Fehler 1. und 2. Art
Vertrauensintervalle

Peter Büchel

HSLU I

Stat: Block 07

Fehler Hypothesentest

- Nullhypothese bei Hypothesentest ist richtig (was aber nicht bekannt)
- Machen n Messungen um Hypothesentest zu überprüfen
- Messungen ergeben extreme Werte
- Durchschnitt x_n liegt im Verwerfungsbereich: Nullhypothese wird verworfen
- Es wurde ein Fehler gemacht: Nullhypothese wird verworfen, obwohl Nullhypothese richtig ist
- Hypothesentest macht keine absolute Aussage, sondern sagt nur das Aussage sehr wahrscheinlich stimmt (p-Wert nahe bei 0)
- Unsicherheit bleibt

Fehler Hypothesentest

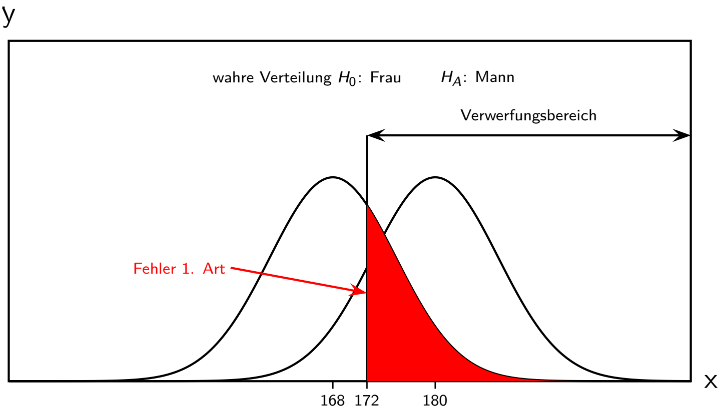
- Schema:

Wahrheit \ Entscheidung	H_0	H_A
H_0	✓	Fehler 1. Art
H_A	Fehler 2. Art	✓

- Entscheidung für H_0, aber H_A wäre richtig $\longrightarrow$ Fehler 2. Art
- Entscheidung für H_A, aber H_0 wäre richtig $\longrightarrow$ Fehler 1. Art

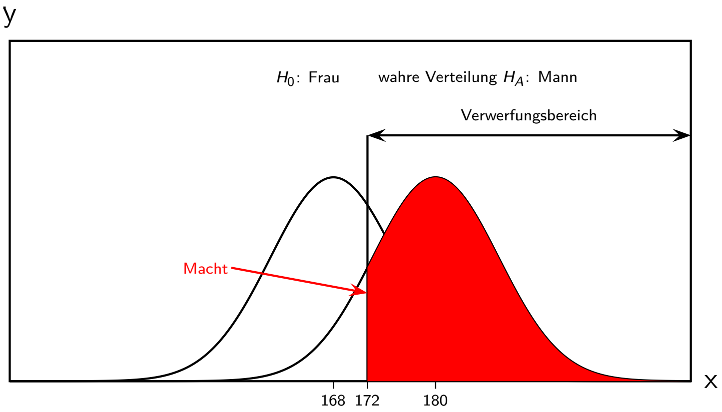
Fehler 1. Art

- Entscheidung für H_A, aber H_0 wäre richtig $\longrightarrow$ Fehler 1. Art
- Entspricht gerade Signifikanzniveau
- Skizze:



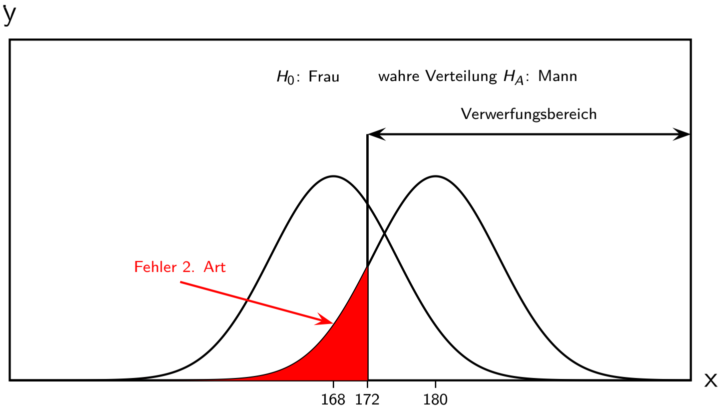
Macht

- H_A wird angenommen und H_A richtig $\quad\rightarrow\quad$ das was wir wollen
- Der wahre Parameter für H_A muss bekannt sein $\quad\rightarrow\quad$ hier $\mu_A = 180$
- Skizze:



Fehler 2. Art

- Entscheidung für H_0, aber H_A wäre richtig $\longrightarrow$ Fehler 2. Art
- Der wahre Parameter für H_A muss bekannt sein $\quad\rightarrow\quad$ hier $\mu_A = 180$
- Fehler 2. Art $= 1 -$ Macht
- Skizze:



Welche Fehlerart ist wichtiger?

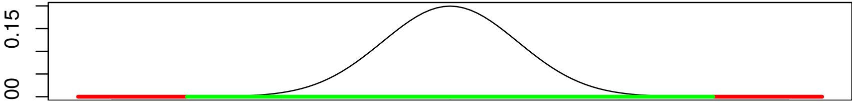
- Fehler 1. Art hat traditionell mehr Gewicht als Fehler 2. Art
- Wissenschaftler arbeiten genau und haben Angst, einen Humbug zu publizieren, der sich dann als falsch herausstellt
- Denn wenn Wissenschaftler einen Effekt (signifikante Abweichung von Nullhypothese) beobachten, möchten sie sicher sein, dass es sich nicht bloss um Zufall handelt
- Fehler 1. Art soll vermieden werden
- Nimmt in Kauf, dass man manchmal wichtigen Effekt verpasst
- Fehler 2. Art ist also zweitrangig

- Fehler 1. Art wird direkt kontrolliert durch Konstruktion eines Tests, indem Signifikanzniveau α möglichst klein gehalten wird
- Über die W'keit eines Fehlers 2. Art keine solche Kontrolle
- Die beiden Fehlerarten konkurrenzieren sich gegenseitig:

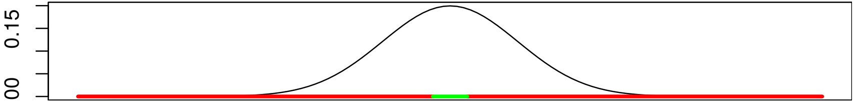
  P(Fehler 2. Art) wird grösser falls α kleiner gewählt wird
- Wahl von α steuert Kompromiss zwischen Fehler 1. und 2. Art
- Weil man aber primär einen Fehler 1. Art vermeiden will, wählt man α klein, z.B. $\alpha = 0.05$
- Je kleiner α, desto kleiner der Verwerfungsbereich
- Vertikale Linie wandert nach rechts $\rightarrow$ Fehler 2. Art wird umso grösser

Wahl von Signifikanzniveau α

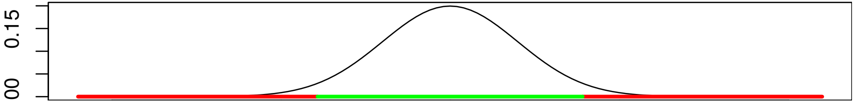
- Graphik: $\alpha = 0.0001$ (nahe bei 0)



- Graphik: $\alpha = 0.8$ (gross)



- Graphik: $\alpha = 0.05$



- Ist α sehr nahe bei null, so Bereich wo *nicht* verworfen wird (grüner Bereich) sehr gross
- D.h.: Es braucht ein sehr Ereignis bis verworfen wird
- Es wird viel zu wenig verworfen
- Im Extremfall $\alpha = 0$: Es wird gar nicht verworfen
- Für α gross: Grüner Bereich sehr klein
- D.h.: Es braucht ein sehr Ereignis bis verworfen wird
- Es wird viel zu wenig verworfen
- Im Extremfall $\alpha = 1$: Es wird immer verworfen
- $\alpha = 0.05$: Kompromiss zwischen den beiden Extremen

Vertrauensintervalle für Normalverteilungen: Einleitung

- Betrachten nochmals Verwerfungsbereich einer normalverteilten Zufallsvariable X mit bekannten σ_X
- Beschränkung vorläufig auf Signifikanzniveau 5 % und zweiseitigem Verwerfungsbereich
- Siehe Jupyter Notebook: `vertrauensintervall_py` (durchmachen)

Vertrauensintervall allgemein

- Das sogenannte *Vertrauensintervall* bei Messdaten besteht aus denjenigen Werten μ, bei denen der entsprechende Test nicht verwirft
- Das sind also alle Parameterwerte des Zufallsmodells, bei denen die Daten recht wahrscheinlich oder plausibel sind
- Dieses Intervall enthält dann das wahre, aber meist unbekannte μ mit einer gegebenen Wahrscheinlichkeit
- Z.B. ist das 95 %-Vertrauensintervall für μ das Intervall, das μ mit einer Wahrscheinlichkeit von 0.95 enthält
- D.h. wenn wir sehr viele Testreihen machen und jeweils das Vertrauensintervall bestimmen, so wird μ in 95 % dieser Intervalle enthalten sein
- Ist das Signifikanzniveau α, so ist nennen wir das Intervall $(1-\alpha)\cdot 100\,\%$-Vertrauensintervall.

Vertrauensintervalle für μ einer Messreihe

- Gehen nun wieder von *Messreihen* aus

- Annahme: Daten Realisierungen von

$$X_1, \ldots, X_n \text{ i.i.d.} \sim \mathcal{N}(\mu, \sigma_X^2)$$

- Müssen wieder unterscheiden, ob σ_X bekannt oder unbekannt ist

Vertrauensintervalle, falls σ_X bekannt

- Diesen Fall schon in Einleitung betrachtet
- Der Mittelwert $\overline{X}_n$ folgt der Verteilung

$$\overline{X}_n \sim \mathcal{N}\left(\mu, \sigma^2_{\overline{X}_n}\right) = \mathcal{N}\left(\mu, \frac{\sigma^2_X}{n}\right)$$

Beispiel

- Schmelzwärme von früher: Normalverteilt mit $\mu = 80$ und $\sigma_X = 0.02$
- Standardabweichung wird hier also als bekannt angenommen
- Mittelwert: $\overline{x}_{13} = 80.02$
- Zweiseitige Vertrauensintervall für Methode A:

$$I = [80.009,\ 80.031]$$

```
from scipy.stats import norm, t
import numpy as np

norm.interval(alpha=0.95, loc=80.02, scale=0.02/np.sqrt(13))
## (80.00912807593181, 80.03087192406818)
```

- Insbesondere liegt 80.00 nicht im Intervall I
- Wert $\mu = 80.00$ ist folglich nicht mit den Daten kompatibel

Vertrauensintervalle, falls σ_X unbekannt

- Ist σ_X unbekannt, so verwenden wir t-Verteilungen und die geschätzte Standardabweichung $\widehat{\sigma}_X$
- Normalverteilung durch t-Verteilung mit Freiheitsgrad $n-1$ ersetzen.

Schmelzwärme Methode A

- Die Standardabweichung lautet

$$\widehat{\sigma}_X = 0.024$$

- Zweiseitige Konfidenzintervall für die mit Methode A gemessene Schmelzwärme:

```
import scipy.stats as st
from scipy.stats import norm, t
import numpy as np

t.interval(alpha=0.95, df=12, loc=80.02, scale=0.024/np.sqrt(13))
## (80.00549694515017, 80.03450305484982)
```

$$I = [80.01, 80.03]$$

- Insbesondere liegt 80.00 nicht im Intervall I

- Der Wert $\mu = 80.00$ ist folglich nicht mit den Daten kompatibel, was wir bereits mit Hilfe des t-Tests ermittelt hatten.