Grundlagen der Biostatistik und Informatik

# Hypothesenprüfungen II.

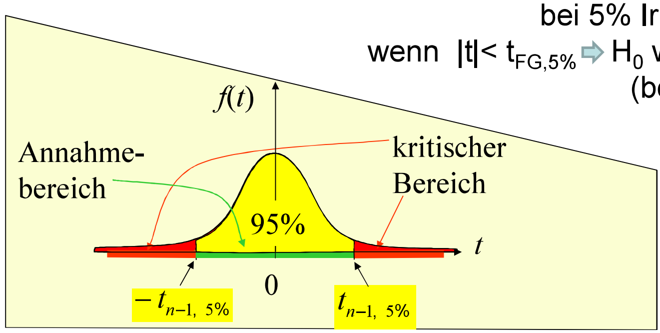
Fehler von erste u. zweite Art,
Anwendungsbedingungen, Verallgemeinerung,
Verhältnis der Schätzungen und Hypothesenprüfungen
Zweistichprobentest

László Smeller

## Widerholung: Grundprinzip der Hypothesenprüfungen

Zur Erinnerung

Zu entscheidende Frage
Indirekter Beweis
Nullhypothese ($H_0$): nur zufällige Änderungen
mathematisch behandelbar
Ein geeigneter Parameter (Prüfgröße) (zB. $t$)
Bei Gültigkeit der $H_0$ $t$ folgt einer gut bestimmten Verteilung
Zu 95% $|t|<t_{FG,5\%}$ ⇨ Wenn $|t|> t_{FG,5\%}$ ⇨ $H_0$ wird abgelehnt bei 5% Irrtumswahrsch.
wenn $|t|< t_{FG,5\%}$ ⇨ $H_0$ wird beibehalten (bei 5% Irrtumsw.).






## Hypothesenprüfung mit Excel

Excel Funktion für *t*-Teste:

ttest(Matrix1; Matrix2; Seiten; Typ)

Messreihe 1
z.B: Temperatur
vor der Eingabe
des Fiebermittels

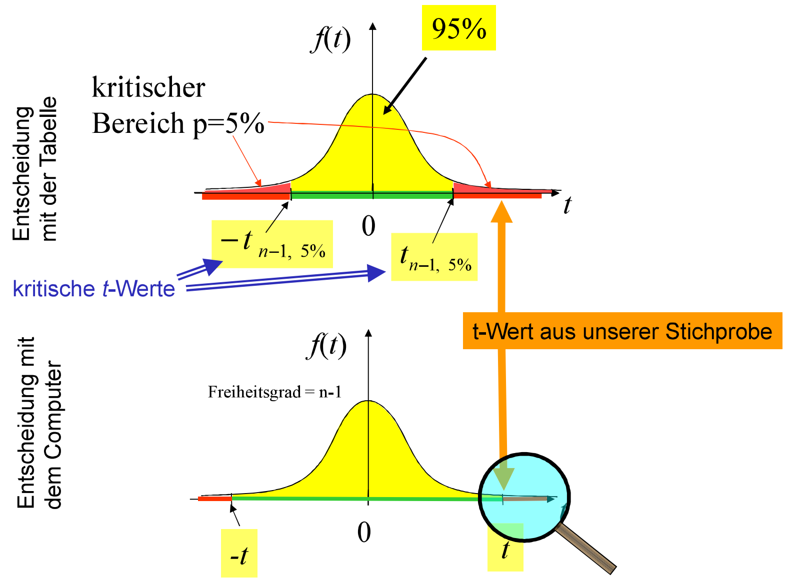
Messreihe 2
z.B: Temperatur
nach der Eingabe
des Fiebermittels

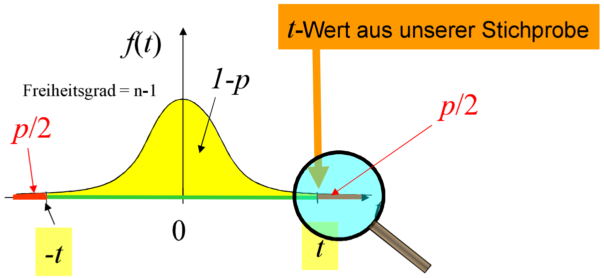
Siehe
später

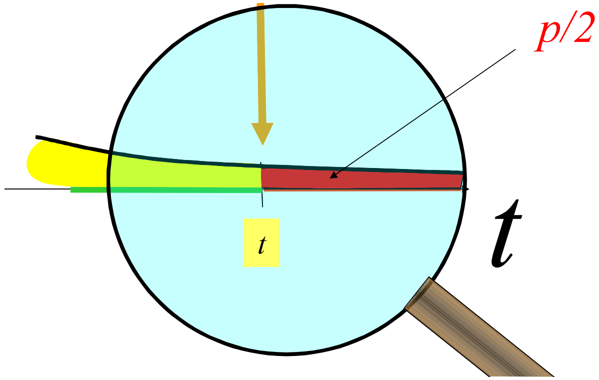
Typ: 1 - gepaart (Eine Stichprobe)
2 - Zwei Stichproben, gleiche Varianz
3 - Zwei Stichproben, ungleiche Varianz

Diese Funktion gibt einen *p* Wert an! (statt ***t***!)

## Die Bedeutung des *p*-Wertes der Excel Funktion






Wenn $t > t_{kritisch}$ dann

p < Signifikanzniveau ($\alpha$)

## Entscheidung mit dem $p$-Wert

1. Fragestellung (mit der Definition der Population!)
   (Bedingung: Normalverteilung)
2. Nullhypothese - Alternativhypothese
3. Festlegung des Signifikanzniveaus ($\alpha$)
4. Messung (Stichprobe mit $n$ Messungen, Repräsentativität!)
5. Berechnung des $p$-Wertes
6. Vergleich von unserem $p$ und dem Signifikanzniveau ($\alpha$)

$p > \alpha$ | $p < \alpha$

7. Die Entscheidung:

| $p > \alpha$ | $p < \alpha$ |
|---|---|
| die **Nullhypothese** wird mit einem $\alpha$ Signifikanzniveau **angenommen**. | die **Nullhypothese** wird mit einem $\alpha$ Signifikanzniveau **Abgelehnt**. |
| ***Anhand unserer Messung kann die Alternativhypothese nicht bewiesen werden.*** | ***Die Alternativhypothese ist angenommen*** **(mit einem Signifikanzniveau von α).** |

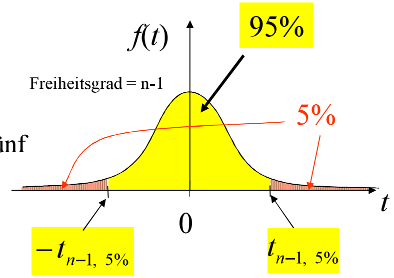
## Die Bedeutung des Signifikanzniveaus

Bei einem unwirksamen Medikament beträgt die Wahrscheinlichkeit dafür, dass $|t| > t_{n-1,\alpha}$ ist, 5%.

(⇨Bei der Untersuchung von hundert unwirksamen Pillen werden zufällig fünf als wirksam gefunden!)

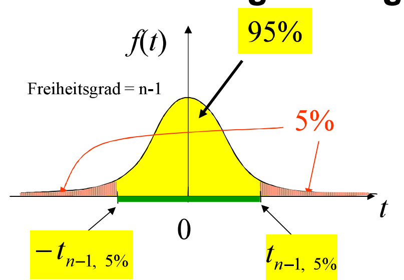
⇩

***Fehler erster Art***

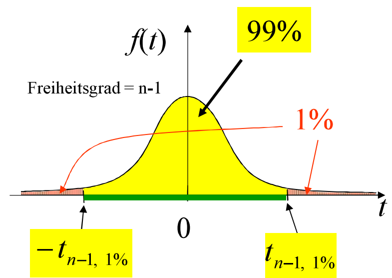



D.h.: wegen des Fehlers erster Art wird eine aus 20 unwirksamen Pillen zufällig und fehlerhaft als wirksam (mit 5% Irrtumswahrscheinlichkeit) gefunden.

## Die Bedeutung des Signifikanzniveaus




Signifikanzniveau=5%




Signifikanzniveau=1%

## Fehler von 1. und 2. Art

***Fehler erster Art ($\alpha$-Fehler):***

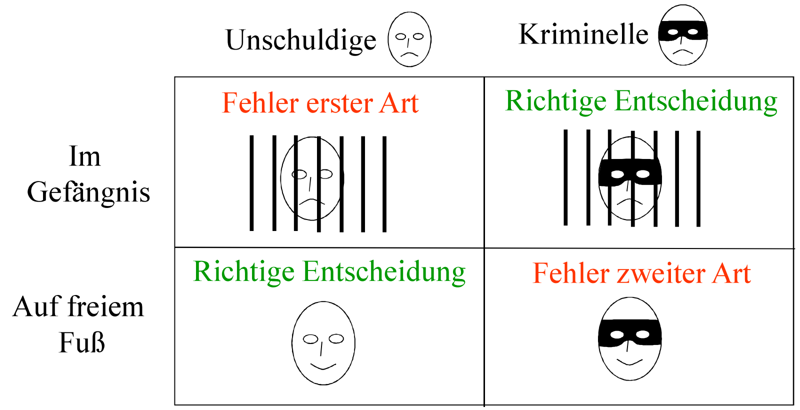
Die Nullhypothese wird zufällig abgelehnt werden, obwohl sie richtig ist!
Die Wahrscheinlichkeit des Fehlers erster Art = Signifikanzniveau
*zB: Unwirksame Pille als wirksam gefunden*

***Fehler zweiter Art ($\beta$-Fehler):***

Die Nullhypothese wird nicht abgelehnt, obwohl sie nicht richtig ist.
Wahrscheinlichkeit =?
*zB: Die Wirkung einer Pille ist so klein, dass man es aus der Messung nicht beweisen kann. ⇨ Man braucht noch mehrere Messungen.*
*⇨ So kleine Wirkung ist oft uninteressant, weil es **klinisch irrelevant** ist.*

## Fehler von 1. und 2. Art

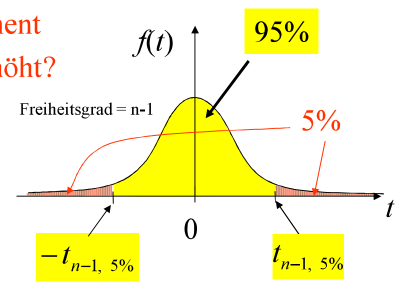
Bei der Gerichtsverhandlung: Nullhypothese: unschuldig



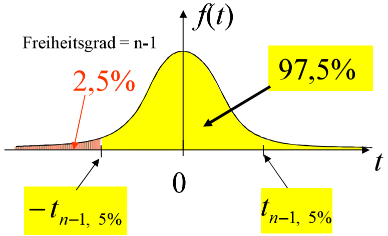

## Einseitige/zweiseitige Teste

Wenn wir ein Fiebermittel testen:
Ist es interessant wenn das Medikament die Körpertemperatur signifikant erhöht?

Zweiseitiger Test:
Nullhyp: das Medikament ändert die Körpertemperatur nicht.

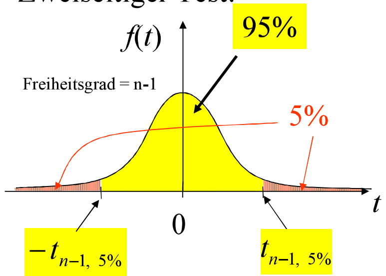



Einseitiger Test
Nullhyp: das Medikament erniedrigt die Körpertemperatur nicht.

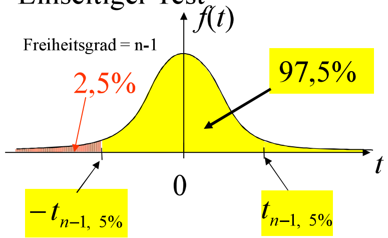



## Einseitige/zweiseitige Teste

Zweiseitiger Test:




Einseitiger Test




| Freiheitsgrad (FG) | $p$ (Irrtumswahrscheinlichkeit, einseitiger Test) | | | | | | |
|---|---|---|---|---|---|---|---|
| | **0,4** | **0,25** | **0,1** | **0,05** | **0,025** | **0,01** | **0,005** |
| | $p$ (Irrtumswahrscheinlichkeit, zweiseitiger Test) | | | | | | |
| | **0,8** | **0,5** | **0,2** | **0,1** | **0,05** | **0,02** | **0,01** |
| **1** | 0,325 | 1,000 | 3,078 | 6,314 | 12,70 | 31,82 | 63,65 |
| **2** | 0,289 | 0,816 | 1,886 | 2,920 | 4,303 | 6,965 | 9,925 |
| **3** | 0,277 | 0,765 | 1,638 | 2,353 | 3,182 | 4,541 | 5,841 |
| **4** | 0,271 | 0,741 | 1,533 | 2,132 | 2,776 | 3,747 | 4,604 |
| **5** | 0,267 | 0,727 | 1,476 | 2,015 | 2,571 | 3,365 | 4,032 |
| **6** | 0,265 | 0,718 | 1,440 | 1,943 | 2,447 | 3,143 | 3,707 |
| **7** | 0,263 | 0,711 | 1,415 | 1,895 | 2,365 | 2,998 | 3,499 |
| **8** | 0,262 | 0,706 | 1,397 | 1,860 | 2,306 | 2,896 | 3,355 |
| **9** | 0,261 | 0,703 | [illegible] | [illegible] | 2,262 | 2,821 | 3,250 |
| **10** | [illegible] | [illegible] | [illegible] | [illegible] | 2,228 | [illegible] | 3,169 |
| **11** | 0,260 | 0,697 | 1,363 | 1,796 | 2,201 | [illegible] | 3,106 |
| **12** | 0,259 | 0,695 | 1,356 | [illegible] | [illegible] | [illegible] | [illegible] |
| **13** | 0,259 | 0,694 | 1,350 | [illegible] | [illegible] | [illegible] | [illegible] |
| **14** | 0,258 | 0,692 | 1,345 | 1,761 | 2,145 | 2,624 | 2,977 |
| **15** | 0,258 | 0,691 | 1,341 | 1,753 | 2,131 | 2,602 | 2,947 |
| **16** | 0,258 | 0,690 | 1,337 | 1,746 | 2,120 | 2,583 | 2,921 |
| **17** | 0,257 | 0,689 | 1,333 | 1,740 | 2,110 | 2,567 | 2,898 |
| **18** | 0,257 | 0,688 | 1,330 | 1,734 | 2,101 | 2,552 | 2,878 |
| **19** | 0,257 | 0,688 | 1,328 | 1,729 | 2,093 | 2,539 | 2,861 |
| **20** | 0,257 | 0,687 | 1,325 | 1,725 | 2,086 | 2,528 | 2,845 |
| **21** | 0,257 | 0,686 | 1,323 | 1,721 | 2,080 | 2,518 | 2,831 |

T.INV.2S(0,05;8)

T.INV(0,025;8)

## Effekt des Stichprobenumfanges

Erhöhung des Stichprobenumfanges:

wenn n $\rightarrow\infty$ dann $s_{\bar{x}} \rightarrow 0$

Wenn $H_0$ ist ungültig, $t$ steigt mit Erhöhung des $n$-es, und $H_0$ wird mit kleinerem Irrtumswahrscheinlichkeit abgelehnt:

$$t = \frac{\bar{x}}{s_{\bar{x}}} \xrightarrow{n\rightarrow\infty} \frac{\mu_1}{0} = \infty$$

Erhöhte Stichprobenumfang führt zu bessere (sicherere) Entscheidung

## Verallgemeinerung: $\mu_0 \neq 0$

Beispiel:

Eine Maschine stellt Pillen mit einem nominalen Wirkstoffsgehalt von 20mg her. Man misst 10 Tabletten und die Wirkstoffgehalte sind (in mg):

20,1 19,8 19,5 17,9 18,8 19,9 18,6 20,3 19,2 19,3

Durchschnitt 19,34 mg, Standardabweichung 0,74 mg, Standardfehler 0,24 mg

Nullhypothese: $\mu_0$ = 20 mg

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$$

Alle weitere Schritte sind wie früher.

$t = -2,80$

## Vergleich von Hypothesenprüfungen und Schätzungen

zB.: Blutdrucksenker: Blutdruckänderungen (mmHg):

-13, 5, -29, -22, 13, -8, -19, -12

Durchschnitt: -10,625 mmHg

Standardfehler: 4,917 mmHg

n=8, FG=7

Schätzung: Konfidenzintervall:

$\bar{x} \pm 2s_{\bar{x}}$ = -10,6±9,8 mmHg -20,4 … – 0,8 mmHg

enthält Null nicht! => Blutdrucksänkender Effekt!

t-Test:

$t$ = -10,625/4,917=-2,161 $|t| < t_{FG=7\,;5\%}$=2,365

?

kein signifikanter Effekt!



## Genaues Konfidenzintervall

$\bar{x} \pm 2s_{\bar{x}}$ ist nur eine **grobe** Annäherung des Konfidenzintervalles.

Das **genaue** Konfidenzintervall für 95% Konfidenzniveau ist:

$$\bar{x} \pm t_{n-1;5\%} s_{\bar{x}}$$

Es zählt nur bei kleinen Stichproben. (n<20)

| Freiheits-grad (FG) | p (Irrtumswahrscheinlichkeit, einseitiger | | | | | |
|---|---|---|---|---|---|---|
| | 0,4 | 0,25 | 0,1 | 0,05 | 0,025 | 0, |
| | p (Irrtumswahrscheinlichkeit, zweiseitige | | | | | |
| | 0,8 | 0,5 | 0,2 | 0,1 | 0,05 | 0, |
| **1** | 0,325 | 1,000 | 3,078 | 6,314 | 12,70 | 31 |
| **2** | 0,289 | 0,816 | 1,886 | 2,920 | 4,303 | 6,9 |
| **3** | 0,277 | 0,765 | 1,638 | 2,353 | 3,182 | 4,5 |
| **4** | 0,271 | 0,741 | 1,533 | 2,132 | 2,776 | 3,7 |
| **16** | 0,258 | 0,690 | 1,337 | 1,746 | 2,120 | 2,5 |
| **17** | 0,257 | 0,689 | 1,333 | 1,740 | 2,110 | 2,5 |
| **18** | 0,257 | 0,688 | 1,330 | 1,734 | 2,101 | 2,5 |
| **19** | 0,257 | 0,688 | 1,328 | 1,729 | 2,093 | 2,5 |
| **20** | 0,257 | 0,687 | 1,325 | 1,725 | 2,086 | 2,5 |
| **21** | 0,257 | 0,686 | 1,323 | 1,721 | 2,080 | 2,5 |
| **22** | 0,256 | 0,686 | 1,321 | 1,717 | 2,074 | 2,5 |
| **23** | 0,256 | 0,685 | 1,319 | 1,714 | 2,069 | 2,5 |
| **24** | 0,256 | 0,685 | 1,318 | 1,711 | 2,064 | 2,4 |
| **25** | 0,256 | 0,684 | 1,316 | 1,708 | 2,060 | 2,4 |
| **26** | 0,256 | 0,684 | 1,315 | 1,706 | 2,056 | 2,4 |
| **27** | 0,256 | 0,684 | 1,314 | 1,703 | 2,052 | 2,4 |
| **28** | 0,256 | 0,683 | 1,313 | 1,701 | 2,048 | 2,4 |
| **29** | 0,256 | 0,683 | 1,311 | 1,699 | 2,045 | 2,4 |
| **30** | 0,256 | 0,683 | 1,310 | 1,697 | 2,042 | 2,4 |
| **40** | 0,255 | 0,681 | 1,303 | 1,684 | 2,021 | 2,4 |
| **60** | 0,255 | 0,679 | 1,296 | 1,671 | 2,000 | 2,3 |
| **120** | 0,254 | 0,677 | 1,289 | 1,658 | 1,980 | 2,3 |
| ∞ | 0,250 | 0,674 | 1,282 | 1,645 | 1,960 | 2,3 |

## Genaues Konfidenzintervall beim Blutdrucksenker

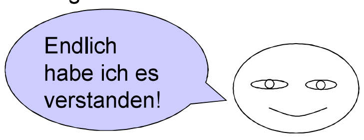
In dem Beispiel des Blutdrucksenkers:

$$\bar{x} \pm t_{n-1;5\%} s_{\bar{x}} =$$

$$\bar{x} \pm t_{7;5\%} s_{\bar{x}} =$$

(-10,6±2,365·4,917) mmHg =
(-10,6±11,6) mmHg
d.h. μ ist in: -22,2 … 0,8 mmHg
mit 95% Wahrscheinlichkeit
=> μ kann 0 sein.

Die Schätzung und der *t*-Test geben dieselbe Ergebnisse!




| Freiheits-grad (FG) | *p* (Irrtumswahrscheinlichkeit, einseitiger Tes | | | | | |
|---|---|---|---|---|---|---|
| | **0,4** | **0,25** | **0,1** | **0,05** | **0,025** | **0,01** |
| | *p* (Irrtumswahrscheinlichkeit, zweiseitiger Te | | | | | |
| | **0,8** | **0,5** | **0,2** | **0,1** | **0,05** | **0,02** |
| **1** | 0,325 | 1,000 | 3,078 | 6,314 | 12,70 | 31,82 |
| **2** | 0,289 | 0,816 | 1,886 | 2,920 | 4,303 | 6,965 |
| **3** | 0,277 | 0,765 | 1,638 | 2,353 | 3,182 | 4,541 |
| **4** | 0,271 | 0,741 | 1,533 | 2,132 | 2,776 | 3,747 |
| **5** | 0,267 | 0,727 | 1,476 | 2,015 | 2,571 | 3,365 |
| **6** | 0,265 | 0,718 | 1,440 | 1,943 | 2,447 | 3,143 |
| **7** | 0,263 | 0,711 | 1,415 | 1,895 | 2,365 | 2,998 |
| **8** | 0,262 | 0,706 | 1,397 | 1,860 | 2,306 | 2,896 |
| **9** | 0,261 | 0,703 | 1,383 | 1,833 | 2,262 | 2,821 |
| **10** | 0,260 | 0,700 | 1,372 | 1,812 | 2,228 | 2,764 |
| **11** | 0,260 | 0,697 | 1,363 | 1,796 | 2,201 | 2,718 |
| **12** | 0,259 | 0,695 | 1,356 | 1,782 | 2,179 | 2,681 |
| **13** | 0,259 | 0,694 | 1,350 | 1,771 | 2,160 | 2,650 |
| **14** | 0,258 | 0,692 | 1,345 | 1,761 | 2,145 | 2,624 |
| **15** | 0,258 | 0,691 | 1,341 | 1,753 | 2,131 | 2,602 |
| **16** | 0,258 | 0,690 | 1,337 | 1,746 | 2,120 | 2,583 |
| **17** | 0,257 | 0,689 | 1,333 | 1,740 | 2,110 | 2,567 |
| **18** | 0,257 | 0,688 | 1,330 | 1,734 | 2,101 | 2,552 |
| **19** | 0,257 | 0,688 | 1,328 | 1,729 | 2,093 | 2,539 |
| **20** | 0,257 | 0,687 | 1,325 | 1,725 | 2,086 | 2,528 |



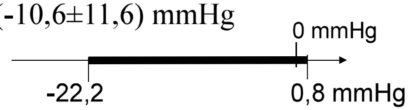
## Vergleich von Hypothesenprüfungen und Schätzungen: p-Wert oder Konfidenzintervall?

p-Wert — Konfidenzintervall

**Blutdrucksenker**

$|t| < t_{7;5\%}$ oder $p > 5\%$ — (-10,6±11,6) mmHg

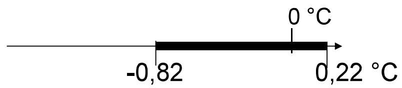
$H_0$ wird angenommen




**Fiebermittel:**

$|t| < t_{4;5\%}$ oder $p > 5\%$ — (-0,30±0,52)°C

$H_0$ wird angenommen




**Kniebeugungen**

$|t| > t_{5;5\%}$ oder $p < 5\%$ — (13,0±4,6)1/Min

$H_0$ wird abgelent




## Placeboeffekt

Placebo: Scheinmedikament, d.h. Pille ohne Wirkstoff

Placeboeffekt: Positive psychische und körperliche Reaktion, die nicht auf die spezifische Wirksamkeit einer Behandlung zurückzuführen ist, sondern auf den psychosozialen Kontext der Behandlung.

Vergleich von zwei Gruppen: Zweistichprobentest