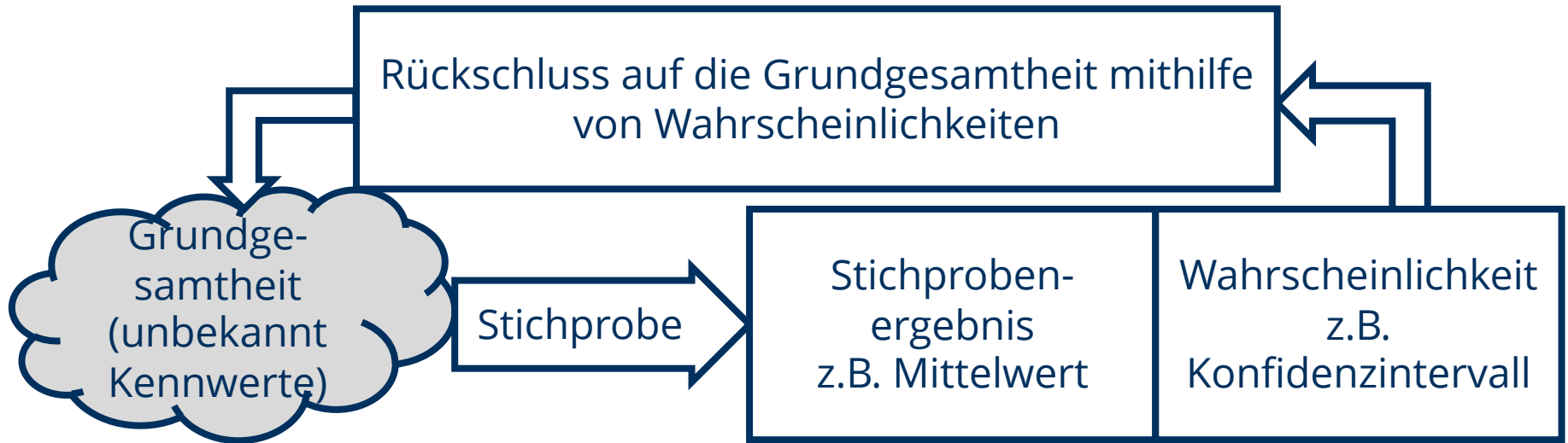


Hypothesen und Auswahl für statistische Tests

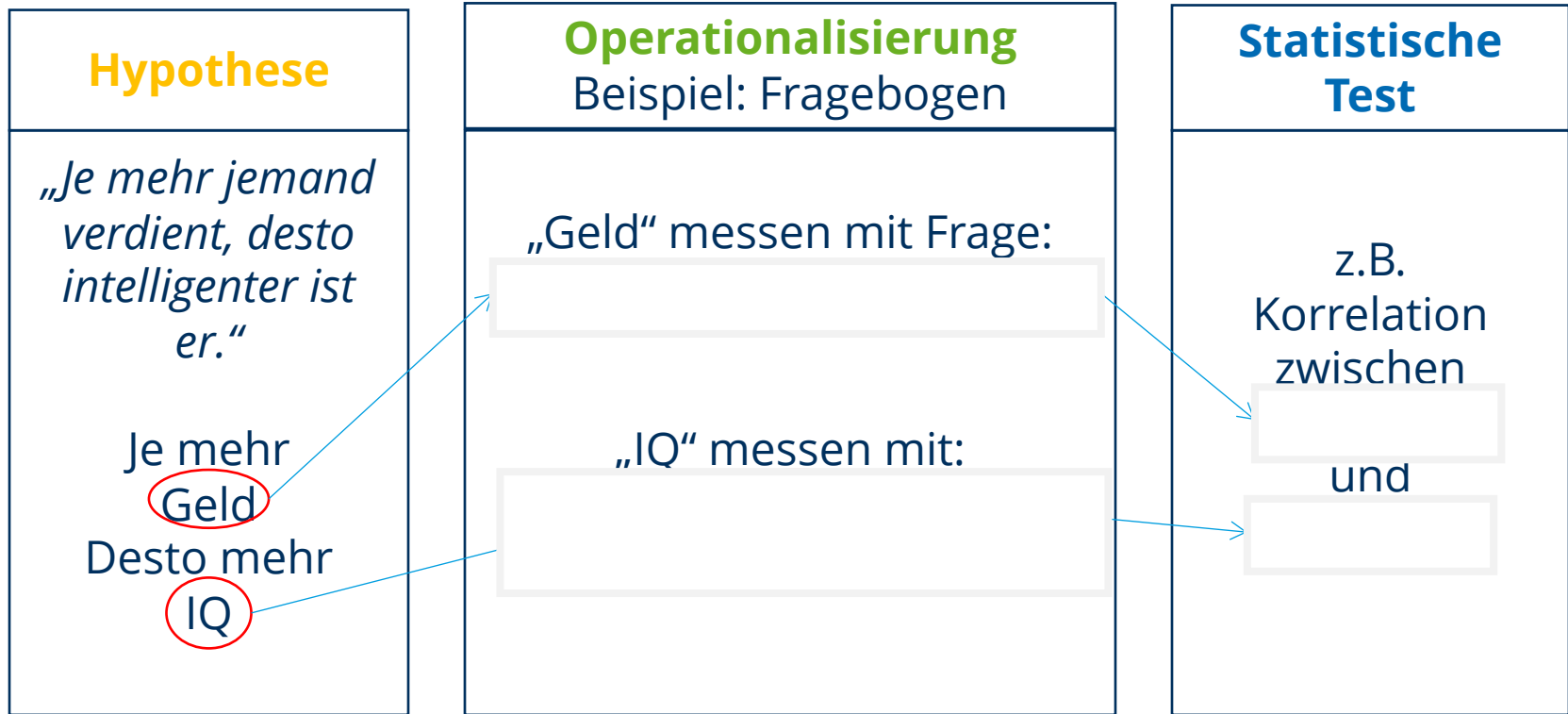
Schließende Statistik arbeitet mit statistischen Tests



Hypothesenarten

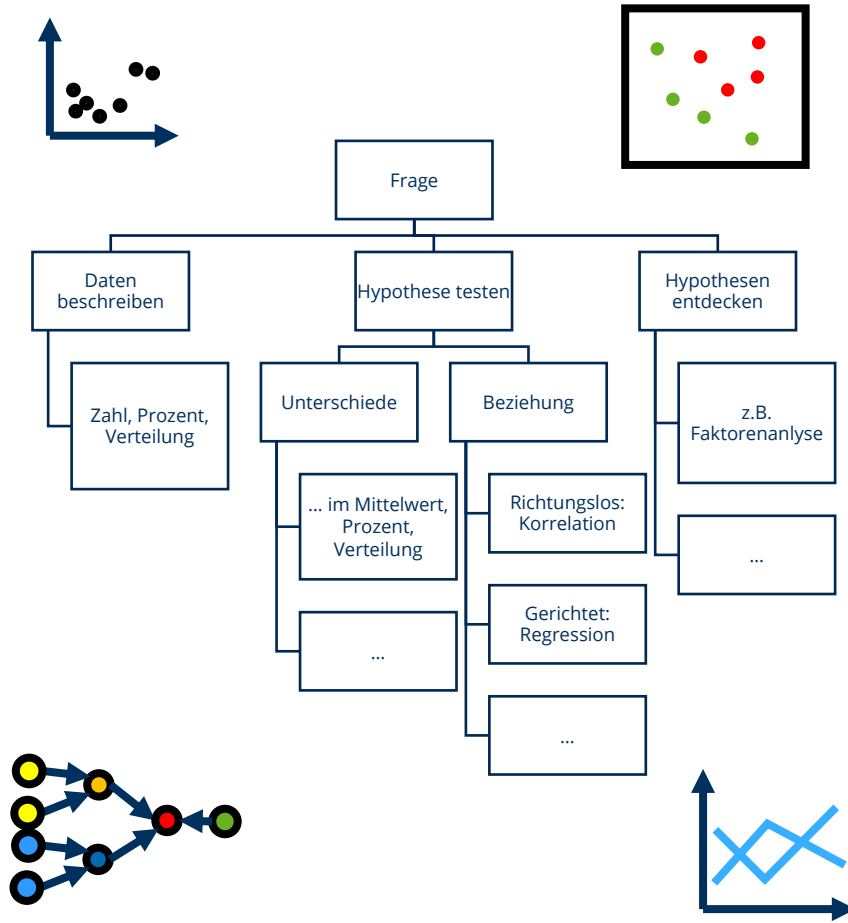
| Hypothese | Erklärung | Beispiel |
|--------------------------|---|----------|
| Es-gibt-Hypothesen | Bestätigt wenn ein Fall gefunden wurde. | |
| Wenn-A-dann-B-Hypothesen | Strenge Kausalität | |
| Unterschiedshypothesen | Unterschiede in zwei (oder mehr) Gruppen | |
| Zusammenhangshypothesen | Stetiger Zusammenhang | |
| Fälle gruppieren | Ähnliche Fälle finden | |
| Variablen gruppieren | Ähnliche Variablen finden | |

Von der **Hypothesen** über die **Operationalisierung** zum statistischen Test



Es gibt so viele statistische Analyseverfahren! Ahhhh..

Wer kann einen statistisches Analyseverfahren auswählen?



- Wer ...
- ... einen Überblick über die grundlegenden statistischen Verfahren hat.
- ... Kenntnisse über die Voraussetzungen und Anforderungen der statistischen Tests hat.
- ... Kreativität hat, um aus den vorhanden Variablen neue zu bilden um die Daten für einen statistischen Test vorzubereiten.
- ... den infrage kommenden Datensatz versteht und etwas kennt.

Was die Wahl eines statistischen Analyseverfahrens beeinflusst:

- die **Fragestellung und Hypothesen**
- das **Skalenniveau** (nominal-, ordinal-, metrischskaliert)
- die **Anzahl** der zu analysierenden **Variablen** (uni-, bi- oder multivariate Analysen)
- die **Stichprobe** (Größe; Art (unabhängig/abhängig))
- Vorhandene Variablen/Daten/Fälle
- Eigenschaften der Verteilung der Werte einer Variablen
- die zu untersuchenden Kennwerten/Verteilungen (z.B. Mittelwert, Häufigkeit, Korrelation, Rangsummen, Anteile, Varianz)
- der Kausalität des Zusammenhangs (abhängig und unabhängige Variablen)
- der Form des Zusammenhangs (linear, nichtlinear)
- der Anzahl der Gruppen: 1, 2 oder viele
- den Eigenschaften der Verteilung der Werte einer Variablen
- Qualität der erhobenen Daten und des Erhebungsinstruments
- Wechselwirkungen zwischen Variablen
- Welche Statistikprogramme stehen zur Verfügung
- Was versteht der Leser, was will er haben
- ... und welche statistischen Verfahren kenne ich überhaupt

Übungen: Von der Frage zum Auswertungsverfahren

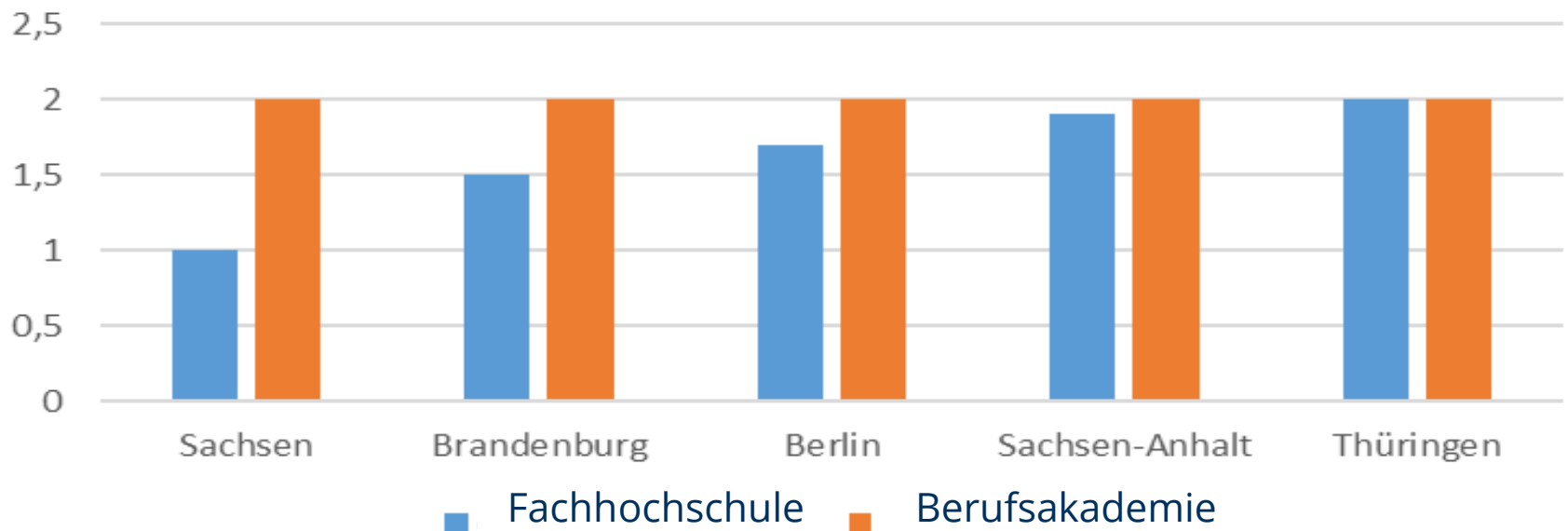
| Frage | Variablen | Auswertungsverfahren |
|--|-----------|----------------------|
| Welche Einstellungen haben Mitarbeiter zur flexiblen Arbeitszeit? | | |
| Empfinden Kunden einen oder zwei Ansprechpartner besser? | | |
| Größere Anbieter haben höhere Preise. | | |
| Meetings die kürzer und besser vorbereitet sind, sind erfolgreicher. | | |

Hypothesen- oder Signifikanztests als Blackboxes

Wo ist der signifikante Unterschied?

Achtung: Die Daten sind
(aus didaktischen Gründen)
frei erfunden!

Einkommensunterschiede zwischen Absolventen in verschiedenen Bundesländern

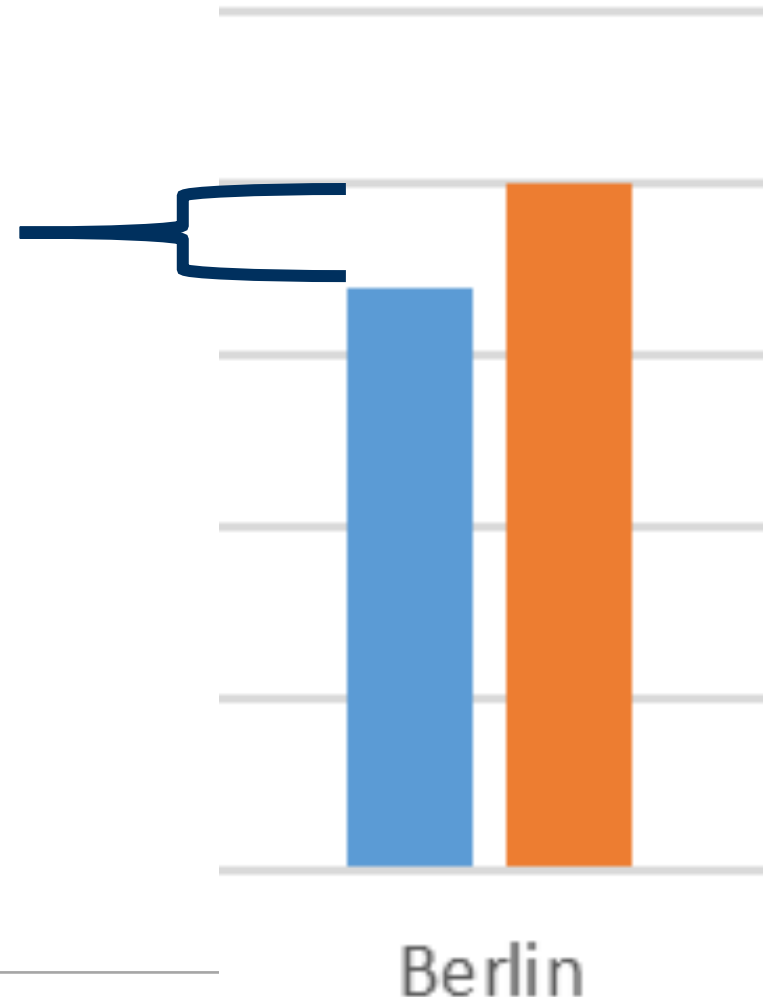


Die Ausgangsfrage bei Signifikanztests

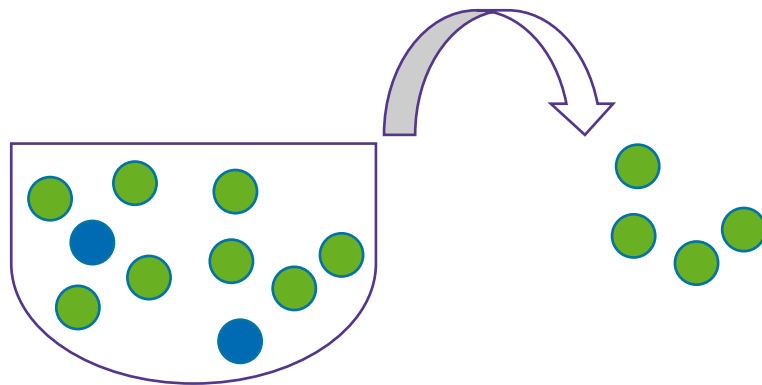
Zur Info! Signifikanz wird hier am Beispiel Mittelwerte verglichen. Weil dies anschaulicher ist als andere Fragestellungen.

Ist der Unterschied wesentlicher Natur, oder ist das Ergebnis **zufällig** zustande gekommen?

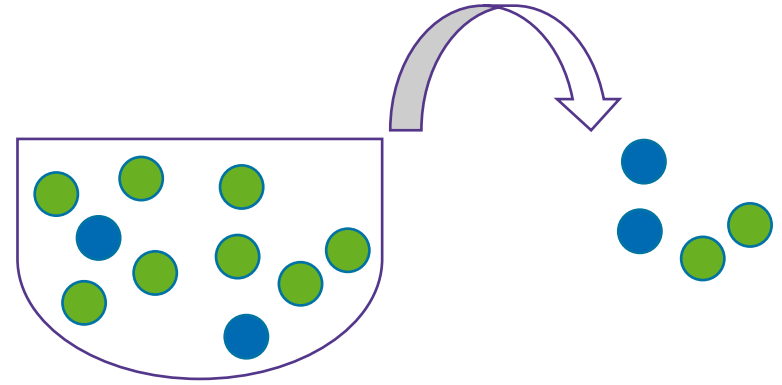
Das ist DIE FRAGE, die statistische Tests beantworten.



Warum gibt es zufällige Ereignisse?



1. Stichprobe



2. Stichprobe

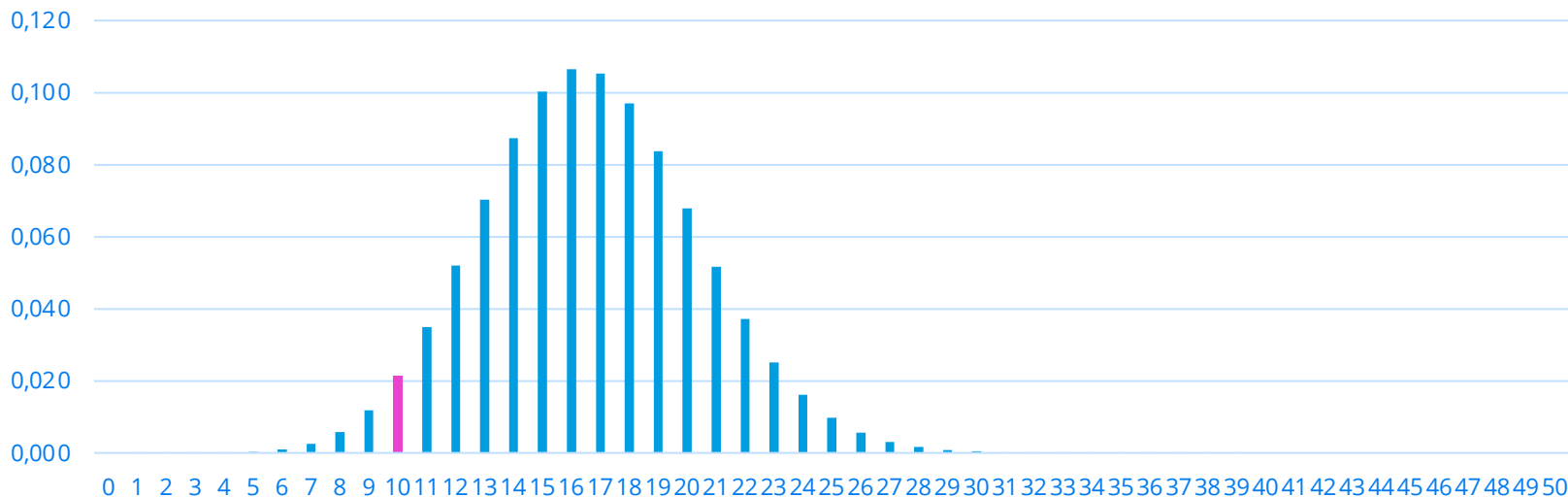
Irrtumswahrscheinlichkeit festlegen. Ein Beispiel für den Einstieg.

Sie möchten wissen, ob ein Würfel fair ist. Sie würfeln 100 Mal und haben 10 mal die 6. Die Wahrscheinlichkeit dafür ist 4,3 %*.

WIR DEFINIEREN, dass Ereignisse mit einer Wahrscheinlichkeit unter 5 % zu gering sind.

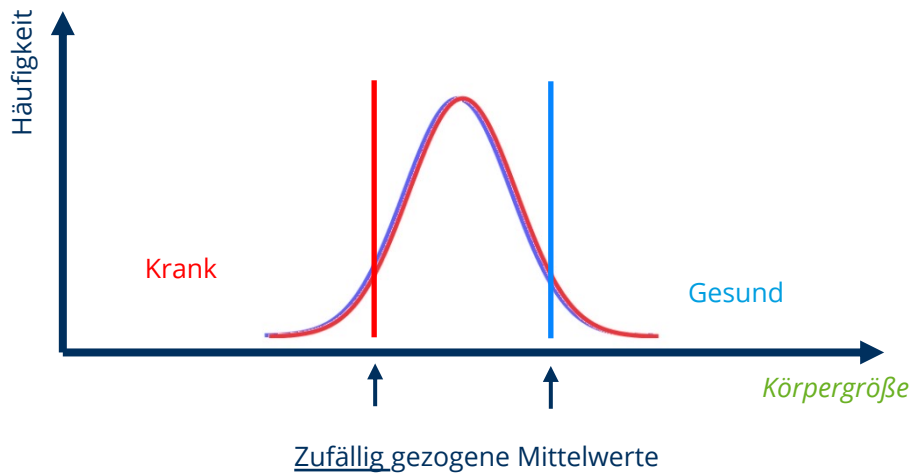
Wir schlussfolgern, dass der Würfel gezinkt ist. (Wissen tuen wir es aber nicht).

Prinzipiell kann so ein Ergebnis zufällig auftreten, es ist nur eher unwahrscheinlich.

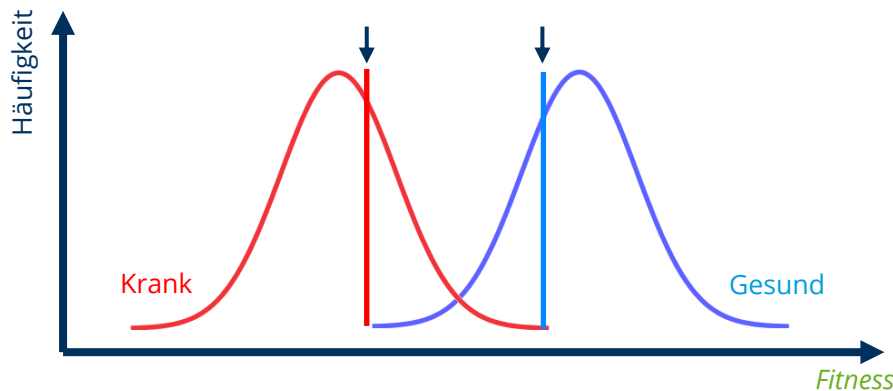


* Zur Berechnung dieser Wahrscheinlichkeit vgl. später Binominalverteilung. Jetzt könnten sie aber schon den Erwartungswert errechnen: $100 \cdot \frac{1}{6} = 16,66$.

Zufällige Ereignisse bei zwei Stichproben



Ergibt sich eine Differenz der Mittelwerte in zwei Gruppen kann nicht sofort geschlossen werden, dass die beiden Gruppen sich „wirklich“ von einander unterscheiden.

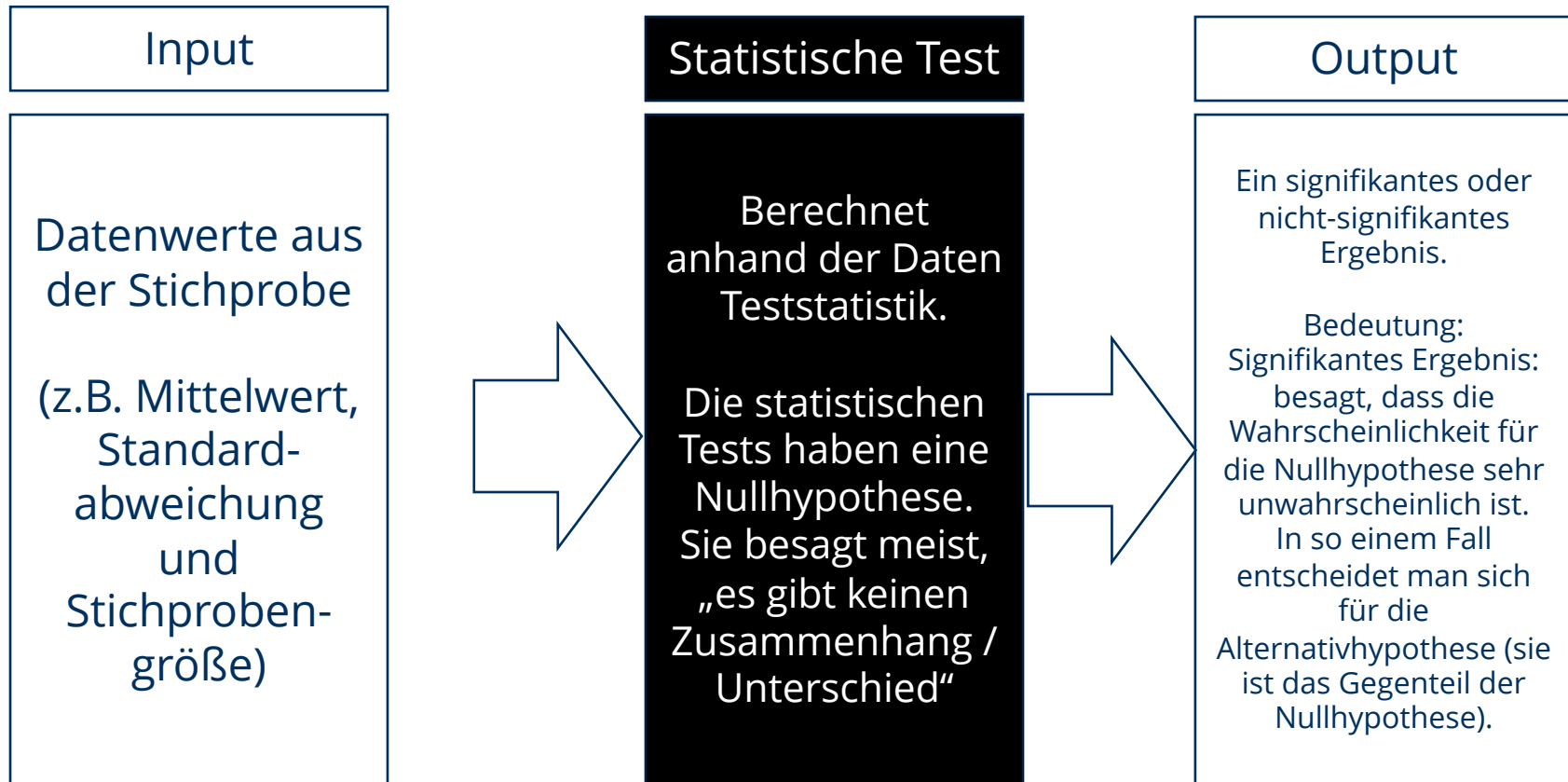


Ob ein Unterschied „signifikant“ ist kann nur ein Signifikanztest beantworten.

Dieser Test ermittelt auf Grund der Kennwerte der Verteilung und der Stichprobengröße, ob die gefundene Differenz signifikant ist.

Was brauchen statistische Tests und was liefern sie?

Zur Info! Statistische Tests werden hier am Beispiel des Mittelwertvergleichs erläutert.



Statistische Tests: Null und Alternativhypothese

Nullhypothese

• Ein statistischer Test prüft, ob kein Unterschied vorhanden ist/kein Zusammenhang da ist.

• H_0 = Nullhypothese:

• Bsp.: Es gibt **keinen** Unterschied in der Ausprägung einer Variable zwischen verschiedenen Gruppen.

• Als Formel: $\mu_0 = \mu_1$

Alternativhypothese

• Die Alternativhypothese ist das Gegenteil der Nullhypothese.

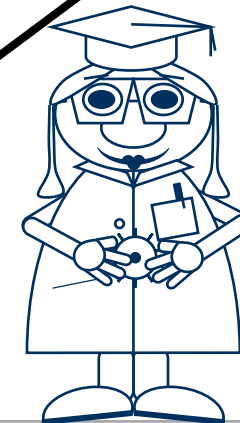
• H_1 = Alternativhypothese:

• Bsp.: Es gibt **einen** Unterschied in der Ausprägung einer Variable zwischen verschiedenen Gruppen.

• Als Formel: $\mu_0 \neq \mu_1$

Achtung!
Die Tests
überprüfen die
Nullhypothese. Die
Alternativhypothese
wird **nicht** von den
Tests überprüft.

Stat.
Test



Hypothesen für statistische Tests im Überblick

Forschungshypothese:

Im Durchschnitt produziert Maschine A längere Nudeln als Maschine B.

Nullhypothese H_0 : Der Mittelwert der Nudeln unterscheidet sich nicht.

$$(\mu_a = \mu_b)$$

Alternativhypothese allgemein H_1 : Die Mittelwerte unterscheiden sich signifikant.

$$(\mu_a \neq \mu_b)$$

Was sagt der Output eines statistischen Tests

Zur Info! Statistische Tests werden hier am Beispiel des Mittelwertvergleichs erläutert.



Allg.

Statistischer Prüfgröße: z.B. t-Wert
Weiter Werte
z.B. Freiheitsgrade

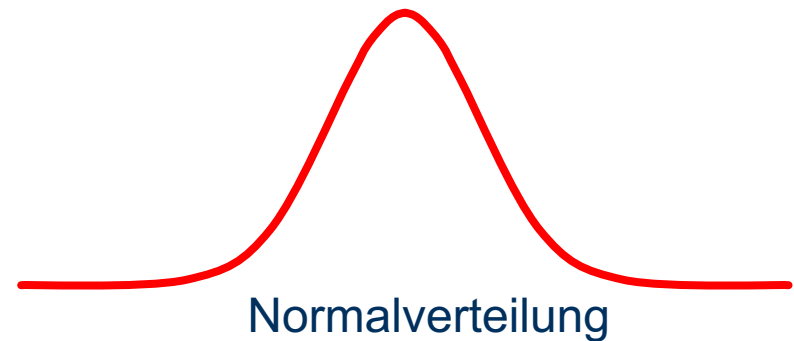
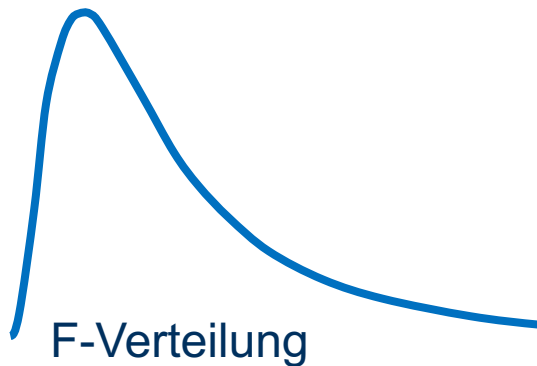
(Ermittlung des zugehörigen Signifikanzniveaus
siehe folgende Folien)

Soft
ware

In statistischer Software wird auch der
statistische Wert angezeigt aber es wird sofort
das zutreffende **Signifikanzniveau** ermittelt.

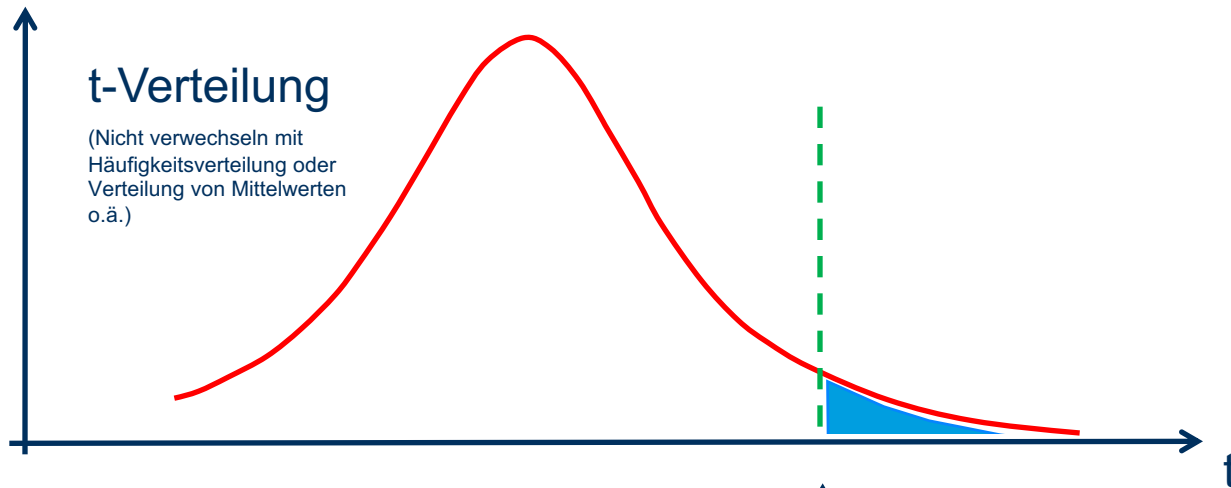
Hintergrund für statistische Tests

Jeder Wert eines statistischen Test kann – einer für den Test bekannten – Verteilungsform zugeordnet werden. Viele Tests werden anhand dieser Verteilung benannt, haben aber auch alternative Namen.



Bedeutung des t-Wertes

Zur Info! Statistische Tests werden hier am Beispiel des Mittelwertvergleichs erläutert.



Die Fläche gibt an:

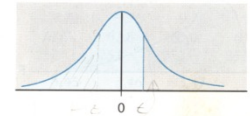
Wie wahrscheinlich ist das Auftreten der gefundenen Ergebnisse, wenn die Nullhypothese in Wirklichkeit zutrifft.

Die Fläche kann als Wahrscheinlichkeit, interpretiert werden.

t-Wert (Beispiel)
Mit Hilfe des statistischen Tests (t-Test) empirisch ermittelte Wert.

Arbeiten mit der Tabelle der Verteilungsfunktion

Tabelle D. Verteilungsfunktion der t-Verteilungen und zweiseitige Signifikanzgrenzen für Produkt-Moment-Korrelationen (zit. nach Glass, G. V., Stanley, J. C.: Statistical methods in education and psychology, p. 521. New Jersey: Prentice-Hall, Englewood Cliffs 1970)



④ Umrechnung:
 $1 - 0,95 = 0,05$

Das ist das relevante Signifikanzniveau.

Beispiel:
 $df = 16$
 $t\text{-Wert} = 2,0$

Freiheitsgrade:
 $df = n_1 + n_2 - 2$

| Fläche* df | 0,55 | 0,60 | 0,65 | 0,70 | 0,75 | 0,80 | 0,85 | 0,90 | 0,95 | 0,975 | 0,990 | 0,995 | 0,9995 | $t_{0,05}$ | $t_{0,01}$ |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|---------|------------|------------|
| 1 | 0,158 | 0,325 | 0,510 | 0,727 | 1,000 | 1,376 | 1,963 | 3,078 | 6,314 | 12,706 | 31,821 | 63,657 | 636,619 | 0,997 | 1,000 |
| 2 | 0,142 | 0,289 | 0,445 | 0,688 | 0,978 | 1,327 | 1,888 | 2,924 | 5,965 | 11,917 | 31,000 | 61,255 | 609,658 | 0,990 | 0,990 |
| 3 | 0,137 | 0,277 | 0,424 | 0,676 | 0,965 | 1,313 | 1,871 | 2,901 | 5,841 | 11,704 | 30,829 | 60,925 | 606,312 | 0,878 | 0,959 |
| 4 | 0,134 | 0,271 | 0,414 | 0,669 | 0,956 | 1,304 | 1,863 | 2,891 | 5,810 | 11,591 | 30,747 | 60,810 | 605,917 | 0,811 | 0,917 |
| 5 | 0,132 | 0,267 | 0,408 | 0,664 | 0,950 | 1,299 | 1,858 | 2,885 | 5,784 | 11,500 | 30,685 | 60,754 | 605,874 | 0,754 | 0,874 |
| 6 | 0,131 | 0,265 | 0,404 | 0,661 | 0,947 | 1,296 | 1,856 | 2,883 | 5,770 | 11,467 | 30,659 | 60,732 | 605,834 | 0,707 | 0,834 |
| 7 | 0,130 | 0,263 | 0,402 | 0,659 | 0,945 | 1,294 | 1,855 | 2,882 | 5,760 | 11,440 | 30,645 | 60,719 | 605,798 | 0,666 | 0,798 |
| 8 | 0,130 | 0,262 | 0,399 | 0,658 | 0,943 | 1,293 | 1,854 | 2,881 | 5,752 | 11,420 | 30,635 | 60,710 | 605,765 | 0,632 | 0,765 |
| 9 | 0,129 | 0,261 | 0,398 | 0,657 | 0,942 | 1,292 | 1,853 | 2,880 | 5,745 | 11,403 | 30,628 | 60,704 | 605,735 | 0,602 | 0,735 |
| 10 | 0,129 | 0,260 | 0,397 | 0,656 | 0,941 | 1,291 | 1,852 | 2,879 | 5,740 | 11,390 | 30,622 | 60,700 | 605,708 | 0,576 | 0,708 |
| 11 | 0,129 | 0,260 | 0,396 | 0,655 | 0,940 | 1,290 | 1,851 | 2,878 | 5,736 | 11,380 | 30,618 | 60,697 | 605,684 | 0,553 | 0,684 |
| 12 | 0,128 | 0,259 | 0,395 | 0,654 | 0,939 | 1,289 | 1,850 | 2,877 | 5,733 | 11,372 | 30,615 | 60,694 | 605,661 | 0,532 | 0,661 |
| 13 | 0,128 | 0,259 | 0,394 | 0,653 | 0,938 | 1,288 | 1,849 | 2,876 | 5,731 | 11,365 | 30,612 | 60,691 | 605,641 | 0,514 | 0,641 |
| 14 | 0,128 | 0,258 | 0,393 | 0,652 | 0,937 | 1,287 | 1,848 | 2,875 | 5,729 | 11,360 | 30,609 | 60,688 | 605,623 | 0,497 | 0,623 |
| 15 | 0,128 | 0,258 | 0,393 | 0,651 | 0,936 | 1,286 | 1,847 | 2,874 | 5,728 | 11,356 | 30,607 | 60,686 | 605,606 | 0,482 | 0,606 |
| 16 | 0,128 | 0,258 | 0,392 | 0,650 | 0,935 | 1,285 | 1,846 | 2,873 | 5,727 | 11,353 | 30,605 | 60,684 | 605,590 | 0,468 | 0,590 |
| 17 | 0,128 | 0,257 | 0,392 | 0,649 | 0,934 | 1,284 | 1,845 | 2,872 | 5,726 | 11,351 | 30,603 | 60,682 | 605,575 | 0,456 | 0,575 |
| 18 | 0,127 | 0,257 | 0,392 | 0,648 | 0,933 | 1,283 | 1,844 | 2,871 | 5,725 | 11,349 | 30,601 | 60,680 | 605,561 | 0,444 | 0,561 |
| 19 | 0,127 | 0,257 | 0,391 | 0,647 | 0,932 | 1,282 | 1,843 | 2,870 | 5,724 | 11,347 | 30,599 | 60,678 | 605,549 | 0,433 | 0,549 |
| 20 | 0,127 | 0,257 | 0,391 | 0,646 | 0,931 | 1,281 | 1,842 | 2,869 | 5,723 | 11,345 | 30,597 | 60,676 | 605,537 | 0,423 | 0,537 |
| 21 | 0,127 | 0,256 | 0,391 | 0,645 | 0,930 | 1,280 | 1,841 | 2,868 | 5,722 | 11,343 | 30,595 | 60,674 | 605,526 | 0,413 | 0,526 |
| 22 | 0,127 | 0,256 | 0,390 | 0,644 | 0,929 | 1,279 | 1,840 | 2,867 | 5,721 | 11,341 | 30,593 | 60,672 | 605,515 | 0,404 | 0,515 |
| 23 | 0,127 | 0,256 | 0,390 | 0,643 | 0,928 | 1,278 | 1,839 | 2,866 | 5,720 | 11,339 | 30,591 | 60,670 | 605,505 | 0,396 | 0,505 |
| 24 | 0,127 | 0,256 | 0,390 | 0,642 | 0,927 | 1,277 | 1,838 | 2,865 | 5,719 | 11,337 | 30,589 | 60,668 | 605,496 | 0,388 | 0,496 |
| 25 | 0,127 | 0,256 | 0,390 | 0,641 | 0,926 | 1,276 | 1,837 | 2,864 | 5,718 | 11,335 | 30,587 | 60,666 | 605,487 | 0,381 | 0,487 |
| 26 | 0,127 | 0,256 | 0,390 | 0,640 | 0,925 | 1,275 | 1,836 | 2,863 | 5,717 | 11,333 | 30,585 | 60,664 | 605,478 | 0,374 | 0,478 |
| 27 | 0,127 | 0,256 | 0,389 | 0,639 | 0,924 | 1,274 | 1,835 | 2,862 | 5,716 | 11,331 | 30,583 | 60,662 | 605,470 | 0,367 | 0,470 |
| 28 | 0,127 | 0,256 | 0,389 | 0,638 | 0,923 | 1,273 | 1,834 | 2,861 | 5,715 | 11,329 | 30,581 | 60,660 | 605,463 | 0,361 | 0,463 |
| 29 | 0,127 | 0,256 | 0,389 | 0,637 | 0,922 | 1,272 | 1,833 | 2,860 | 5,714 | 11,327 | 30,579 | 60,658 | 605,456 | 0,355 | 0,456 |
| 30 | 0,127 | 0,256 | 0,389 | 0,636 | 0,921 | 1,271 | 1,832 | 2,859 | 5,713 | 11,325 | 30,577 | 60,656 | 605,449 | 0,349 | 0,449 |
| 40 | 0,126 | 0,255 | 0,388 | 0,634 | 0,918 | 1,268 | 1,829 | 2,856 | 5,708 | 11,318 | 30,570 | 60,650 | 605,404 | 0,330 | 0,393 |
| 60 | 0,126 | 0,254 | 0,387 | 0,632 | 0,915 | 1,265 | 1,826 | 2,853 | 5,703 | 11,310 | 30,562 | 60,643 | 605,325 | 0,250 | 0,325 |
| 120 | 0,126 | 0,254 | 0,386 | 0,629 | 0,911 | 1,261 | 1,822 | 2,849 | 5,697 | 11,300 | 30,553 | 60,635 | 605,232 | 0,178 | 0,232 |
| z | 0,126 | 0,253 | 0,385 | 0,624 | 0,874 | 1,036 | 1,282 | 1,645 | 1,960 | 2,326 | 2,576 | 3,291 | | | |

② Die nächst kleinere Zahl zum gefundenen t-Wert (kritischer Wert) suchen.

①

③

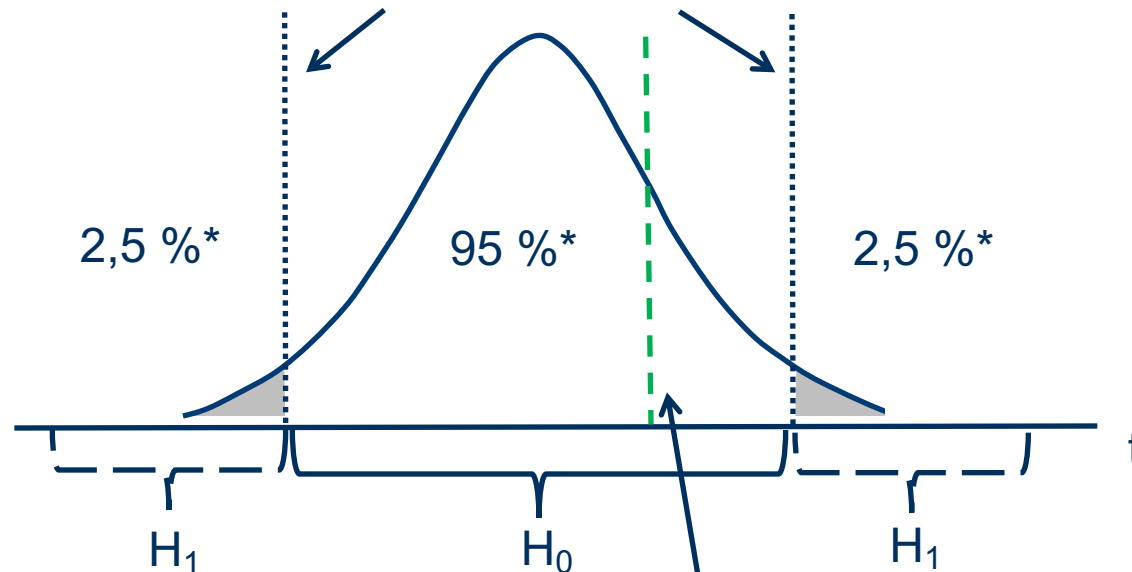
Ablezen der zugehörigen Fläche (Fläche von links).

Quelle der Abbildung: Bortz, J. (1999): Statistik für Sozialwissenschaftler. Berlin u.a.: Springer, S. 775

* Die Flächenanteile für negative t-Werte ergeben sich nach der Beziehung $p(-t_{df}) = 1 - p(t_{df})$

Festlegung des Signifikanzniveaus

(Willkürliche) Festlegung eines „kritischen“ Wertes ab dem der statistische Wert als „unwahrscheinlich“ gelten soll (Signifikanzniveau) z.B. insg. 0,05 (zwei mal 2,5)



* Die Prozente beschreiben, wie wahrscheinlich ein Ergebnis wäre, falls H_0 in Wirklichkeit gilt.

t-Wert (Beispiel)
Mit Hilfe des statistischen Tests (t-Test) empirisch ermittelte Wert.

Noch genauere Hypothesen: gerichtete und ungerichtete Alternativhypothesen

Bekannt:

Alternativhypothese allgemein H1: Die Mittelwerte unterscheiden sich signifikant. ($\mu_a \neq \mu_b$)

Neu:

1. gerichtete Alternativhypothese H1: Der Mittelwert der Nudeln aus Maschine A ist signifikant größer als aus Maschine B.

$$(\mu_a > \mu_b)$$

2. gerichtete Alternativhypothese H1: Der Mittelwert der Nudeln aus Maschine A ist signifikant kleiner als aus Maschine B.

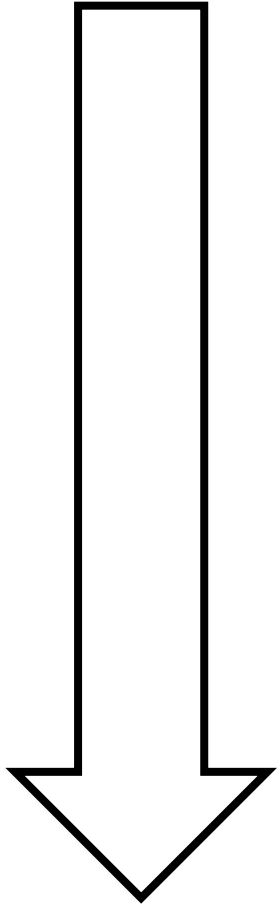
$$(\mu_a < \mu_b)$$

Es macht einen Unterschied, ob Sie eine gerichtete oder eine ungerichtete Alternativhypothese haben. Bei einer ungerichteten Hypothese

Hypothesen Quiz

| | a | b |
|---|---|---|
| Die Forschungshypothese stimmt meist mit a.) der Nullhypothese oder b.) der Alternativhypothese überein. | | |
| Richtig formuliert lautet eine Nullhypothese a.) $\mu_a = \mu_b$ oder b.) $\mu_a \neq \mu_b$ | | |
| Die Formulierung der Forschungshypothese kommt a.) vor oder b.) nach dem Aufstellen der Nullhypothese. | | |
| Wenn Sie wissen wollen, ob zwei Sorten Reis sich in der Korngröße unterscheiden heißt die Alternativhypothese a.) Sorte A und B unterscheiden sich nicht signifikant voneinander. b.) Sorte A unterscheidet sich signifikant von B. | | |

Praktisches Vorgehen bei einem Signifikanztests



Daten vorbereitet

Hypothese formulieren

Heraussuchen des richtigen Tests

Ermittlung der Testgröße und des Signifikanzniveau (mit Hilfe von Software).

Entscheidung fällen ob H_0 beibehalten wird oder zugunsten der H_1 zurückgewiesen wird.

Ergebnisse formulieren (siehe folgende Folien)

Formulierung der (nicht) signifikanter Ergebnisse

Signifikant: Wenn signifikantes Ergebnis ($p < 0,05$) → **Ablehnung/Zurückweisung** der Nullhypothese zugunsten der Alternativhypothese

- Beispiel: „Es konnte ein signifikanter Unterschied im IQ zwischen Hauptschülern und Gymnasiasten festgestellt werden.“

Nicht signifikant: Wenn kein signifikantes Ergebnis → **Beibehaltung** der Nullhypothese

- Beispiel: „Lehrer die sich in ihrer Freizeit mit Wirtschaft beschäftigen, zeigen kein signifikant höheres Interesse am Fach Wirtschaft als anderer Lehrer.“

Übung: Interpretation von signifikanten / nicht signifikanten Ergebnissen

Formulieren Sie folgende drei Ergebnis ausführlich:

Vergleich des Mittelwerts der Variable: „Motivation“
Gruppe 1: Auszubildende die selbstgesteuert lernen dürfen
Gruppe 2: Auszubildende mit einem festen Lehrplan
Ergebnis: **signifikant**

Vergleich des Mittelwerts der Variable: „Chancen für ein Job“
Gruppe 1: Schüler mit sechs Wochen Praktikum
Gruppe 2: Schüler mit einer Woche Praktikum
Ergebnis: **signifikant**

Vergleich des Mittelwerts der Variable: „Noten im Fach Statistik.“
Gruppe 1: Studentengruppe 1
Gruppe 2: Studentengruppe 2
Ergebnis: **insignifikant**

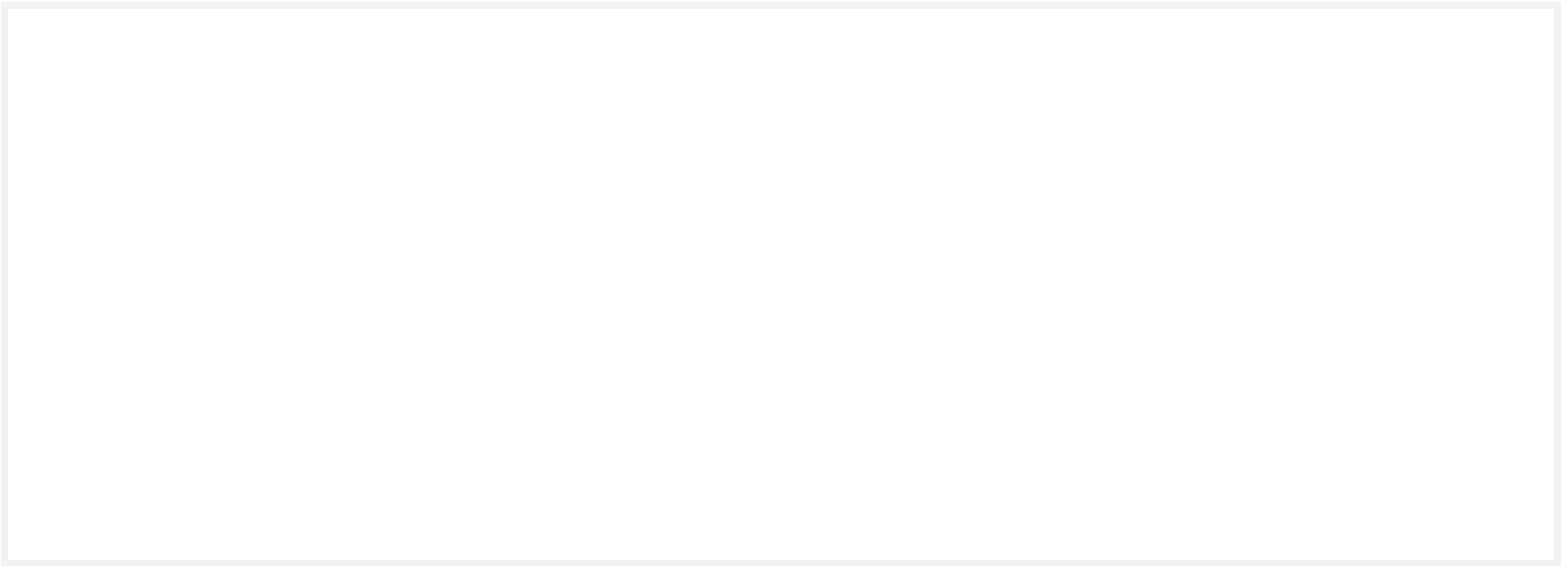
Übung Signifikanz

Was bedeutet Folgendes.

- In einem Artikel heißt es „Die Ergebnisse unserer Analyse sind signifikant ($p = 0,001$)“
- Drei Wissenschaftsteam machen alle drei jeweils eine Studie zur gleichen Frage. Sie erheben unabhängig Daten voneinander. Team A hat ein nicht signifikantes Ergebnis, Team B ein signifikantes Ergebnis ($p = 0,05$) und Team C ein signifikantes Ergebnis mit einem Signifikanzniveau von $p = 0,01$.
- Eine Forscherin erzählt, dass der Zusammenhang zwischen X und Y hochsignifikant ist.

Vorsicht bei der Interpretation von (Nicht-) Signifikanten Ergebnissen

Beispiel: Sie untersuchen den Unterschied in Anzahl der beruflichen Kontakte zwischen zwei Mitarbeiterteams (Außendienst / Innendienst). Sie finden keinen Unterschied. Gibt es wirklich keinen Unterschied?



Übung: Interpretation von signifikanten Ergebnissen

| Gruppe nach Jahren | n | Mittelwert der Gelenkigkeit (1-10) |
|--------------------|-----|------------------------------------|
| 3 - 22 | 65 | 9,4 |
| 23 - 42 | 234 | 9,5 |
| 43 - 62 | 645 | 8,3 |
| 63 - 82 | 26 | 5,5 |

Achtung die Daten sind frei erfunden.

Ein Forscherteam hat untersucht, ob es einen Zusammenhang zwischen dem Alter und der Gelenkigkeit gibt, bei Menschen, die intensiv klettern. Sie haben insgesamt 970 Leute untersucht. Sie haben alle Teilnehmer in Gruppen geteilt und alle Gruppen miteinander verglichen.

Sie finden keine signifikanten Unterschiede zwischen allen Gruppen.

1. Gibt es wirklich keine Unterschiede?
2. Warum wurden keine Unterschiede gefunden?
3. Wie könnte man das anders untersuchen?

Warnung vor zu vielen Signifikanztest (nicht vor zu viel einzelnen Studien)

Das Signifikanniveau sagt uns: Falls In Wirklichkeit kein Zusammenhang/Unterschied vorhanden ist, kann das gefundene Ergebnis trotzdem mal zufällig auftreten.

Bsp.: Dazu wird 20 Mal eine Population, in der **kein** Zusammenhang zwischen zwei Merkmalen vorhanden ist, untersucht. Dazu ziehen wir 20 mal eine Stichprobe aus dieser Population.

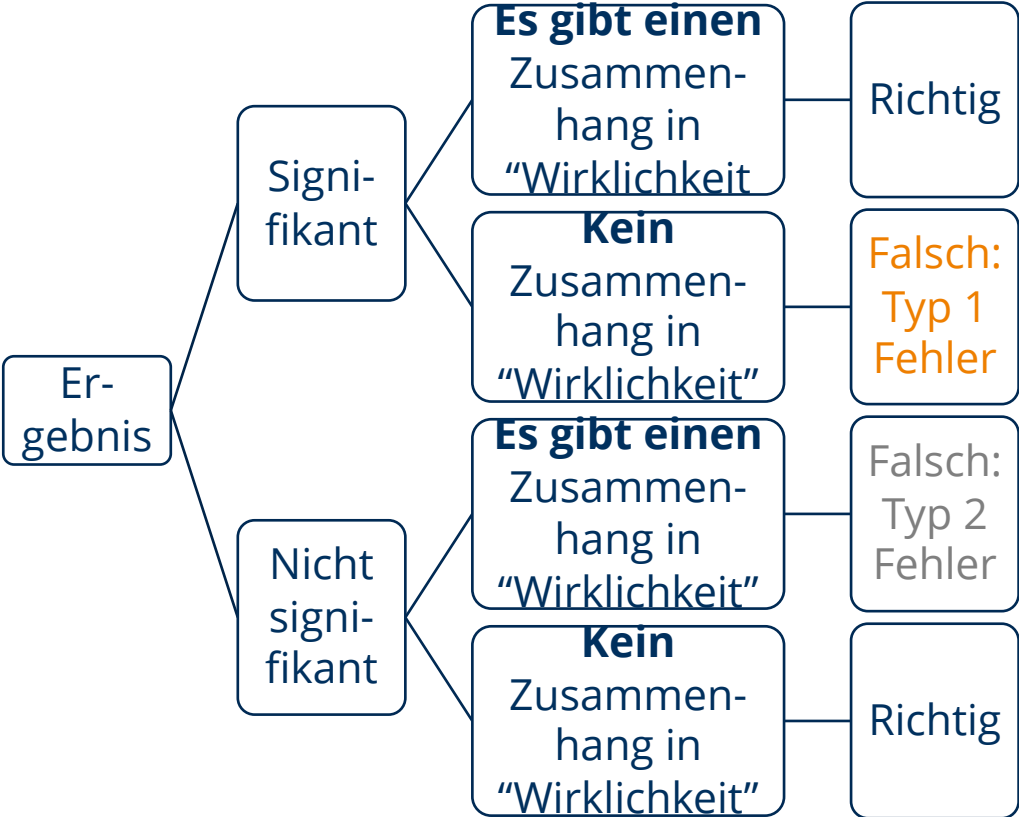
Das Signifikanzniveau wird auf 5 % festgelegt.

- Beim 20. Mal erbringt der Test ein signifikantes Ergebnis. (Obwohl in Wirklichkeit kein Zusammenhang da ist)
- Dies ist auch ein Grund, warum man wissenschaftliche Ergebnisse immer mehrmals überprüfen sollte.

| Stichprobe | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|----------------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|
| Ergebnis signifikant | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | JA |

Fehler von Signifikanztests

| 100 Stichproben Orange 5 % der Stichproben (Bei einem 0.05 Sign. Level werden 5 % signifikant selbst wenn in Wirklichkeit kein Zusammenhang da ist.) | | | | | | | | | |
|--|----|----|----|----|----|----|----|----|-----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
| 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 |
| 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 |
| 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 |
| 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 |
| 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 |
| 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 |
| 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100 |



Fehler bei Signifikanztests: eine andere Darstellung

Für H1 entscheiden, obwohl H0 gilt.

| | | In der Population/ In "Wahrheit" gilt | |
|--|----------------|--|---|
| | | H ₀ | H ₁ |
| Entscheidung aufgrund der Stichprobe zugunsten | H ₀ | Richtige Entscheidung | β -Fehler Fehler 2. Art |
| | H ₁ | α -Fehler Fehler 1. Art (Hängt von Signifikanzniveau ab) | Richtige Entscheidung |

Bei Verringerung des α -Niveaus (Signifikanzniveau) wird die Wahrscheinlichkeit für einen β -Fehler größer.



Übung Fehler bei Signifikanztest 1. und 2. Art

Normalerweise kennen wir die Wahrheit nicht, welchen Fehler wir machen, daher hier konstruierte Beispiele:

Eine Personalabteilung einer großen Firma weiß genau, dass bei der Einstellung von neuen Mitarbeiterinnen und Mitarbeitern keine Gehaltsunterschiede zwischen Frauen und Männern gemacht werden. Nun gibt es eine Studie des Betriebsrat, die zu einem anderen Ergebnis kommt. Welcher Art des Fehler wurde hier gemacht?

Eine Firma produziert handgefertigte Porzellanteller. Je teurer, desto aufwendiger die Malerei. Sie schauen sich die Teller an und kommen zu dem Schluss, dass es keinen Zusammenhang zwischen Preis und Aufwand gibt. Welche Art des Fehlers haben Sie gemacht?

Wir wissen alle, die Kinder kommen nicht von den Störchen. Aber es gibt diese falsche Annahme von einen Zusammenhang. Kann hier ein Fehler 2. Art gemacht wurden sein?

Fehler 1. Art und 2. Art sind unterschiedlich gefährlich, je nach Kontext

Sie messen die durchschnittliche Anzahl von Streichhölzern. Und wollen wissen, ob die Anzahl signifikant von den versprochenen 38 abweicht. Nun hängt es vom Kontext ab, wie gefährlich ein Fehler 1. und 2. Art ist.

A. Sie machen diese Untersuchung privat nur aus Neugier. Dann ist jeder Art von Fehler **O folgenlos, O bedeutend oder O tödlich?**

B. Sie machen diese Untersuchung für die Streichholzproduktion, dann wird jede Art des Fehlers **O folgenlos, O bedeutend oder O tödlich?**

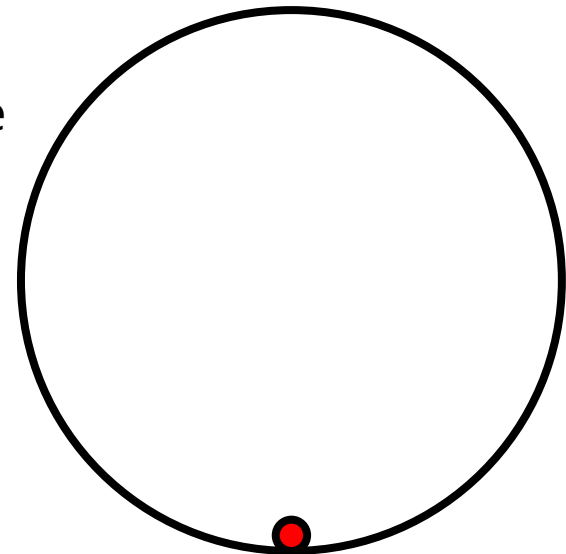
C. Sie machen diese Untersuchung für eine Antarktisexpedition, wo jedes Streichholz zählt. Dann ist jeder Fehler **O folgenlos, O bedeutend oder O tödlich.**

https://docs.google.com/forms/d/e/1FAIpQLSdQe6iv4cGuKjg6ObHp7NcBuLDQxS1IV2wYMEvZalqymZL1Pg/viewform?usp=sf_link



Beachten bei statistischen Tests: Größer der Stichprobe

- **Zu kleine Stichproben:** Selbst starke Zusammenhänge werden von statistischen Test nicht als signifikant angesehen.
 - größer Stichprobe
 - Wiederholung der Studie
- **Größere Stichproben:** schwächste Zusammenhänge werden signifikant
 - Relevanz ist neben Signifikanz auch wichtig

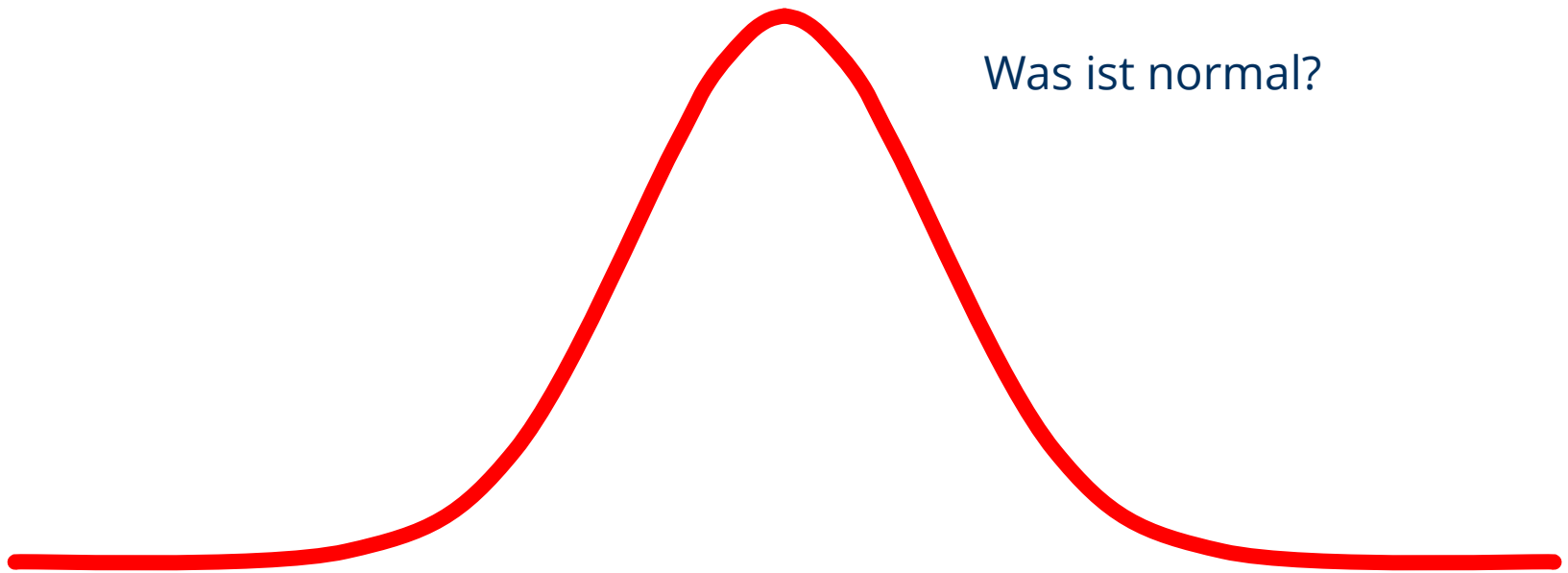


Gefahr bei statistischen Tests

- Alle möglichen Zusammenhänge werden erst mal getestet und dann erst baut man sich die **Theorie**
 - Erst Theorie, dann testen
 - Aber: BigDataAnalysen gehen einen anderen Weg
(Big Data hat aber viele Daten und oft sogar die Grundgesamtheit komplett)
- Statistische Test werden ohne Beachtung von **Voraussetzungen** getestet.
 - Viele dieser Voraussetzungen sind einleuchtend oder auf einen Blick zu überprüfen (Anzahl der Variablen/Anzahle der Fälle)
 - Andere Voraussetzungen muss man erst mit bestimmten statistischen Prozeduren testen (Test auf Normalverteilung)
- Anhand einer einzelnen statistischen Analyse wird auf die „Wahrheit“ geschlossen und gravierende Schlussfolgerungen gezogen
 - Wiederholung der Studie ist notwendig

Siehe auch Quatember, Andreas (2008): Statistik ohne Angst vor Formeln. Studienbuch für Wirtschafts- und Sozialwissenschaftler. Pearson Studium. 2. aktualisierte Auflage, S. 185ff.

Denken mit der Normalverteilung



Was ist normal?

Unser erster statistische Signifikanztest:

**auch Student t-Test oder Mittelwertsvergleich bei
unabhängigen Stichproben oder kurz t-Test**

t-Test bei zwei unabhängigen/ungepaarten Stichproben

| Vorname | Leistungstest | Gruppe |
|---------|---------------|--------|
| Ilsa | 33 | A |
| Peter | 44 | A |
| Lisa | 22 | B |
| Klaus | 46 | B |

Frage des t-Test: Sind die Mittelwerte in zwei Gruppen unterschiedlich groß?

H_0 = Mittelwerte in den zwei untersuchten Gruppen sind gleich

Skalenniveau:

abhängige Variable: metrisch;

unabhängige Variable:

nominal/ordinal mit zwei

Ausprägungen (z. B. Geschlecht; Gruppe A/B)

Metrische Variable für deren Mittelwert wir uns interessieren

Nominale Variable für die Gruppenzugehörigkeit

T-Test bei zwei unabhängigen/ungepaarten Stichproben: Die Formel

Das ist das Verfahren für den T-Test

(alle anderen Test sind sehr ähnlich):

Hier ist die Formel:

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

- Setzen sie alle Werte ein und rechnen Sie so den T-Wert aus.
- Den T-Wert vergleichen Sie in einer Tabelle.
- Dazu benötigen Sie die Freiheitsgrade: $F = n_x + n_y - 2$
- Tabelle: https://de.wikipedia.org/wiki/Studentsche_t-Verteilung
- Sie gehen in die Zeile mit dem ermittelten Freiheitsgrad.
- Dann gehen Sie in die Spalte mit dem gewünschtem Signifikanzniveau (meist 0,05 / 95 %), das ist der kritische Wert.
- Ist der kritische Wert kleiner als der ermittelte t-Wert, haben Sie ein signifikantes Ergebnis. (Negative Vorzeichen multiplizieren Sie hier mit -1, ignorieren es also)

| F | 0.95 |
|----|--------|
| 1 | 12,706 |
| 2 | 4,303 |
| 3 | 3,182 |
| 4 | 2,776 |
| 5 | 2,571 |
| 6 | 2,447 |
| 7 | 2,365 |
| 8 | 2,306 |
| 9 | 2,262 |
| 10 | 2,228 |
| 11 | 2,201 |
| 12 | 2,179 |
| 13 | 2,16 |
| 14 | 2,145 |
| 15 | 2,131 |
| 16 | 2,12 |
| 17 | 2,11 |

Ausschnitt aus der Tabelle der t-Quantile für 0,05

t-Test: Formulierung der Ergebnisse eines signifikanten Ergebnisses

- Sehr exakte Formulierung: *Die Nullhypothese, dass beide Geschlechter das gleiche Gewicht haben, kann nicht aufrechterhalten werden. Sie wird zugunsten der Alternativhypothese – das Gewicht unterscheidet sich zwischen Männern und Frauen – zurückgewiesen. Der zum t-Test zugehörige p-Wert (0,014) liegt unter dem Signifikanzniveau von 0,05*
- Umgangssprachliche Formulierung ohne Erwähnung von H_0 und H_1): *Die Ergebnisse weisen darauf hin, dass Frauen und Männer signifikant unterschiedliche viel Gewicht auf die Waage bringen. Der Mittelwert des Gewichts von Männer (84,58 Kilogramm) ist signifikant höher als der von Frauen (71,33 Kilogramm).*

Übung t-Test 1

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

Sie haben folgende Werte von Wettkampfteilnehmern. Es handelt sich um die erzielten Wettkampfpunkte in zwei Trainingsgruppen (A, B). Gruppe A hat Hallentraining gemacht und Gruppe B Outdoor-Training.

- Formulieren Sie richtige Null- und Alternativhypothesen.
- Errechnen Sie Mittelwerte, Varianz, den t-Wert und die Freiheitsgrade ($F = n_x + n_y - 2$)
- Vergleichen Sie t-Wert und Freiheitsgrade in der Tabelle.
- Überprüfen Sie, ob die Werte der beiden Gruppen sich signifikant voneinander unterscheiden (Signifikanzniveau von 0,05).
- Interpretieren Sie das Ergebnis inhaltlich. Welche Schlussfolgerung für die Trainingspraxis lassen sich ableiten?

| Gruppe A | Gruppe B |
|----------|----------|
| 56 | 12 |
| 45 | 34 |
| 74 | 85 |
| 84 | 75 |
| 34 | 54 |
| 57 | 37 |
| 64 | 56 |
| 78 | 37 |
| 64 | 65 |
| 35 | 42 |
| 97 | 78 |
| 65 | 68 |
| 43 | 98 |

Übung T-Test 2

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

Sie haben Kaufabsicht gegenüber einem neuen Produkt bei zwei Kundengruppen miteinander verglichen (Werte von 1-5; niedrigere Werte weniger Kaufabsicht). In jeder Gruppe waren 60 Personen. Der Mittelwert der Gruppe A = 2,1 und der Gruppe B = 2,7. Die Varianz in Gruppe A beträgt 1,06 und in Gruppe B 1,53

1. Errechnen Sie den t-Wert anhand der Formel und die Freiheitsgrade.
2. Legen Sie ein Signifikanzniveau fest, wählen Sie z.B. 0,01 oder 0,001.
3. Ermitteln Sie in der Tabelle*, ob der t-Wert zu einem signifikanten oder nicht signifikanten Ergebnis führt.
4. Interpretieren Sie das Ergebnis rein statistisch und indem sie auch auf praktische Schlussfolgerung eingehen.

t-Test bei zwei unabhängigen/ungepaarten Stichproben in Excel I

2 Variablen sind notwendig

1. Variable enthält die Gruppen (z.B. Frau/Mann), das gesamte Datenblatt muss danach sortiert werden

2. metrische Variable, deren Mittelwert uns interessiert

→ Daten → Datenanalyse → Zweistichproben-t-Test (Gleicher Varianzen/Unterschiedlicher Varianzen)

Ob die Varianzen gleich sind müsste vorher getestet werden

... Fortsetzung ...

Das Excel-Add-In
Datenanalyse
muss aktiviert sein

t-Test bei zwei unabhängigen/ungepaarten Stichproben in Excel II

In „Bereich Variable A“ die Daten der 2. Variable ziehen, die alle zur 1. Gruppe (z.B. Frauen) gehören.

In „Bereich Variable B“ die Daten der 2. Variable ziehen, die zur 2. Gruppe gehören.

→ Ok

| A | B | C | D | E | F | G | H | I |
|------------|---------|---|---|---|---|---|---|---|
| Geschlecht | Gewicht | | | | | | | |
| 1 | 76 | | | | | | | |
| 1 | 78 | | | | | | | |
| 1 | 65 | | | | | | | |
| 1 | 69 | | | | | | | |
| 1 | 78 | | | | | | | |
| 1 | 62 | | | | | | | |
| 0 | 79 | | | | | | | |
| 0 | 70 | | | | | | | |
| 0 | 80 | | | | | | | |
| 0 | 87 | | | | | | | |
| 0 | 90 | | | | | | | |
| 0 | 98 | | | | | | | |
| 0 | 88 | | | | | | | |

Zweistichproben t-Test: Gleicher Varianzen

Eingabe

Bereich Variable A:

Bereich Variable B:

Hypothetische Differenz der Mittelwerte:

Beschriftungen

Alpha:

Ausgabe

Ausgabebereich:

Neues Tabellenblatt:

Neue Arbeitsmappe

OK

Abbrechen

Hilfe

t-Test: Output

| Zweistichproben t-Test unter der Annahme gleicher Varianzen | | |
|---|------------|------------|
| | Variable 1 | Variable 2 |
| Mittelwert | 71,333 | 84,571 |
| Varianz | 48,667 | 81,952 |
| Beobachtungen | 6 | 7 |
| Gepoolte Varianz | 66,823 | |
| Hypothetische Differenz der Mittelwerte | 0 | |
| Freiheitsgrade (df) | 11 | |
| t-Statistik | -2,911 | |
| P(T<=t) einseitig | 0,007 | |
| Kritischer t-Wert bei einseitigem t-Test | 1,796 | |
| P(T<=t) zweiseitig | 0,014 | |
| Kritischer t-Wert bei zweiseitigem t-Test | 2,201 | |

Mittelwerte beider Gruppen

p-Wert. hier liegt der p-Wert unter (unserem festgelegtem) Signifikanzniveau von 0,05 → signifikante Unterschiede

In der Verschriftlichung unbedingt die Mittelwerte und den p-Wert veröffentlichen. Profis veröffentlichen auch t-Statistik und Freiheitsgrade.

t-Test: Verletzung der Voraussetzung. Praktischer Hinweis

Bei dem vorangegangenen Beispiel hatten wir angenommen die Varianzen des Gewichts bei Frauen und Männern sind gleich!

Tatsächlich scheinen sie nicht gleich gewesen zu sein, weshalb unter Datenanalyse eigentlich der Test „Zweistichproben t-Test: Unterschiedlicher Varianzen“ durchgeführt hätte werden sollen!

... Aber: der t-Test ist sehr robust gegen Verstöße der Voraussetzungen und meist kommen beide t-Testvarianten zu sehr ähnlichen Ergebnissen, so dass nur in Zweifelsfällen die Missachtung der Voraussetzungen die Ergebnisse beeinflusst.

Gerichtete Alternativhypothesen und einseitiger Signifikanztest.

Die Standardalternativhypothese lautet: Die Mittelwerte unterscheiden sich nicht.

Die gerichtete Alternativhypothese ist etwas genauer und behauptet nun, dass der Mittelwert der einen Stichprobe größer/kleiner als der Mittelwert der anderen Stichprobe ist.

Bei gerichteten Alternativhypothesen kann der p-Wert halbiert werden. Es handelt sich dann um einen einseitigen Signifikanztest. Der Standardfall ist der zweiseitige Signifikanztest.

Beispiel mit p-Wert von 0,07:

$$0,07 / 2 = 0,035$$

Zweiseitig: 0,07 liegt über 0,05 und ist nicht signifikant

Einseitig: 0,035 liegt unter 0,05 und ist signifikant

Haben wir ordinal skalierte Daten aber die gleiche Frage wie beim t-Test, verwendet man einen nichtparametrische Test: Man-Whitney U-Test (unabhängige Stichproben) Wilcoxon-Rangsummen-Test (abhängige Stichproben) die in Excel leider kompliziert umzusetzen sind.

Übung gerichtete Alternativhypothese

Betrachten Sie die Ausgabe für einen t-Test in Excel unten. Welchen Wert interpretieren Sie wenn sie eine gerichtete Hypothese haben? Wie ist noch mal der Zusammenhang zwischen einseitigen und zweiseitigen p-Wert?

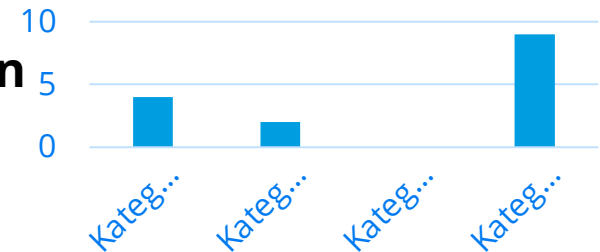
Sie wetten mit Ihrer Nachbarin, dass im Durchschnitt die gelben Rosen im Garten größer werden als die roten Rosen. Mit Ihrer Tante wetten Sie nur, dass die Länge der gelben Rosen sich signifikant von den roten Rosen unterscheiden. Mit wem haben sie eine gerichtete und mit wem eine ungerichtete Hypothese ausgemacht?

| Zweistichproben t-Test unter der Annahme gleicher Varianzen | | |
|---|------------|------------|
| | Variable 1 | Variable 2 |
| Mittelwert | 71,333 | 84,571 |
| Varianz | 48,667 | 81,952 |
| Beobachtungen | 6 | 7 |
| Gepoolte Varianz | 66,823 | |
| Hypothetische Differenz der Mittelwerte | 0 | |
| Freiheitsgrade (df) | 11 | |
| t-Statistik | -2,911 | |
| P(T<=t) einseitig | 0,007 | |
| Kritischer t-Wert bei einseitigem t-Test | 1,796 | |
| P(T<=t) zweiseitig | 0,014 | |
| Kritischer t-Wert bei zweiseitigem t-Test | 2,201 | |

Besondere Voraussetzung beim t-Test

(Annähernd) Normalverteilung der abhängigen Variable

Sinn: Ansonsten könnte es sein, dass der Mittelwert keine sinnvolle Aussage über die Datenverteilung macht. (Siehe Beispiel in den Daten)



Gleichheit der Varianzen

Wie testet man auf Gleichheit der Varianzen?

Mit dem Levene-Test:

Sinn: Sind die Varianzen beider Stichproben gleich? (Gleich bedeutet nicht identisch)

H_0 = Die Varianzen beider Stichproben sind gleich.

Skalenniveau: metrisch



Beispiel in den Abbildungen:
Hier sind die Varianzen nicht gleich.
Die unterer Stichprobe hat viel kleinere Varianz

T-Test Online

Aufgabe: Öffnen Sie eine interaktive Seite zum t-Test und führen sie einen t-Test dort durch, vergleichen Sie mit den Ergebnissen aus Excel.

<https://statpages.info/>

Aufgabe: Ordnen Sie die Schritte die beim T-Test gemacht werden in eine logische Reihenfolge: https://miro.com/app/board/uXjVO18-aGg=

Übersicht weitere Test: Sie kennen einen Test, können Sie trotzdem insgesamt vier Test ihren Anwendungen zuordnen?

Es gibt verschiedene Arten von t-Tests

1. Vergleich eines **vorgegeben** Mittelwerts/Anteilswerts aus der Grundgesamtheit mit dem Mittelwert/Anteilswert aus der Stichprobe (für Mittelwert: Gaußtest)
2. t-Test bei unabhängigen Stichproben
3. Vergleich von Differenzen von **Anteilswerten**
4. Vergleich von gepaarten/**abhängigen** Stichproben

Ordnen Sie die Beispiele zu:

- a. Partei A hat signifikant weniger Sitze im Parlament als Partei B.
- b. Ein Kaufhaus misst vor und nach der Einführung einer Werbung die Anzahl an täglichen Kunden im Geschäft für jede Abteilung (Süßigkeiten, Haushalt, ...)
- c. Sie wollen wissen, ob dunkle oder Zartbitterschokoladenkugeln Ihnen mehr schmeckt.
- d. In der Bevölkerung hat einen IQ von 100. Eine Direktorin möchte nun wissen, ob ihre Schüler über diesem Wert liegen.

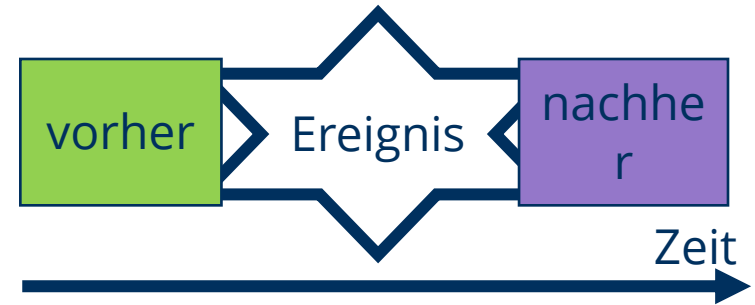
Zweistichprobentest bei gepaarten / abhängigen Stichproben

Zweistichprobentest bei gepaarten/abhängigen Stichproben

Sinn: Unterscheidet sich der Mittelwert von zwei Variablen.

H_0 = Mittelwerte in den zwei Variablen sind gleich.

Skalenniveau: Beide Variablen metrisch



| Name | Leistungstest vorher | Leistungstest nachher |
|-------|----------------------|-----------------------|
| Ilisa | 2 | 1 |
| Peter | 4 | 6 |
| Lisa | 9 | 7 |
| Hans | 2 | 5 |

Zweistichproben Test bei gepaarten/abhängigen Stichproben

$$Z = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

$$\bar{d} =$$

Mittelwert der Differenz aller Paare

$$s_d =$$

Standardabweichung der Differenzen

| Name | Leistung stest vorher | Leistung stest nachher | Differenz |
|----------------------|-----------------------|------------------------|-----------|
| Ilsa | 2 | 1 | 1 |
| Peter | 4 | 6 | -2 |
| Lisa | 9 | 7 | 2 |
| Hans | 2 | 5 | -3 |
| Mittelwert \bar{d} | | | -0,5 |
| Standardabweichung | | | 2,38 |
| Z-Wert | | | |

Der ermittelte Z-Wert wird nun wieder in einer Tabelle – hier Z-Verteilung, für die Standardnormalverteilung nach geschaut. Er entspricht ca. 34,46 % Perzentil und liegt damit über den Signifikanzniveau von 0,05 d.h. die Differenzen sind nicht signifikant. Es gibt keinen signifikanten Unterschied zwischen dem Leistungstest „vorher“ und „nachher“. Das „Ereignis“ hatte keine Auswirkung auf die Leistung.

t-Test bei gepaarten/abhängigen Stichproben: in Excel

Das Excel-Add-In
Datenanalyse
muss aktiviert
sein

→ Daten → Datenanalyse → Zweistichproben-t-Test bei abhängigen Stichproben

| A | B |
|-------------------------|--------------------------|
| Ergebnisse vor dem Test | Ergebnisse nach dem Test |
| 11 | 21 |
| 12 | 22 |
| 13 | 23 |
| 14 | 24 |
| 15 | 26 |

Zweistichproben t-Test bei abhängigen Stichproben

Eingabe

Bereich Variable A:

Bereich Variable B:

Hypothetische Differenz der Mittelwerte:

Beschriftungen

Alpha:

Ausgabe

Ausgabebereich:

Neues Tabellenblatt:

Neue Arbeitsmappe

OK
Abbrechen
Hilfe

t-Test bei gepaarten/abhängigen Stichproben: Interpretation

| Zweistichproben t-Test bei abhängigen Stichproben (Paarvergleichstest) | | |
|--|--------------------------------|---------------------------------|
| | <i>Ergebnisse vor dem Test</i> | <i>Ergebnisse nach dem Test</i> |
| Mittelwert | 13 | 23,2 |
| Varianz | 2,5 | 3,7 |
| Beobachtungen | 5 | 5 |
| Pearson Korrelation | 0,986 | |
| Hypothetische Differenz der Mittelwerte | 0 | |
| Freiheitsgrade (df) | 4 | |
| t-Statistik | -51 | |
| P(T<=t) einseitig | 0,000 | |
| Kritischer t-Wert bei einseitigem t-Test | 2,132 | |
| P(T<=t) zweiseitig | 0,000 | |
| Kritischer t-Wert bei zweiseitigem t-Test | 2,776 | |

Die Mittelwerte beider Variablen (Stichproben)

Der p-Wert des t-Tests

Die H_0 wird abgelehnt, da der zum t-Test zugehörige p-Wert unter dem Signifikanzniveau von 0,05 liegt.