Hypothesenprüfungen

Dr László Smeller

Vergleich der Schätzungen und Hypothesenprüfungen

Schätzungen:

Frage: Wie groß (ist eine physikalische Größe) $\mu=?$

z.B.: Körperhöhe, Blutdruck, Blutzuckerkonzentration...

Antwort: Punktschätzung: Ein Wert

Intervallschätzung: Ein Intervall + Konfidenzniveau (Sicherheitswahrscheinlichkeit)

Hypothesenprüfungen:

Frage: Eine Entscheidungsfrage (ist es wahr oder nicht?)

zB: hat ein Medikament eine Wirkung oder nicht?

Mathematisch: ist $\mu=\mu_0$?

Antwort: Ja oder Nein + Konfidenzniveau (Sicherheitswahrsch.)

Signifikanzniveau (Irrtumswahrsch.)

Typische Aufgaben der Hypothesenprüfung

1. Hat ein Medikament/Behandlung eine Wirkung?
   - 1a. Verursacht es eine Änderung (zB. Blutdruckänderung, d.h.: ist der Blutdruck kleiner nach der Eingabe?)
   - 1b. Gibt es einen Unterschied zwischen den unbehandelten und behandelten Gruppen?
2. Gibt es eine Korrelation zwischen zwei Parametern (zB. Körperhöhe und Gewicht, …)
3. Gibt es eine Korrelation zwischen zwei Eigenschaften (die kategorisierbare sind, zB: Raucher – Nichtraucher, Lungenkrebs - kein Lungenkrebs)

Typische Fragen - typische Größen

1. Hat ein Medikament/Behandlung eine Wirkung?

   Änderung von einer **numerischen (kontinuierlichen) Größe**

   (zB. Blutdruck, Körpertemperatur, Blutzuckerkonzentration, …)

   - 1a. Änderung nach einem Einfluss an einer Stichprobe
   - 1b. Unterschied zwischen zwei Stichproben
2. Gibt es eine Korrelation zwischen **zwei numerischen Größen**

   (zB. Körperhöhe und Gewicht, …)
3. Gibt es eine Korrelation zwischen zwei (oder mehreren) **kategorischen Merkmalen** (zB: Raucher – Nichtraucher, Lungenkrebs-kein Lungenkrebs)

Grundprinzip der Hypothesenprüfungen

Die Wirkung
Die Korrelation } können unterschiedlich sein

⇩

unprognostizierbar

⇩

mathematisch unbehandelbar

☹

Keine Wirkung
Keine Korrelation } nur zufällige Effekte treten auf

⇩

mathematisch gut beschreibbar!

☺

Die Nullhypothese und die Alternativhypothese

Nullhypothese (H_0):

Es gibt **keine Wirkung** $\mu=\mu_0$

Alle **Abweichungen** von dem theoretischen Wert sind rein **zufällig**.

Alternativhypothese (H_1)

Es gibt eine Wirkung $\mu\neq\mu_0$

Die Abweichungen sind nicht zufällig, sondern systematisch!

Eine von H_0 und H_1 wird unbedingt auftreten! $p(H_0 \text{ oder } H_1)=1$

Grundprinzip der Hypothesenprüfungen

Sei es vorausgesetzt, dass wir keine Wirkung haben! (H_0)

Wenn unsere Ergebnisse dieser Voraussetzung nicht entsprechen, dann haben wir wahrscheinlich eine Wirkung/Korrelation.

Keine Wirkung:
zB: Fiebermittel: Wenn es keine Wirkung gibt, ist die Temperaturänderung nach der Eingabe = 0.

1. Beispiel: Fibermittel

Seien die Temperaturen vor und nach der Eingabe gemessen.
Die Messergebnisse (in °C):

Temperatur-änderung

T_{vor}	T_{nach}	$x=T_{nach}-T_{vor}$
39,7	39,2	-0,5
38,8	38,4	-0,4
37,9	38,7	0,8
39,2	38,7	-0,5
Durchschnitt $\bar{x}$		-0,15

Nullhypothese: das Fiebermittel is unwirksam.
Die Temperaturänderungen (x_i) sind zufällig.
Der Erwartungswert der Temperaturänderungen ist Null.

1. Beispiel: Fibermittel

Die Nullhypothese entspricht μ = 0

Wenn die Nullhypothese gültig ist, befindet sich $\bar{x}$ nicht weit von μ.

Ist $\bar{x} = -0{,}15$ klein genug, um die Nullhypothese anzunehmen?

oder

Ist $\bar{x} = -0{,}15$ groß genug, um die Nullhypothese abzulehnen?

Aber wo ist die Grenze? Wie groß muss der Durchschnitt sein, um die Nullhypothese abzulehnen?

Bemerkung: Auch wenn die Nullhypothese gültig ist, kann zufällig $\bar{x}$ sehr groß sein. Aber mit einer sehr kleinen Wahrscheinlichkeit!

1. Beispiel: Fibermittel

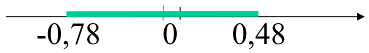
Schätzung des Erwartungswertes:

μ befindet sich mit 95% Wahrscheinlichkeit im Konfidenzbereich von $\bar{x} \pm 2s_{\bar{x}}$, oder genauer in: $\bar{x} \pm t_{n-1,\ 5\%} s_{\bar{x}}$.

In diesem Beispiel: $s = 0{,}635$ °C $\quad s_{\bar{x}} = 0{,}317$ °C

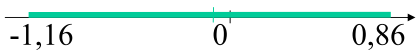
Konfidenzintervall für den Erwartungswert:

grob: $\bar{x} \pm 2s_{\bar{x}} = -0{,}15 \pm 0{,}63$



genau:

$\bar{x} \pm t_{n-1,\ 5\%} s_{\bar{x}} = -0{,}15 \pm 1{,}01$



⇒ μ kann 0 sein! Die Nullhypothese kann nicht abgelehnt werden.

Es ist nicht unwahrscheinlich, dass die H_0 gültig ist.

2. Beispiel: Kniebeugungen

Pulszahl vor und nach 10 Kniebeugungen.
Wird die Pulszahl geändert nach der Kniebeugungen?
H_0: keine Änderung μ=0

p_{vor}	p_{nach}	x=Δp
65	79	14
68	77	9
72	91	19
63	70	7
74	88	14
69	84	15
	Durchsch.	13
	Stabw.	4,34
	Stfehler	1,77

95% Konfidenzintervall für μ:

$13{,}0 \pm 2{,}57 \cdot 1{,}77\ ^1/_{\text{min}}$

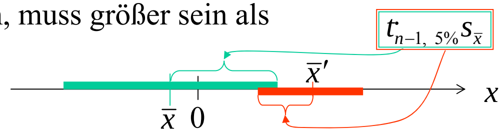
$13{,}0 \pm 4{,}5\ ^1/_{\text{min}}$

$(t_{5;\ 5\%} = 2{,}57)$

0 liegt nicht in diesem Bereich, d.h. wahrscheinlich: μ≠0
Die Nullhypothese kann abgelehnt werden. 95% !

Der t-Wert

Die Abweichung, die nicht mehr als zufällig betrachtet werden kann, muss größer sein als



Eine neue Größe wird definiert: $t = \dfrac{\bar{x}}{s_{\bar{x}}}$ ($\bar{x}$ in $s_{\bar{x}}$ Einheiten gemessen)

oder mathematisch:

$$t = \frac{\bar{x}}{s_{\bar{x}}} = \frac{\dfrac{\sum_{i=1}^{n} x_i}{n}}{\dfrac{s}{\sqrt{n}}} = \frac{\dfrac{\sum_{i=1}^{n} x_i}{n}}{\dfrac{\sqrt{\dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}}{\sqrt{n}}} = f(x_1, x_2, \ldots, x_n)$$

t für unseren „Fibermittel“:
$t = -0{,}15/0{,}317 = -0{,}47$
für Kniebeugungen $t' = 7{,}34$

Der t-Wert

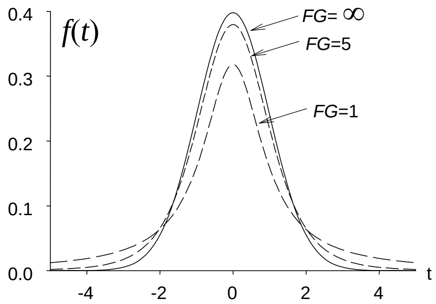
$t = f(x_1, x_2, \ldots, x_n)$

Wenn die Nullhypothese gültig ist, alle x_i Werte folgen einer Verteilung mit $\mu = 0$.

Wenn diese Verteilung eine Normalverteilung ist, kann die Verteilung von t berechnet werden: t-Verteilung

Wenn die Nullhypothese gültig ist, der aus unserer Stichprobe ausgerechnete t-Wert folgt einer t-Verteilung (Student-Vert.).

Bedingung: x muss normalverteilt sein.

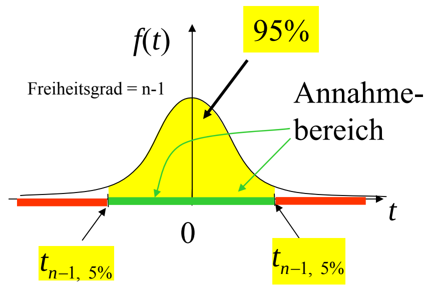


Die Anwendung der t-Verteilung

Die Wahrscheinlichkeit, dass bei Richtigkeit der Nullhypothese

$-t_{n-1,\ 5\%} < t < +t_{n-1,\ 5\%}$

gilt, beträgt 95%.



Bei richtiger Nullhypothese ist der aus der Stichprobe ausgerechnete t-Wert mit 95% Wahrscheinlichkeit in dem Annahmebereich. Wir können diesen kleinen t-Wert mit zufälligen Abweichungen erklären. $\Rightarrow$ Wir müssen keine Wirkung voraussetzen.

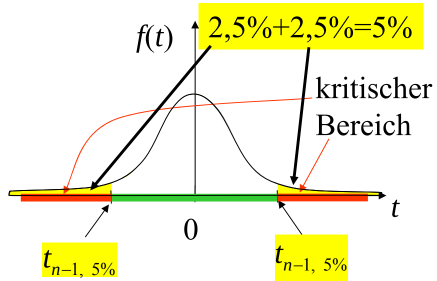
Die Nullhypothese kann nicht abgelehnt werden.

Die Anwendung der t-Verteilung

Die Wahrscheinlichkeit, dass bei Richtigkeit der Nullhypothese

$t < -t_{n-1,\ 5\%}$ oder $t > +t_{n-1,\ 5\%}$

gilt, beträgt 5%.



D.h.: Es ist sehr unwahrscheinlich (<5%), dass wir ***bei richtiger Nullhypothese*** einen so großen t-Wert bekommen. $\Rightarrow$ Wir haben wahrscheinlich eine Wirkung, ***die Nullhypothese kann abgelehnt werden, die Alternativhypothese ist wahrscheinlich richtig.***

Das 5% nennt man als ***Signifikanzniveau*** oder ***Irrtumswahrscheinlichkeit.***

Ablauf der Hypothesenprüfung bei einem t-Test

1. Fragestellung (mit der Definition der Population!)
2. Nullhypothese - Alternativhypothese
3. Festlegung des Signifikanzniveaus (p)
4. Messung (Stichprobe mit n Messungen, Repräsentativität!)
5. Berechnung des t-Wertes
6. Vergleich von unserem t und dem Grenzwert ($t_{n-1,p}$)

$|t| < t_{n-1,p}$ die Nullhypothese kann nicht abgelehnt werden.

Anhand unserer Messung kann die Alternativhypothese nicht bewiesen werden.

$|t| > t_{n-1,p}$ die Nullhypothese kann mit einem p Signifikanzniveau abgelehnt werden

Die Alternativhypothese ist wahrscheinlich richtig.

! Bedingung: x ist normalverteilt! !!!

Beispiel des Fiebermittels

$\bar{x} = -0{,}15 \quad s_x = 0{,}317 \qquad n = 4$

$$t = \frac{-0{,}15}{0{,}317} = -0.47$$

$$-3{,}18 = -t_{3,\ 5\%} < t < +t_{3,\ 5\%} = 3{,}18$$

$\Rightarrow$ ***t*** liegt in dem Annahmebereich, die Nullhypothese kann nicht abgelehnt werden, $\Rightarrow$

$\Rightarrow$ **Das Medikament ist unwirksam**

(mit p=5% Irrtumswahrscheinlichkeit)

FG \p	0.05	0.02	0.01
2	4.30266	6.96455	9.92499
3	3.18245	4.54071	5.84085
4	2.77645	3.74694	4.60408
5	2.57058	3.36493	4.03212
6	2.44691	3.14267	3.70743
7	2.36462	2.99795	3.49948
8	2.30601	2.89647	3.35538
9	2.26216	2.82143	3.24984
10	2.22814	2.76377	3.16926
11	2.20099	2.71808	3.10582
12	2.17881	2.68099	3.05454
13	2.16037	2.65030	3.01228
14	2.14479	2.62449	2.97685
15	2.13145	2.60248	2.94673
20	2.08596	2.52798	2.84534
50	2.00856	2.40327	2.67779
70	1.99444	2.38080	2.64790
100	1.98397	2.36421	2.62589
unendlich	1.95996	2.32635	2.57583

Beispiel der Kniebeugungen

$\bar{x} = 13 \quad s_x = 1{,}77 \qquad n = 6$

$$t = \frac{13}{1{,}77} = 7.34$$

$$t > +t_{5,\ 5\%} = 2{,}57$$

$\Rightarrow$ ***t*** liegt in dem kritischen Bereich, die Nullhypothese kann abgelehnt werden, $\Rightarrow$

$\Rightarrow$ **Die „Behandlung“ ist wirksam**

(mit p=5% Irrtumswahrscheinlichkeit)

Auch bei 2 % und bei 1% Irrtumswahrscheinlichkeit!

FG \p	0.05	0.02	0.01
2	4.30266	6.96455	9.92499
3	3.18245	4.54071	5.84085
4	2.77645	3.74694	4.60408
5	2.57058	3.36493	4.03212
6	2.44691	3.14267	3.70743
7	2.36462	2.99795	3.49948
8	2.30601	2.89647	3.35538
9	2.26216	2.82143	3.24984
10	2.22814	2.76377	3.16926
11	2.20099	2.71808	3.10582
12	2.17881	2.68099	3.05454
13	2.16037	2.65030	3.01228
14	2.14479	2.62449	2.97685
15	2.13145	2.60248	2.94673
20	2.08596	2.52798	2.84534
50	2.00856	2.40327	2.67779
70	1.99444	2.38080	2.64790
100	1.98397	2.36421	2.62589
unendlich	1.95996	2.32635	2.57583

Entscheidung mit dem p-Wert

Computern berechnen den *p*-Wert. z.B. Funtion **ttest** in Excel.

1. Fragestellung (mit Definition der Population!)
2. Nullhypothese - Alternativhypothese
3. Festlegung des Signifikanzniveaus (p_0)
4. Messung (Stichprobe mit *n* Messungen, Repräsentativität!)
5. Berechnung des *p*-Wertes
6. Vergleich von unserem *p* und dem festgelegten Signifikanzniveau (p_0)

$p > p_0$ die Nullhypothese kann nicht abgelehnt werden.

Anhand unserer Messung kann die Alternativhypothese nicht bewiesen werden.

$p < p_0$ die Nullhypothese kann mit einem p_0 Signifikanzniveau abgelehnt werden

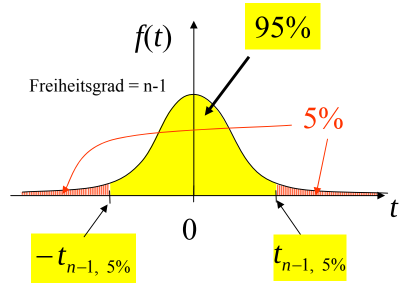
Die Alternativhypothese ist wahrscheinlich richtig.

Die Bedeutung des Signifikanzniveaus

Bei einem unwirksamen Medikament beträgt die Wahrscheinlichkeit dafür, dass $|t| > t_{n-1,p}$ ist, 5%.
(⇨Bei der Untersuchung von hundert unwirksamen Pillen werden zufällig fünf als wirksam gefunden!)

⇩

Fehler erster Art



Fehler von 1. und 2. Art

Fehler erster Art:

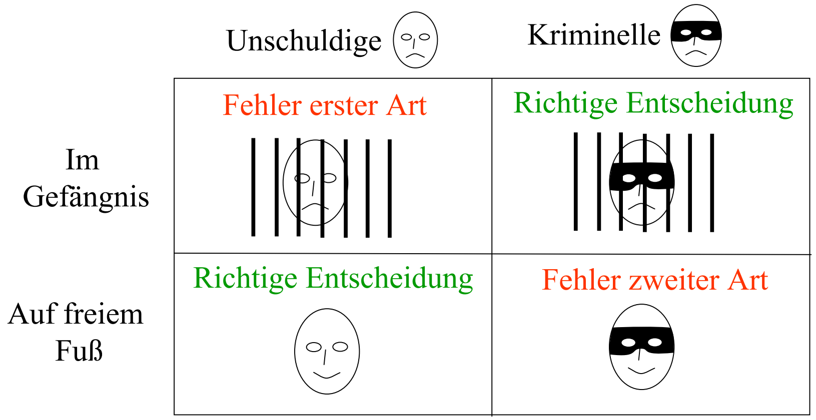
Die Nullhypothese wird zufällig abgelehnt werden, obwohl sie richtig ist!
Die Wahrscheinlichkeit des Fehlers erster Art = Signifikanzniveau
zB: Unwirksame Pille als wirksam gefunden

Fehler zweiter Art:

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*

Fehler von 1. und 2. Art

Nullhypothese: unschuldig

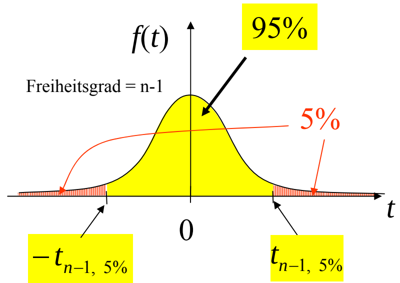


Einseitige/zweiseitige Teste

Ist es interessant wenn das Medikament die Körpertemperatur erhöht?

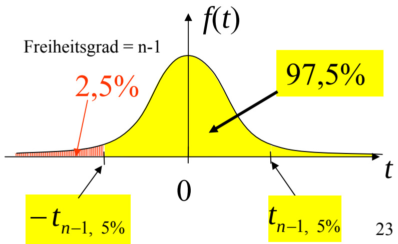
Zweiseitiger Test:

Nullhyp: das Medikament ändert die Körpertemperatur nicht.



Einseitiger Test

Nullhyp: das Medikament erniedrigt die Körpertemperatur nicht.



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

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

$t = -2{,}80$