BA KW | Vorlesung

Einführung in die Statistik

Signifikanztest

Prof. Thomas Hanitzsch

Testtheorie: Grundidee

- Ziel:
  - Anhand wahrscheinlichkeitstheoretischer Überlegungen ziehen wir von einer oder mehreren Stichproben Rückschlüsse auf die zugehörige(n) Grundgesamtheit(en)
- Vorgehensweise:
  - Die Tests zur Prüfung von Hypothesen werden durchgeführt, indem eine aus den Stichprobenwerten gewonnene Größe (Prüfgröße) mit einem kritischen Wert verglichen wird, der die bei einer gegebenen Wahrscheinlichkeit maximal zulässige Abweichung vom erwarteten Wert repräsentiert
  - Wie beim Konfidenzintervall beziehen sich die empirisch gewonnene Prüfgröße und der kritische Wert auf standardisierte Verteilungen mit bekannten Wahrscheinlichkeiten (Standardnormalverteilung, t-Verteilung etc.)

Hypothesen

- Unterschieds- und Zusammenhangshypothesen:
  - Annahme über systematische Unterschiede vs. systematische Zusammenhänge zwischen mindestens zwei Variablen
- Gerichtete (einseitige) und ungerichtete (zweiseitige) Hypothesen:
  - ungerichtete Hypothesen konstatieren einen Zusammenhang
  - gerichtete Hypothesen treffen überdies Aussagen über die Richtung des vermuteten Zusammenhangs
- Spezifische und unspezifische Hypothesen:
  - spezifische Hypothesen treffen Aussagen über die Größe des vermuteten Zusammenhangs

Statistische Hypothesen und Hypothesentest

- Die Nullhypothese (H_0) ist eine Behauptung, von der zunächst angenommen wird, dass sie stimmt
- Die Alternativhypothese (H_1) ist die formale Gegenhypothese und markiert üblicherweise die zu prüfende bzw. zu beweisende Vermutung des Forschers
  - H_1 und H_0 sind somit zueinander komplementäre Aussagen
- Ziel des Hypothesentests ist es, die Nullhypothese zugunsten der Alternativhypothese aufzugeben
  - H_0 wird beibehalten, solange die Testergebnisse nicht deutlich gegen sie sprechen
  - Hierfür wird eine Fehlerobergrenze festgelegt (siehe Signifikanzniveau)

Statistische Hypothesen und Hypothesentest

- Beispiel A
  - Der letzte Zensus der Minion-Population im Ewigen Eis von 1968 hat ergeben, dass mindestens 30 Prozent der Minions einäugig sind. → Nullhypothese
  - Minionforscher Professor Flux vermutet, dass der Anteil der einäugigen Minions seither zurück gegangen ist und mithin kleiner als 30 Prozent ist. → Alternativhypothese
- Beispiel B
  - Minionforscher Professor Flux vermutet, dass der Anteil der einäugigen Minions kleiner als 30 Prozent ist. → Alternativhypothese
  - Der Anteil der einäugigen Minions liegt bei mindestens 30 Prozent. → Nullhypothese

Hypothesentests: Vorgehen

1. Formulierung der Nullhypothese (H_0) und Alternativhypothese (H_1)
2. Entscheidung über Testsituation (linksseitig, rechtsseitig oder zweiseitig)
3. Festlegung des Signifikanzniveaus
4. Konstruieren der Entscheidungsregel (d.h. Bestimmung der kritischen Werte und des Annahmebereichs von H_0)
5. Ziehung der Stichproben und Berechnung der Prüfgröße
6. Anwendung der Entscheidungsregel (d.h. Vergleich von Prüfgröße mit kritischem Wert)
7. Testentscheidung (Annahme bzw. Ablehnung der Nullhypothese)

Testsituation: ein- und zweiseitige Tests

- Einseitiger Test:
  - Gerichtete Alternativhypothese H_1:
    Es wird ein Zusammenhang vermutet, der in eine bestimmte Richtung geht
    - Linksseitiger Test: $H_0: \mu \geq \mu_0$ $H_1: \mu < \mu_0$
    - Rechtsseitiger Test: $H_0: \mu \leq \mu_0$ $H_1: \mu > \mu_0$
- Zweiseitiger Test:
  - Ungerichtete Alternativhypothese H_1:
    Es wird ein Zusammenhang vermutet, wobei dessen Richtung unbedeutend ist
    - Zweiseitiger Test: $H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$

α-Fehler und β-Fehler

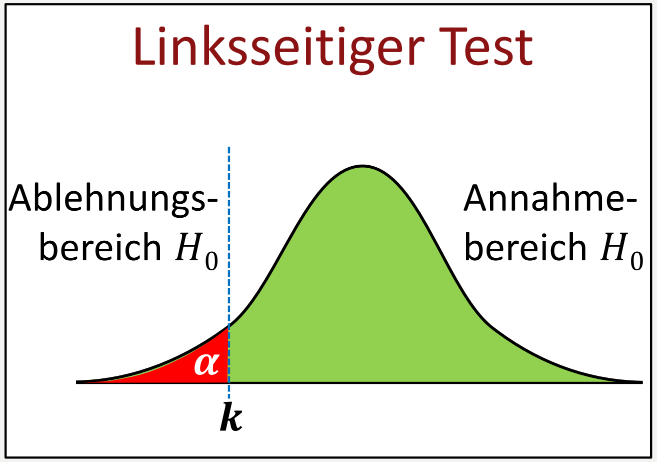
		Für die *Wirklichkeit* gilt:	
		H_0 ist wahr	H_1 ist wahr
Test-Entscheidung (basierend auf der Stichprobe)	H_0 beibehalten	**Ok**	**β-Fehler (Fehler 2. Ordnung)**
	H_0 (zugunsten von H_1) ablehnen	**α-Fehler (Fehler 1. Ordnung)**	**Ok**

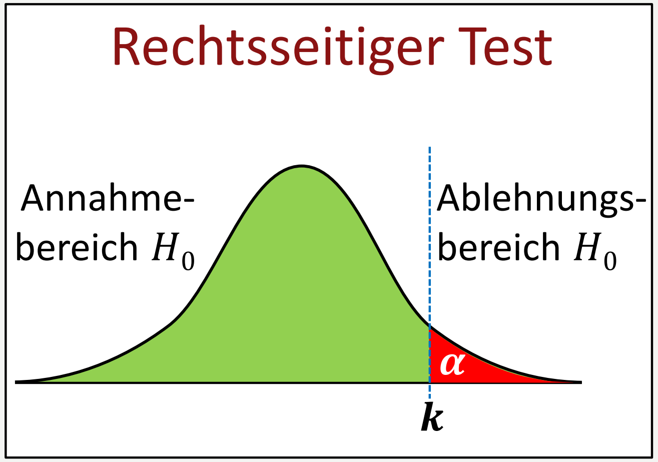
Signifikanzniveau und Signifikanztest

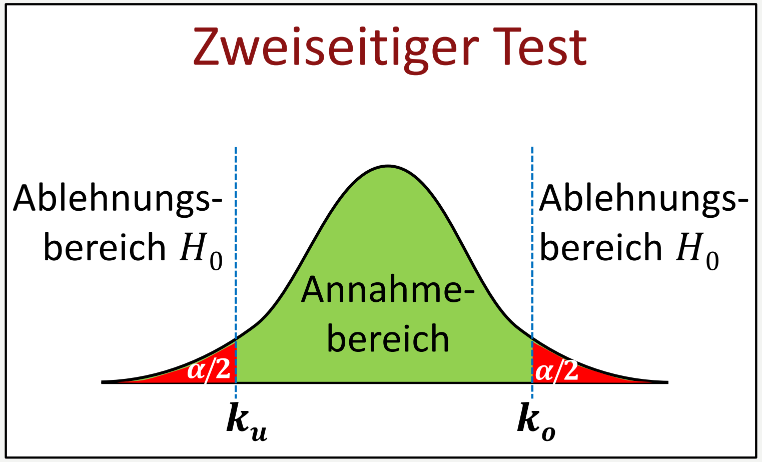
- Das Signifikanzniveau α:
  - Steht für die Wahrscheinlichkeit der Menge von Testergebnissen, für die wir bereit sind, H_0 zurückzuweisen, obwohl sie zutreffend ist
  - α bezeichnet daher die vom Forscher festgelegte Wahrscheinlichkeit, mit welcher die Ablehnung von H_0 im Rahmen eines Signifikanztests zu einem Fehler 1. Art führt
  - Meist werden das 5%- und das 1%-Signifikanzniveau verwendet ($\alpha = 0{,}05$ bzw. $\alpha = 0{,}01$)
- Das Prinzip des Signifikanztests:
  - Es wird berechnet, in welchem Bereich bei einer festgelegten Wahrscheinlichkeit $(1 - \alpha)$ die Testergebnisse aus mehreren gleichartigen Studien (die jeweils einem zufälligen Auswahlfehler unterliegen) liegen müssen, wenn H_0 zutrifft
  - Im Signifikanztest wird geprüft, ob das empirische Testergebnis innerhalb oder außerhalb dieses Bereiches liegt

Annahme- und Ablehnungsbereich

Annahmebereich	Ablehnungsbereich
Menge der Testergebnisse, die wir (noch) als mit der Nullhypothese H_0 vereinbar sehen	Menge der Testergebnisse, die wir als mit der Nullhypothese H_0 unvereinbar sehen







Der Signifikanztest – linksseitig: Beispiel

- Die ARD behauptet, dass mindestens 50 Prozent der Deutschen den „Tatort" mögen. Medienforscher Franz bezweifelt das und sagt, es sind weniger. ($\alpha = 0{,}05$; $N = 100$)

$$\boldsymbol{H_0{:}p \geq 0,5} \qquad \boldsymbol{H_1{:}p < 0,5}$$

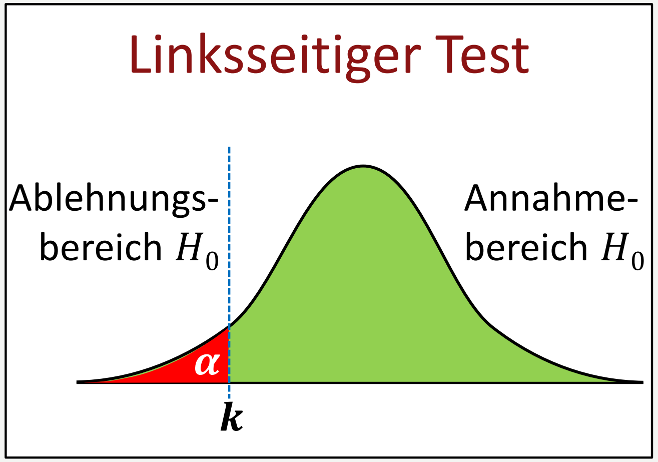


Tabelle der Binomialverteilung:
„kumulierte Wahrscheinlichkeiten"

$\boldsymbol{k}$	$\boldsymbol{P(X \leq k) \leq \alpha}$
41	0,044

- Annahmebereich: [42 ; 100]
- Ablehnungsbereich: [0 ; 41]

Der z-Test

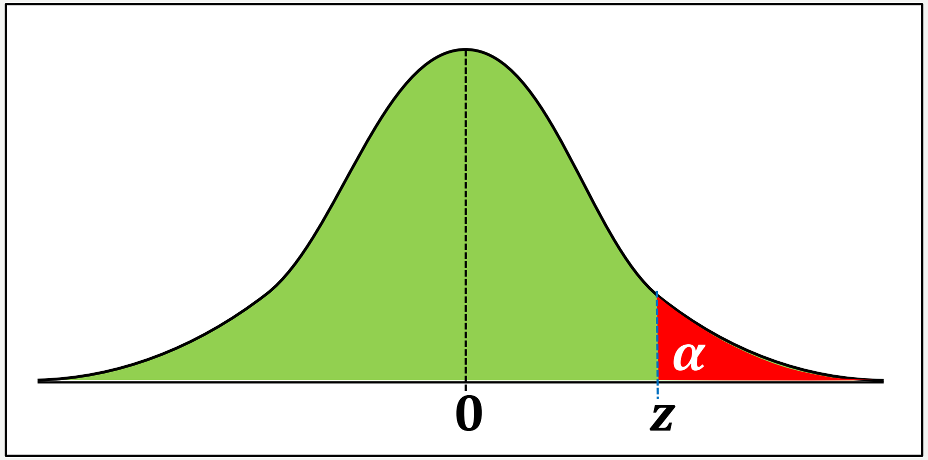
- Approximation durch Normalverteilung bei großen Stichproben
  → es gilt: $N \cdot p_0 \cdot (1 - p_0) > 9$

- Bestimmung der kritischen Werte:
  - Linksseitig: $k = N \cdot p_0 - z_{1-\alpha} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$
  - Rechtsseitig: $k = N \cdot p_0 + z_{1-\alpha} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$
  - Zweiseitig: $k_u = N \cdot p_0 - z_{1-\frac{\alpha}{2}} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$
    $k_o = N \cdot p_0 + z_{1-\frac{\alpha}{2}} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$

Der z-Test

- Häufig verwendete α-Quantile der Normalverteilung:

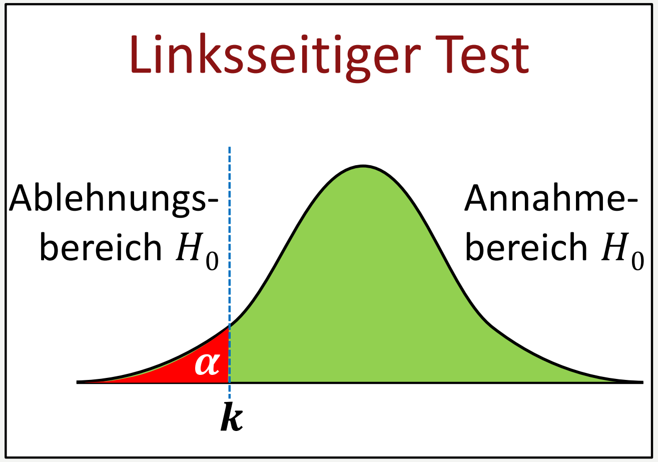
$P(Z \leq z)$	z
95,0%	1,64
97,5%	1,96
99,0%	2,33
99,5%	2,58



Der z-Test – linksseitig: Beispiel

- Die ARD behauptet, dass mindestens 50 Prozent der Deutschen den „Tatort“ mögen. Medienforscher Franz bezweifelt das und sagt, es sind weniger. ($\alpha = 0{,}05$; $N = 100$)

$$\boldsymbol{H_0: p \geq 0,5} \qquad \boldsymbol{H_1: p < 0,5}$$



$$k = N \cdot p_0 - z_{1-\alpha} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$$
$$k = 100 \cdot 0{,}5 - 1{,}64 \cdot \sqrt{100 \cdot 0{,}5 \cdot (1 - 0{,}5)}$$
$$k = 50 - 8{,}2 = \mathbf{41,8}$$

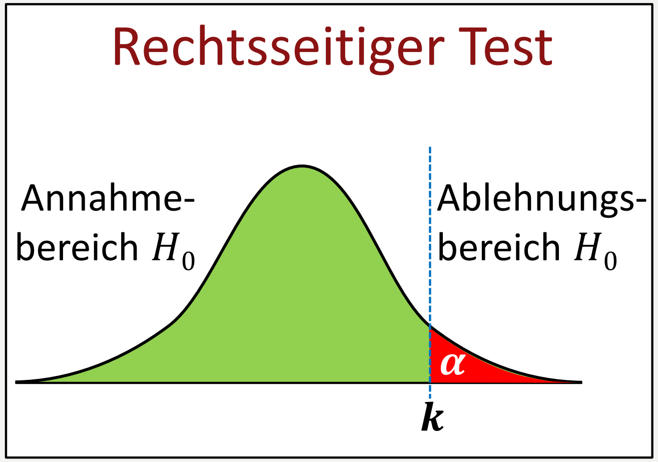
- Annahmebereich: [42 ; 100]
- Ablehnungsbereich: [0 ; 41]

Der z-Test – rechtsseitig: Beispiel

- Der Löwen-Sportvorstand behauptet, nicht mehr als 70 Prozent der Münchner sind Bayern-Fans. Kollege Rummenigge sagt, es sind mehr. ($\alpha = 0{,}05$; $N = 50$)

$\boldsymbol{H_0: p \leq 0,7}$

$\boldsymbol{H_1: p > 0,7}$



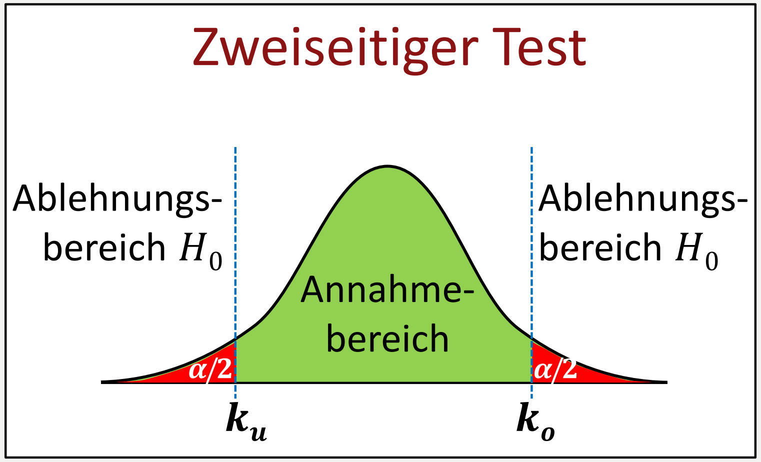
$$k = N \cdot p_0 + z_{1-\alpha} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$$
$$k = 50 \cdot 0{,}7 + 1{,}64 \cdot \sqrt{50 \cdot 0{,}7 \cdot (1 - 0{,}7)}$$
$$k = 35 + 5{,}31 = \mathbf{40,3}$$

- Annahmebereich: [0 ; 40]
- Ablehnungsbereich: [41 ; 50]

Der z-Test – zweiseitig: Beispiel

- Anna-Maria spielt im Casino und vermutet, dass sie es mit einem gezinkten Würfel zu tun hat. ($\alpha = 0{,}05$; $N = 100$)

$$\boldsymbol{H_0{:}p} = {}^{1}/_{6} \qquad \boldsymbol{H_1{:}p} \neq {}^{1}/_{6}$$



$$k_u = N \cdot p_0 - z_{1-\frac{\alpha}{2}} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$$

$$k_u = 100 \cdot {}^{1}/_{6} - 1{,}96 \cdot \sqrt{100 \cdot {}^{1}/_{6} \cdot (1 - {}^{1}/_{6})} = \mathbf{9,4}$$

$$k_o = N \cdot p_0 + z_{1-\frac{\alpha}{2}} \cdot \sqrt{N \cdot p_0 \cdot (1 - p_0)}$$

$$k_o = 100 \cdot {}^{1}/_{6} + 1{,}96 \cdot \sqrt{100 \cdot {}^{1}/_{6} \cdot (1 - {}^{1}/_{6})} = \mathbf{24,0}$$

- Annahmebereich: [10 ; 24]
- Ablehnungsbereich: [0 ; 9] + [25 ; 100]

Einfacher t-Test für den Mittelwert

- Prüft, ob der empirische Mittelwert $\overline{x}$ aus einer Grundgesamtheit mit dem Erwartungswert μ entstammt

$$\boldsymbol{H_0}: \boldsymbol{\overline{x} = \mu} \qquad \boldsymbol{H_1}: \boldsymbol{\overline{x} \neq \mu}$$

$$t = \sqrt{N} \cdot \left(\frac{\overline{x} - \mu}{s} \right)$$

- T ist unter H_0 t-verteilt mit $df = N - 1$ Freiheitsgraden
- Zweiseitiger Test:
  - Kritischer Wert t_{krit} entspricht dem $\left(1 - \frac{\alpha}{2}\right)$-Quantil der t-Verteilung
  - H_0 wird abgelehnt, wenn $|t| > t_{krit}$

Einfacher t-Test für den Mittelwert: Beispiel

- Eine repräsentative Studie in China konnte zeigen, dass Zehntklässler 2016 durchschnittlich fünf Mal mit anderen Passanten auf dem Gehweg zusammengestoßen waren, da sie zu sehr in ihr Handy vertieft waren. ($\mu = 5$)
- Eine aktuelle Befragung einer Zufallsauswahl von 50 Schülern an einer Schule im chinesischen Chongqing weist nun einen Wert von durchschnittlich zehn Kollisionen aus. ($\bar{x} = 10$; $s = 2{,}5$; $N = 50$)
- Lässt sich anhand dieses Ergebnisses – bei einer Irrtumswahrscheinlichkeit von 5% – sagen, dass sich die Schüler an der untersuchten Schule systematisch vom Durchschnitt der chinesischen Schüler unterscheiden? ($\alpha = 0{,}05$)

$\mu = 5$
$\overline{x} = 10$
$s = 2{,}5$
$\alpha = 0{,}05$
$N = 50$

Einfacher t-Test für den Mittelwert: Beispiel

- Hypothesen: $\boldsymbol{H_0: \overline{x} = \mu}$ $\boldsymbol{H_1: \overline{x} \neq \mu}$
  - $t = \sqrt{N} \cdot \left(\frac{\overline{x}-\mu}{s}\right) = \sqrt{50} \cdot \left(\frac{10-5}{2{,}5}\right) = 7{,}07 \cdot 2 = \mathbf{14,14}$
  - $df = N - 1 = 49$ → $t_{krit} = \mathbf{2,02}$ (zweiseitig)
  - $t = 14{,}14 > t_{krit}$ → H_0 wird abgelehnt
- Die mittlere Anzahl der Kollisionen in der untersuchten Schule weicht signifikant vom chinesischen Durchschnitt ab ($t = 14{,}14$; $df = 49$; $p < 0{,}05$).

Der p-Wert

- Bemisst die Wahrscheinlichkeit, dass die Teststatistik den beobachteten oder einen noch extremeren Wert (d.h. noch weiter entfernt von H_0 liegend) annimmt, obwohl H_0 richtig ist
- Wenn $p < \alpha$, dann wird H_0 abgelehnt und die Evidenz spricht zunächst (d.h. bis gegenteilige Evidenz vorliegt) für die Annahme von H_1
- Konvention:
  - $p < 0{,}05$ (5%): Ergebnis ist *signifikant*
  - $p < 0{,}01$ (1%): Ergebnis ist *hoch signifikant*

Statistische Hypothesen: Aufgabe 1

- Entwickeln Sie jeweils eine Null- und Alternativhypothese für folgende Fragestellungen. Geben Sie jeweils eine gerichtete und eine ungerichtete Hypothese an.
  a) Bei der Bundestagswahl 2017 gab es deutschlandweit eine Wahlbeteiligung von 72%. Sie vermuten, dass sich die Wahlbeteiligung unter den Zuschauern von „Maybritt Illner" hiervon unterscheidet.
  b) Die Macher*innen der Sendung „ZDF Magazin Royale" geben an, dass seit Verleihung des Grimme-Preis 2023 jede dritte Person in Deutschland ihre Sendung zumindest dem Namen nach kennt. Sie vermuten, dass dies nicht korrekt ist.

Der z-Test: Aufgabe 2

- Überprüfen Sie die folgenden Annahmen mittels eines Signifikanztests.
  a) Bei der Bundestagswahl 2017 gab es deutschlandweit eine Wahlbeteiligung von 72%. Sie stellen die Hypothese auf, dass die Wahlbeteiligung unter „Maybrit Illner"-Zuschauern höher ist als in der Gesamtbevölkerung. In einer Befragung 90 Zuschauern gaben 75 Personen an, gewählt zu haben. Das Signifikanzniveau legen Sie auf 1% fest.
  b) Die Macher*innen der Sendung „ZDF Magazin Royale" geben an, dass seit Verleihung des Grimme-Preis 2023 jede dritte Person in Deutschland ihre Sendung zumindest vom Namen her kennt. Sie stellen die Hypothese auf, dass die Bekanntheit der Sendung nicht 33% entspricht. Bei einer Quoten-Erhebung von 200 Personen aus der deutschen Wohnbevölkerung gaben 45 Personen an, dass sie die Sendung zumindest dem Namen nach kennen. Das Signifikanzniveau legen Sie auf 5% fest.

Exkurs: Freiheitsgrade

- Die mit einem Kennwert verbundenen Freiheitsgrade entsprechen der Anzahl der Werte, die bei seiner Berechnung frei variieren können
  - $\overline{x}$ hat N Freiheitsgrade, weil es keinerlei Bedingung gibt, der die N Werte genügen müssen
  - s^2 hat $N-1$ Freiheitsgrade, die in die Berechnung der Quadratsumme ($\sum_{i=1}^{N}(x_i-\bar{x})^2$) eingehen, da die Summe der Abweichungen vom Mittelwert gleich null ist (von N Abweichungen können deshalb nur $N-1$ frei variieren)

➢ Für die Berechnung des t-Wertes werden die Freiheitsgrade vom Variabilitätsmaß (s) übernommen, daher gilt $N-1$