2\pi}} q^{-\frac{1}{2}} \exp\left(-\frac{q}{2}\right)}$$

Zu vergleichen mit: $$\chi^2(x, r=1) = \frac{1}{2^{r/2}\Gamma(r/2)} x^{r/2-1} e^{-x/2}\Bigg|_{n=1}$$

$$= \boxed{\frac{1}{\sqrt{2\pi}} q^{-\frac{1}{2}} \exp\left(-\frac{q}{2}\right)}$$

- In diesem Beispiel folgt $g(q)$ der χ^2-Verteilung für $r = 1$ Freiheitsgrad für <u>beliebige</u> Werte von n.

6 Hypothesentests in der modernen Physik

6.4 Goodness-of-fit Tests

Eine spezielle Klasse von Hypothesentests sind Goodness-of-fit (GoF) Tests. Sie dienen dazu, zu überprüfen, ob ein statistisches Modell die Daten einer Messung hinreichend beschreiben kann.

Goodness-of-fit (GoF) Test

- Mit Hilfe des Goodness-of-fit (GoF) Tests überprüfen Sie, ob und in wieweit Ihr Modell die Daten Ihrer Messung beschreiben kann.

Der Goodness-of-fit Test (GoF) basiert auf der Schätzfunktion $t_{\mathrm{GoF}}(\vec{x})$. Die Einschätzung des Tests erfolgt auf Grundlage des p-Wertes (p-value) der Stichprobenverteilung $g(t_{\mathrm{GoF}}(\vec{x}))$

$$p = \int\limits_{t_{\mathrm{GoF}}^{\mathrm{obs}}}^{\infty} g(t'_{\mathrm{GoF}})\,\mathrm{d}t'_{\mathrm{GoF}} \qquad \text{mit } 0 \leq t_{\mathrm{GoF}}(\vec{x}) < \infty$$

mit der Interpretation:

t_{GoF} klein: Das Modell kann die Daten gut beschrieben

t_{GoF} groß: Das Modell kann die Daten nicht gut beschrieben

Der p-Wert bezeichnet die Wahrscheinlichkeit einen Wert von $t_{\mathrm{GoF}} \geq t_{\mathrm{GoF}}^{\mathrm{obs}}$ zu erhalten, wenn das Modell wahr ist.

Vorgehensweise beim GoF Test

- Die Berechnung des p-Wertes impliziert folgendes Vorgehen:
  - (Einmalige) Evaluation von $t_{\mathrm{GoF}}^{\mathrm{obs}}$ auf Daten.
  - Bestimmung von $g(t_{\mathrm{GoF}}(\vec{x}))$ entweder analytisch oder durch vielfache Auswertung von t_{GoF} im Rahmen eines Ensemble Tests.
  - Bestimmung von p .
- **NB**: Diese Vorgehensweise ist die beste und sicherste, um ein Modell gegen eine Messung zu testen.
- In seltenen Fällen ist $g(t_{\mathrm{GoF}}(\vec{x}))$ analytisch bekannt, so z.B. beim χ^2-Test nach Pearson, oder in Fällen in denen das Theorem von Wilks anwendbar ist. In diesen Fällen kann die Einschätzung auch direkt auf Grundlage von $t_{\mathrm{GoF}}^{\mathrm{obs}}$ erfolgen.

Pearson (χ^2)-Test

Die Pearson Teststatistik

$$t_{\chi^2} = \sum_{i \leq n} \frac{(f(x_i) - y_i)^2}{f(x_i)}$$

entspricht dem χ^2-Wert für Poisson-verteilte Zufallsgrößen mit $\sigma_i^2 = f(x_i)$.

- Die $f(x_i)$ entsprechen den Modellvorhersagen, die y_i den Messungen/Beobachtungen.
- t_{χ^2} folgt einer χ^2-Verteilung, d.h. $\frac{t_{\chi^2}}{n} \lesssim 1$ impliziert, dass das Modell die Daten beschreiben kann (siehe VL-04 Folie 18).
- **NB**: Dies ist <u>nicht gleichbedeutend</u> damit, dass das Modell der Wahrheit entspricht.
- Der Pearson Test kann einfach und schnell bei vielen Gelegenheiten angewandt werden.

Beispiel-1: Test auf Unabhängigkeit

- **Hypothese**: Bei einer Umfrage zur Rentenreform soll in Abhängigkeit vom Alter die Zustimmung (-1 ablehnend, 0 neutral +1 zustimmend) erfasst werden. Man vermutet, dass das Alter einen Einfluss auf die Zustimmung hat.

  - H_0 : „Die Häufigkeiten in der folgenden Kontingenztabelle der Grundgesamtheiten unterscheiden sich *nicht* von der Erwartung für statistische Unabhängigkeit “ ($H_0\,:\,p_{ij}=p^{\mathrm{e}}_{ij}$).

  - Testfunktion:

$$t_{\chi^2}=\sum_{i=1}^{m}\sum_{j=1}^{k}\frac{h_{ij}-h^{\mathrm{e}}_{ij}}{h^{\mathrm{e}}_{ij}}\propto\chi^2\left((m-1)\cdot(k-1)\right)=\chi^2(6)$$

  - Irrtumswahrscheinlichkeit: $\alpha=0.05\quad\mathrm{mit:}\int\limits_{-\infty}^{12.6}\chi^2(x,n=6)\,\mathrm{d}x=0.05$

Beispiel-1: Test auf Unabhängigkeit

- Stichprobenergebnis (Befragung von 250 Personen):

Einstellung	Alter			
	- 20	21-40	41-65	66 und älter
-1	5	30	60	30
0	6	24	30	15
+1	9	26	10	5

$$\sum \chi^2 =$$

h_{ij}	h^{e}_{ij}
$\left(h_{ij}-h^{\mathrm{e}}_{ij}\right)^2$	$\frac{\left(h_{ij}-h^{\mathrm{e}}_{ij}\right)^2}{h^{\mathrm{e}}_{ij}}$

Einstellung	Alter			
	-20	21-40	41-65	66 und älter
-1				
0				
+1				

Beispiel-1: Test auf Unabhängigkeit

- Stichprobenergebnis (Befragung von 250 Personen):

Einstellung	Alter			
	- 20	21-40	41-65	66 und älter
-1	5	30	60	30
0	6	24	30	15
+1	9	26	10	5

$$\sum \chi^2 = 28$$

h_{ij}	h^{e}_{ij}
$\left(h_{ij} - h^{e}_{ij}\right)^2$	$\frac{\left(h_{ij} - h^{e}_{ij}\right)^2}{h^{e}_{ij}}$

Einstellung	Alter							
	-20		21-40		41-65		66 und älter	
-1	5	10	30	40	60	50	30	25
	25	2,5	100	2,5	100	2	25	1
0	6	6	24	24	30	30	15	15
	0	0	0	0	0	0	0	0
+1	9	4	26	16	10	20	5	10
	25	6,25	100	6,25	100	5	25	2,5

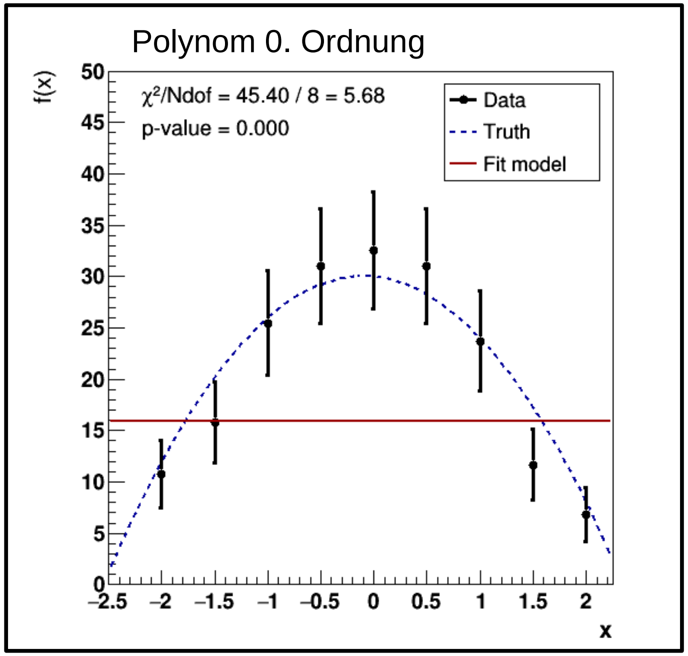
Beispiel: Test auf Unabhängigkeit

- Testentscheidung: H_0 ist widerlegt ($\alpha^* = 0$).

- Es besteht also eine Korrelation zwischen Alter und Zustimmung zur Rentenreform. Welcher Art diese Korrelation ist und wie sie interpretierbar ist, sollte Bestandteil weiterer Untersuchungen sein.

Beispiel-2: Anpassung eines Polynoms

- Anpassung eines Polynoms variierender Ordnung an eine randomisierte Verteilung mit dem zugrundeliegenden Verlauf $f(x) = 30 - x - 5\,x^2$:

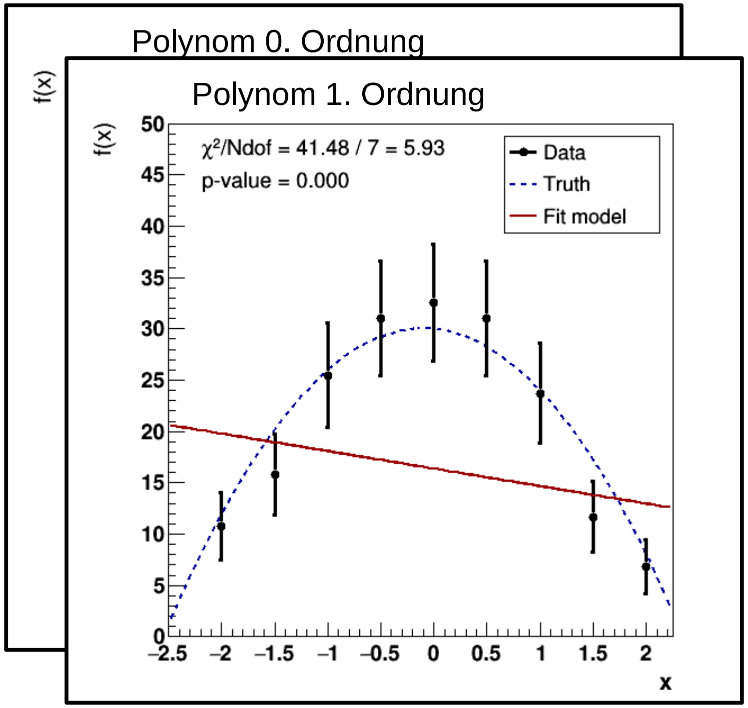


(Vergleiche mit VL-04 Folie 18)

Ein lauffähiges RooT macro finden Sie hier.

Beispiel-2: Anpassung eines Polynoms

- Anpassung eines Polynoms variierender Ordnung an eine randomisierte Verteilung mit dem zugrundeliegenden Verlauf $f(x) = 30 - x - 5\,x^2$:

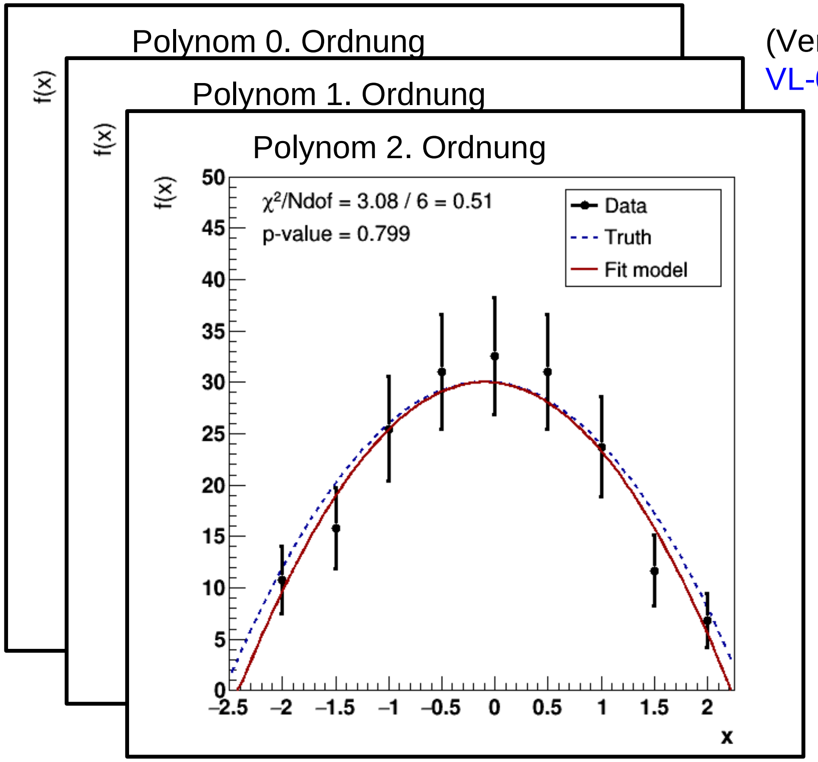


(Vergleiche mit VL-04 Folie 18)

Ein lauffähiges RooT macro finden Sie hier.

Beispiel-2: Anpassung eines Polynoms

- Anpassung eines Polynoms variierender Ordnung an eine randomisierte Verteilung mit dem zugrundeliegenden Verlauf $f(x) = 30 - x - 5\,x^2$:

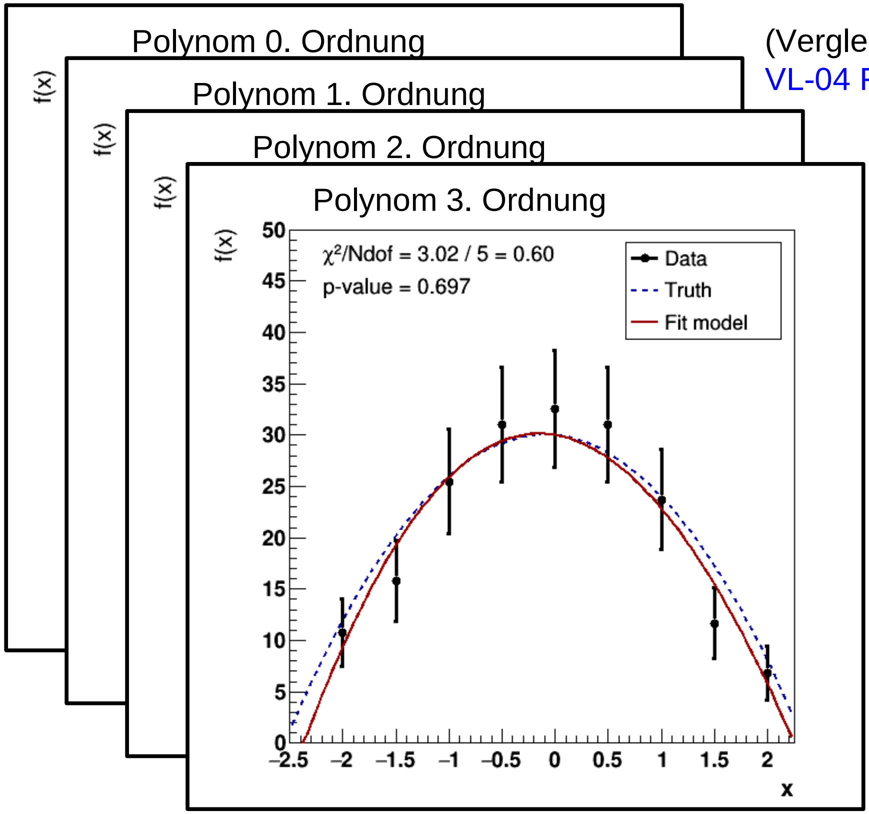


(Vergleiche mit VL-04 Folie 18)

Ein lauffähiges RooT macro finden Sie hier.

Beispiel-2: Anpassung eines Polynoms

- Anpassung eines Polynoms variierender Ordnung an eine randomisierte Verteilung mit dem zugrundeliegenden Verlauf $f(x) = 30 - x - 5\,x^2$:

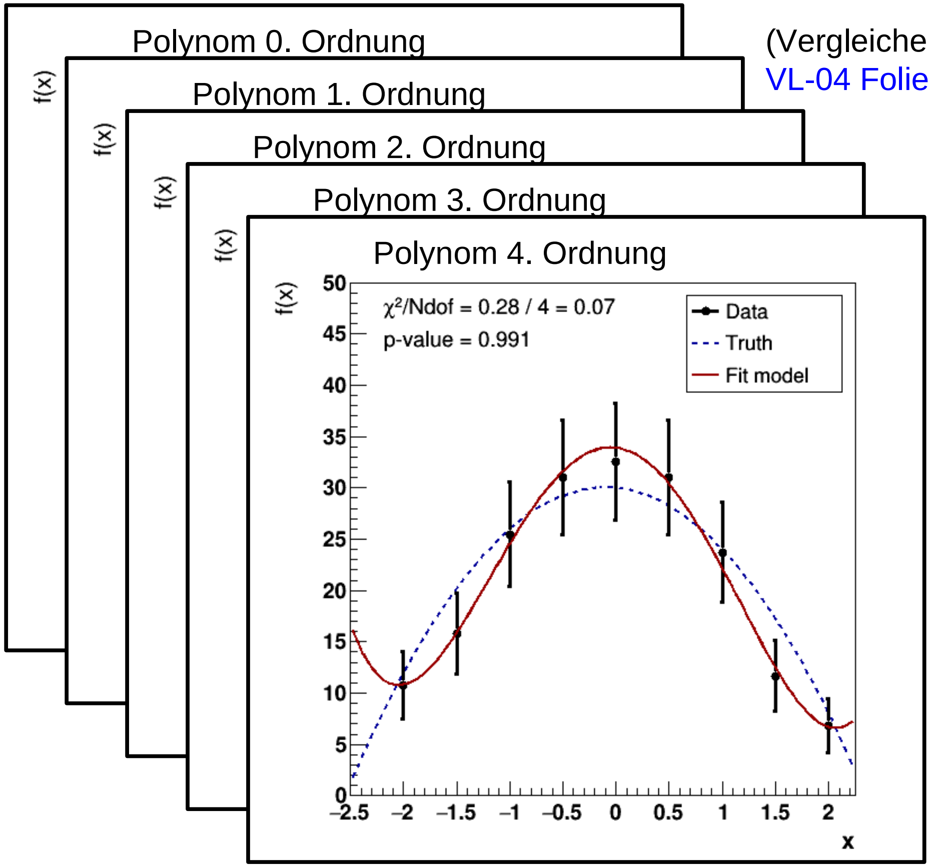


(Vergleiche mit VL-04 Folie 18)

Ein lauffähiges RooT macro finden Sie hier.

Beispiel-2: Anpassung eines Polynoms

- Anpassung eines Polynoms variierender Ordnung an eine randomisierte Verteilung mit dem zugrundeliegenden Verlauf $f(x) = 30 - x - 5\,x^2$:



(Vergleiche mit VL-04 Folie 18)

- Halten Sie bei Anpassungen die Anzahl nicht a priori motivierter Parameter so gering wie möglich.
- Werte von $\chi^2 \ll 1$ können nicht nur auf überschätzte Unsicherheiten sondern auch auf „*overfitting*" hindeuten.
- Beachten Sie: Bei guter Abdeckung Ihrer Unsicherheiten berühren ~32% der Datenpunkte die angepasste Kurve i.a. nicht mir ihren Fehlerbalken (siehe Folie 11).

Ein lauffähiges RooT macro finden Sie hier.

Likelihood-Quotient

- Der sogenannte **saturierte Modelltest** erweitert den χ^2-Test nach Pearson zu einem allgemeinen Likelihood basierten Test.

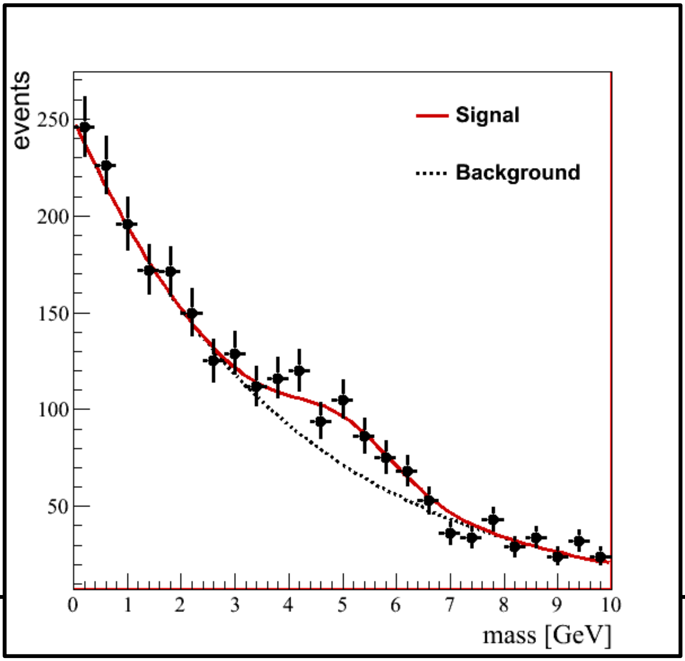
- Die Teststatistik ist der Likelihood-Quotient:

$$\boxed{t_{\mathrm{SAT}} = -2\ln\left(\frac{\mathcal{L}\left(\mathrm{data}|_{\mathrm{test}}\right)}{\mathcal{L}\left(\mathrm{data}|_{\mathrm{SAT}}\right)}\right)}$$

Zu testendes Modell

Modell mit ebenso vielen Parametern wie Messungen (sog. „saturiertes Modell“)

Im rechten Bsp. hätte das saturierte Modell 25 Parameter und würde so durch jeden Messpunkt gehen.



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

Likelihood-Quotient

- Der sogenannte **saturierte Modelltest** erweitert den χ^2-Test nach Pearson zu einem allgemeinen Likelihood basierten Test.

- Die Teststatistik ist der Likelihood-Quotient:

$$\boxed{t_{\mathrm{SAT}} = -2\ln\left(\frac{\mathcal{L}\left(\mathrm{data}|_{\mathrm{test}}\right)}{\mathcal{L}\left(\mathrm{data}|_{\mathrm{SAT}}\right)}\right)}$$

Evaluation der Likelihood unter Annahme normalverteilter Einzelereignisse:

$$\mathcal{L}\left(\mathrm{data}|_{\mathrm{test}}\right) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma_i}}e^{-(y_i-\mu_i)^2/2\sigma_i}$$

$$\mathcal{L}\left(\mathrm{data}|_{\mathrm{SAT}}\right) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma_i}}$$

$$t_{\mathrm{SAT}} = -2\ln\left(\frac{\mathcal{L}\left(\mathrm{data}|_{\mathrm{test}}\right)}{\mathcal{L}\left(\mathrm{data}|_{\mathrm{SAT}}\right)}\right) = \sum_{i=1}^{n}\frac{(y_i-\mu_i)^2}{\sigma_i}$$

- D.h. für Histogramme ohne weitere systematischen Unsicherheiten, für die die Unsicherheiten der Messpunkte <u>normalverteilt</u> sind erhält man $t_{\mathrm{SAT}} \rightarrow t_{\chi^2}$ zurück.

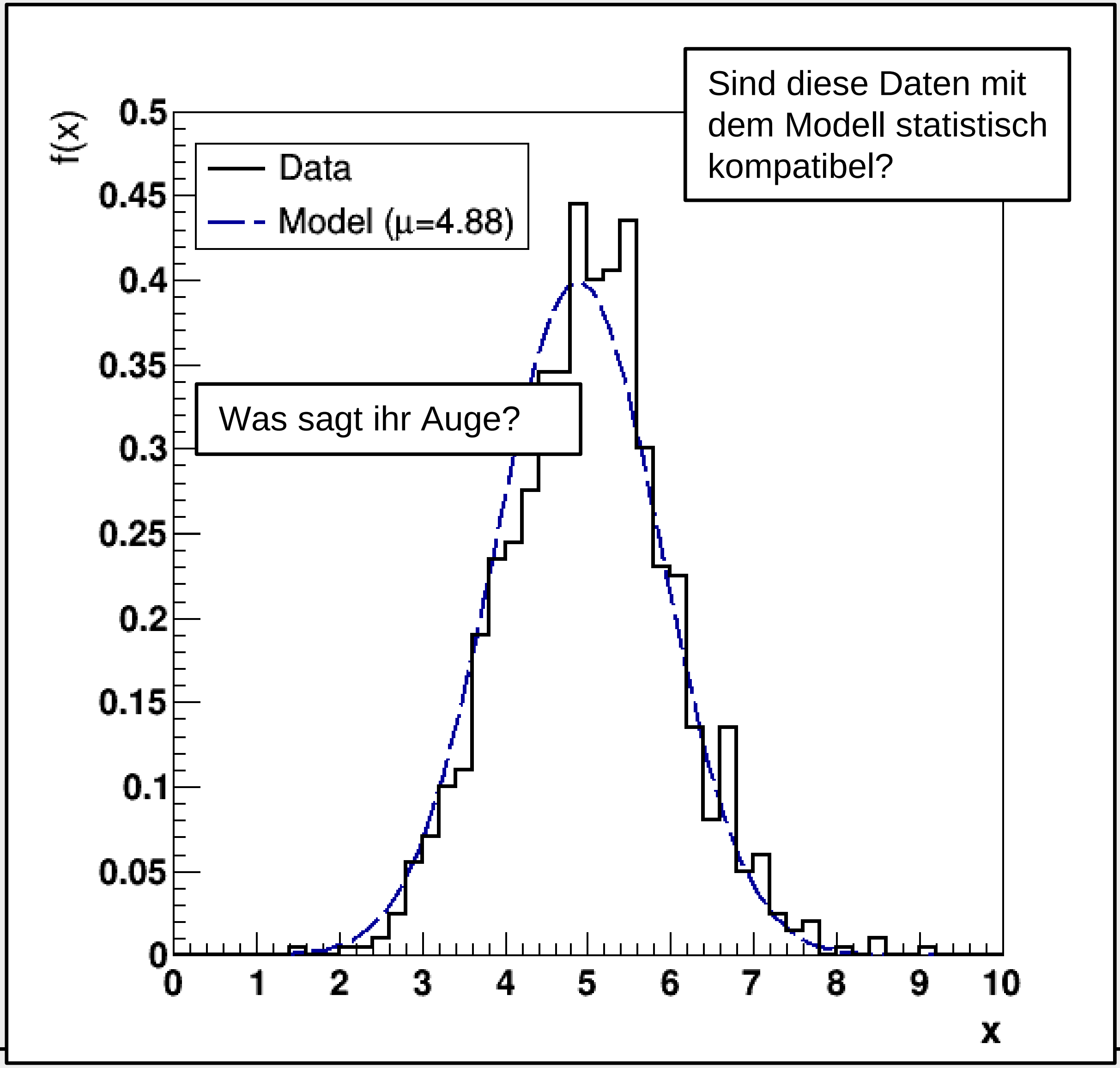
Tests auf der CDF

- Die Teststatistik t_{χ^2} ist nur sensitiv auf die Abstände der Erwartung von den Messpunkten, nicht auf deren „Richtung“. (Warum ist dies so?)

- Es kann somit vorkommen, dass Sie bin-korrelierte Trends nicht mit einem auf t_{χ^2} oder t_{SAT} basierten Test erfassen.

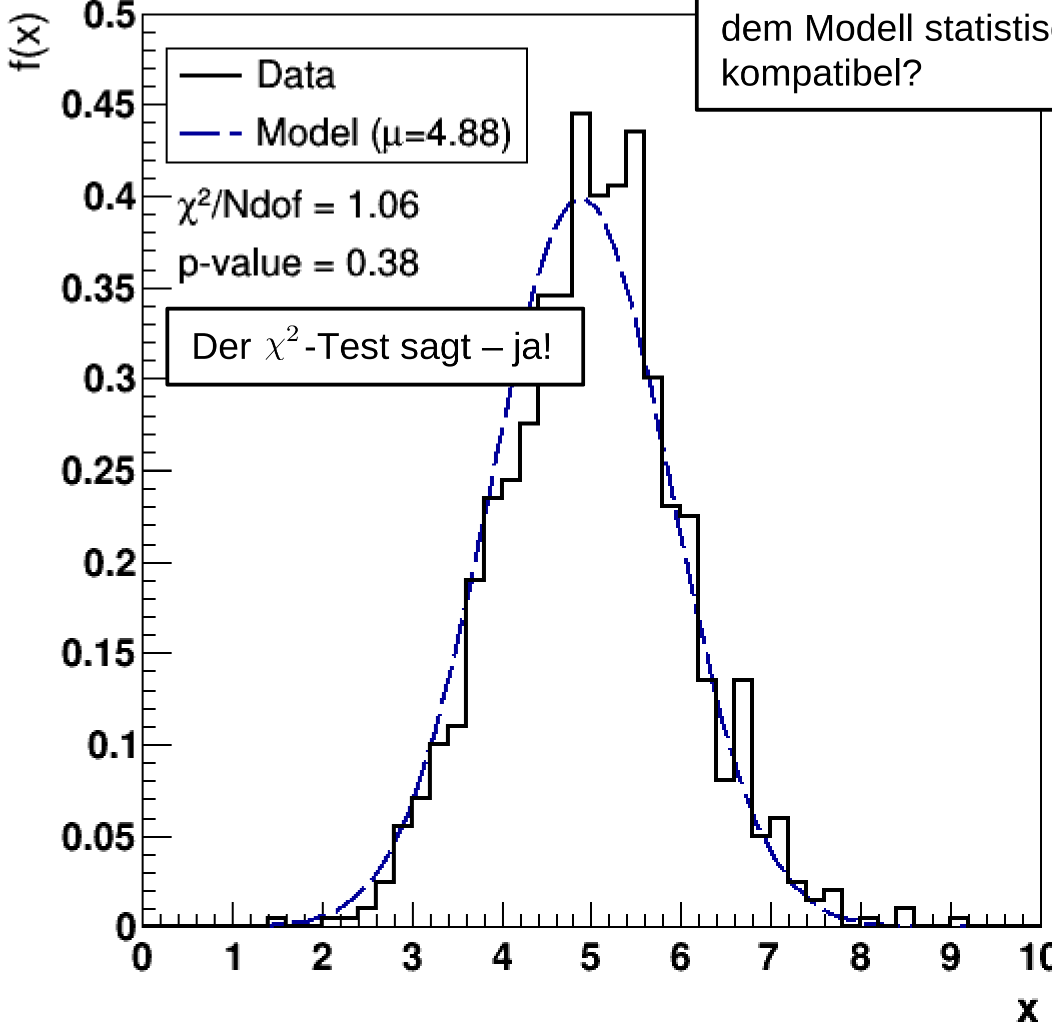
- Dies können Sie mit Tests erreichen, die auf die kummulative (CDF) und die empirische (EDF) Verteilungsfunktion sensitiv sind. Wir diskutieren als Beispiel eines solchen Tests den Kolmogorov-Smirnov-Test (KS).

Vergleich zweier Verteilungsfunktionen



Ein lauffähiges RooT macro finden Sie hier.

Vergleich zweier Verteilungsfunktionen



Sind diese Daten mit dem Modell statistisch kompatibel?

Der χ^2-Test sagt – ja!

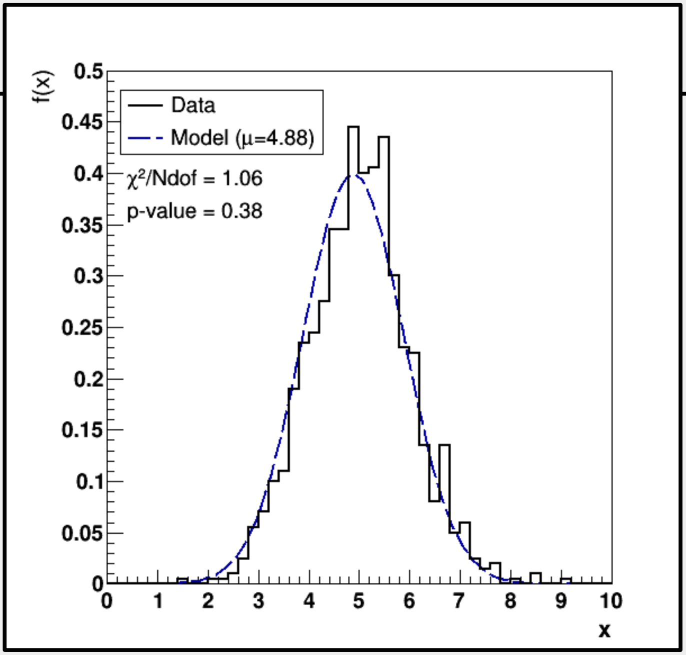
Ein lauffähiges RooT macro finden Sie hier.

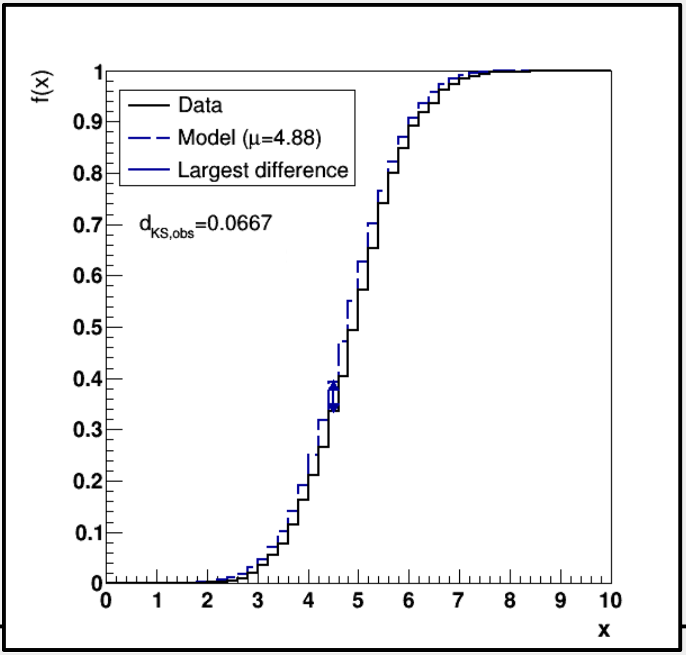
Stat. Kompabilität und Trends

- Für dieses Bsp. wurden die Daten basierend auf einer Normalverteilung mit $\mu = 5$ 1000-mal zufällig ge*sampled*. Der Vergleich erfolgt mit einem Modell mit $\mu = 4.88$.

- Wie Sie sehen, ist das Modell mit $\mu = 4.88$ nach χ^2 mit den Daten kompatibel. Es liegt jedoch, wenn Sie genau hinsehen, auf der linken Flanke systematisch unter- und auf der rechten Flanke oberhalb der Daten.

- Diesen Trend erfassen Sie durch den Vergleich der CDF des Modells mit der EDF der Daten.





Ein lauffähiges RooT macro finden Sie hier.

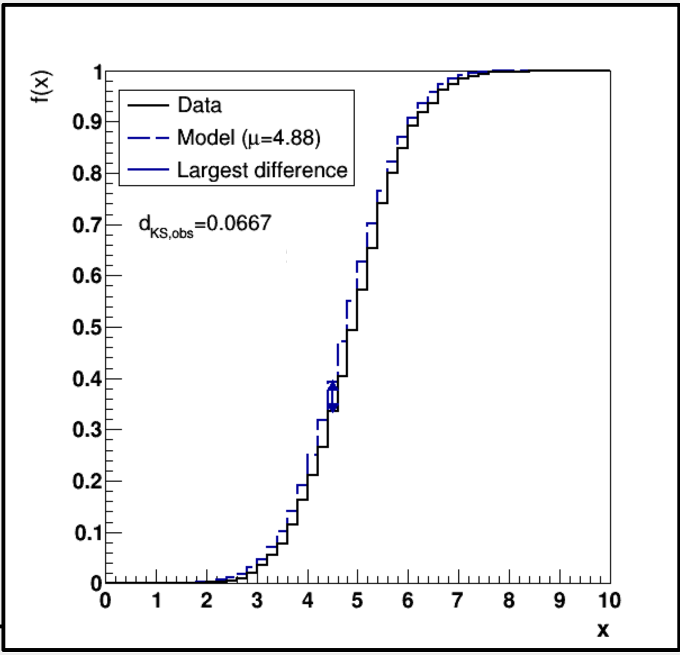
Kolmogorov-Smirnov-Test

Der größte vertikale Abstand zweier kummulativer Verteilungsfunktionen F_0 und F_n

$$d_{\mathrm{KS}} = \|F_n - F_0\|$$

ist die Teststatistik des Kolmogorov-Smirnov-Tests.

- Diesen Trend erfassen Sie durch den Vergleich der CDF des Modells mit der EDF der Daten.
- Den p-Wert für vorgegebene Werte von d_{KS} können Sie aus Tabellen auslesen.



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

Ein lauffähiges RooT macro finden Sie hier.

Kolmogorov-Smirnov-Test

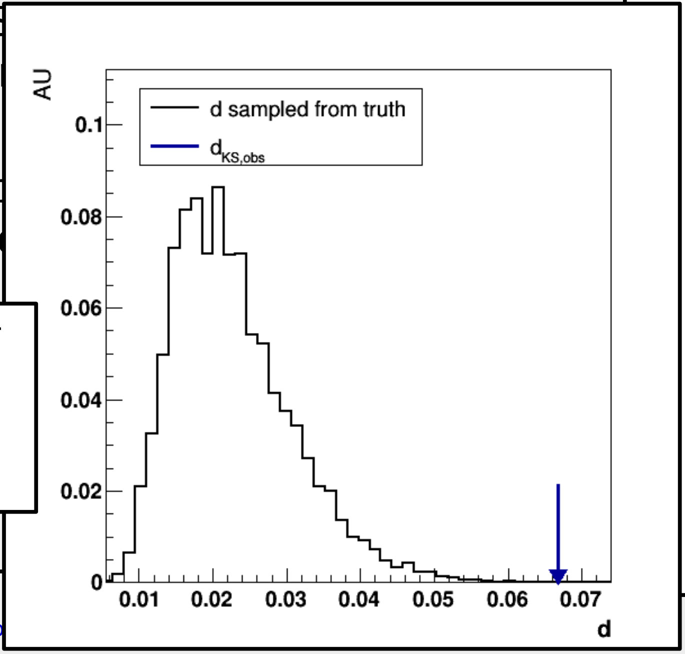
Der größte vertikale Abstand zweier kummulativer Verteilungsfunktionen F_0 und F_n

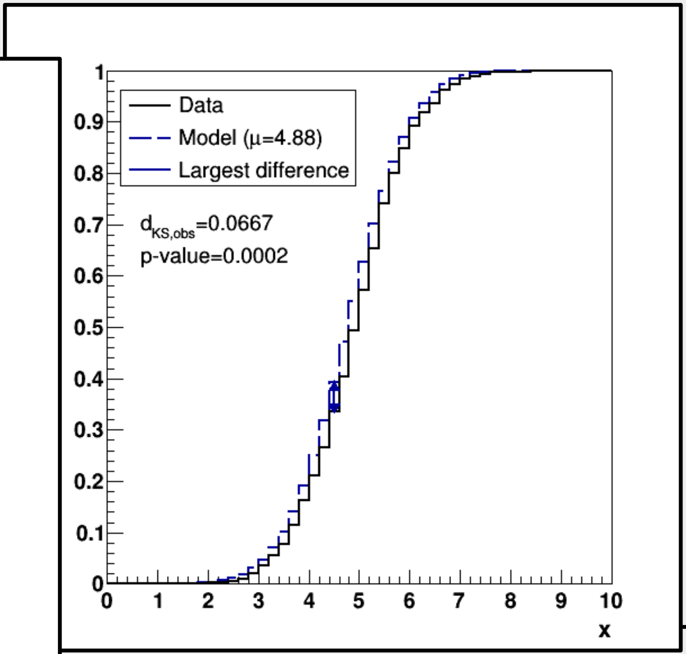
$$d_{\mathrm{KS}} = \|F_n - F_0\|$$

ist die Teststatistik des Kolmogorov-Smirnov-Tests.

- Diesen Trend erfassen S
CDF des Modells mit de
- Den p-Wert für vorgege
Sie aus Tabellen auslese

Die sicherste Art einen p-Wert für einen beobachteten Wert $d_{\mathrm{KS,obs}}$ zu ermitteln, besteht aber auch hier wieder in einem Ensemble Test.



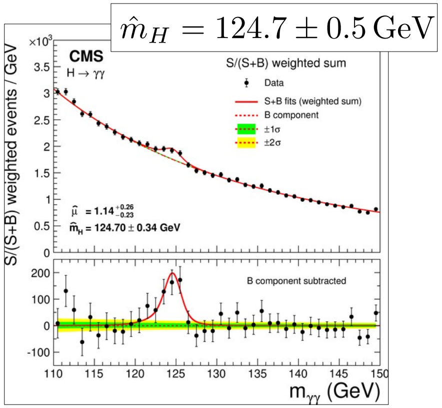


Ein lauffähiges RooT macro finden Sie hier.

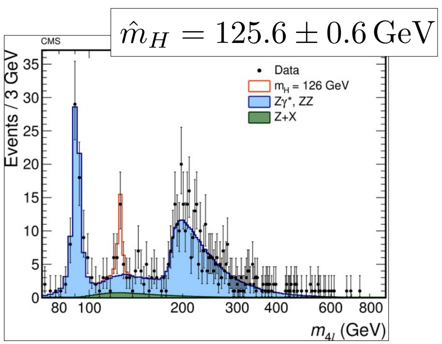
Kompatibilität: Higgs@CMS

Gezeigt sind hier die Run-1 Ergebnisse von CMS.

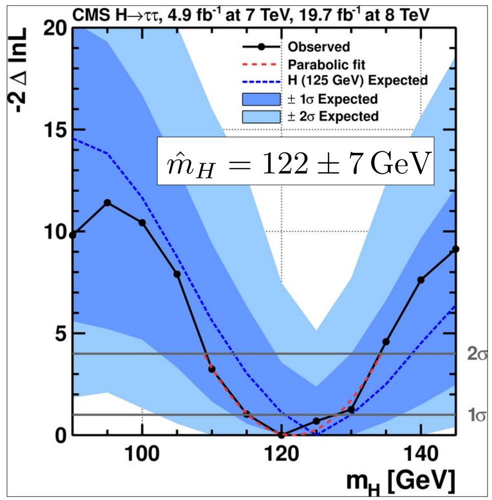
EPJ C 74 (2014) 3076



PRD 89 (2014) 092007

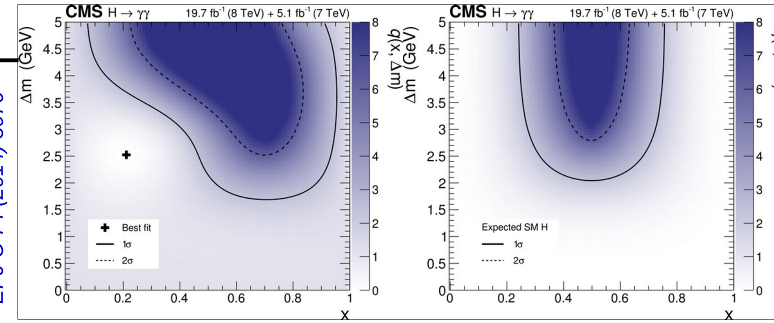


JHEP 05 (2014) 104

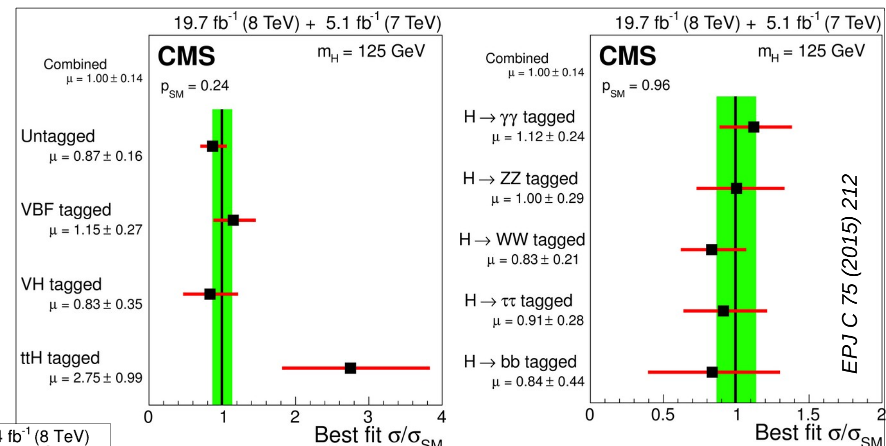


Second close-by resonance in $H \to \gamma\gamma$?

EPJ C 74 (2014) 3076

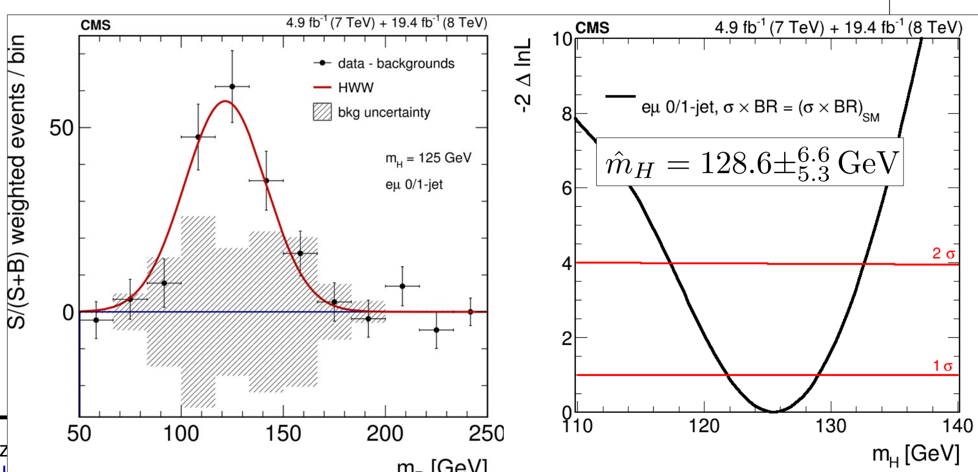


Coupling across production modes or decay channels:



EPJ C 75 (2015) 212

JHEP 01 (2014) 096



EPJ C 75 (2015) 212

- Event categories : 227
- Nuisance parameters: $\mathcal{O}(2500)$
- 16 MB binary file of stat. model (~145 MB in human readable form).

$$\mu = \sigma/\sigma_{SM} = 1.00 \pm 0.14$$

$$p\text{-value} = 84\%$$

Zusammenfassung

- Grundlagen und Begriffe des Hypothesentests.
- Fehler 1. und 2. Art.
- Klassische Beispiele und Anwendungen.
- Goodness of fit (GoF) Test.