



LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Material zur Fortbildung



Die Materialien zur Fortbildung finden Sie hier:

www.tinyurl.com/ytypzvtv



Passwort: **FoBiNeufahrn24**

**Nach der Fortbildung finden Sie hier auch den
Foliensatz zur Fortbildung sowie die Lösungsblätter.**

LMU

LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

**MATHEMATISCHES INSTITUT
DIDAKTIK DER MATHEMATIK**



Das Modul Statistik im Vertiefungskurs Mathematik in der neuen Oberstufe

**Lehrkräftefortbildung,
Neufahrn bei Freising**

Prof. Dr. Karin Binder
1. Juli 2024





LUDWIG-
MAXIMILIANS-
UNIVERSITÄT
MÜNCHEN

Material zur Fortbildung



Die Materialien zur Fortbildung finden Sie hier:

www.tinyurl.com/ytypzvtc



Passwort: **FoBiNeufahrn24**

**Nach der Fortbildung finden Sie hier auch den
Foliensatz zur Fortbildung sowie die Lösungsblätter.**

Das neue Statistik-Modul im Vertiefungskurs



Die Schülerinnen und Schüler ...

- erläutern anhand von Beispielen die wesentlichen Eigenschaften unterschiedlichen **Skalenniveaus** und unterscheiden dabei **nominalskalierte, ordinalskalierte** und **metrischskalierte** Variable.
- stellen **lineare Regression** als Zusammenhang zwischen den Werten von Variablen im Sinne einer Vorhersage dar. **Korrelationskoeffizienten** sind ein Maß für den Zusammenhang zweier Größen. Dass bei **linearen Regression** Fehlinterpertationen auftreten können (z. B. **Simpson-Paradoxon**), weisen sie anhand von geeigneten Beispielen graphisch und rechnerisch nach.
- beschreiben das grundsätzliche Vorgehen beim **Chi-Quadrat-Unabhängigkeitstest** oder beim **t-Test** und bestimmen mithilfe des betrachteten Testverfahrens bei geeigneten Beispielen von Datensätzen **p-Werte** unter Verwendung einer Statistik-Software. Sie interpretieren ihre Ergebnisse im Sachzusammenhang.
- prüfen mediale Darstellungen von Daten auf Korrektheit und analysieren den manipulativen Charakter fehlerhafter Darstellungen anhand typischer Beispiele.

Bitte
was???

Chi-Quadrat- und
t-Test hat hoffentlich
nichts mit
Hypothesentests zu
tun ☹️

Ist das
nicht viel
zu viel?



Gliederung

Teil 1

- Die grundlegende Idee des Statistik-Moduls
- Skalenniveaus
- Forschungsfragen



Teil 2

- χ^2 -Tests



Teil 3

- Mit Simulationen zum p-Wert
- t-Tests



Teil 4

- Fehlvorstellungen zu p-Werten
- Korrelation und Regression
- ggfs. Simpson-Paradoxon



Die grundlegende Idee des Moduls „Statistik“





Warum sogar noch mehr von diesen (schlimmen?) Hypothesentests?

Ein Großteil Ihrer Schülerinnen und Schüler hat später im Studium...



- Statistik für ... Biolog:innen
- ... Wirtschaftswissenschaftler:innen
- ... Sozialwissenschaftler:innen
- ... Mediziner:innen
-

Einer der wichtigsten und umfangreichsten Bestandteile dieser Vorlesungen sind **Hypothesentest**.

Sie können im **Modul „Statistik“** – wie in keinem der anderen Module – einen echten Realitätsbezug und wichtige Kompetenzen für ein erfolgreiches Studium schaffen!



Warum jetzt noch mehr Tests neben dem bereits unterrichteten Binomialtest?

Weber (2020)
Krauss et al. (i.V.)

Die bislang in der Schule unterrichteten Hypothesentests stimmen in drei Aspekten nicht mit Hypothesentests aus der Realität (Wissenschaft und Wirtschaft) überein:

Kontext

Testart

Testprozedere

Realitätsabgleich

Welche **Kontexte** werden
in Schulbüchern bislang betrachtet?

Weber (2020)
Krauss et al. (i.V.)

Schule

11 

Der Bürgermeister einer Stadt plant, den Bau eines Fußballstadions zu bezuschussen, wenn

1

Der Chef einer Großküche bestellt bei seinem Obstlieferanten eine große Lieferung Zwetschgen. Der Lieferant will einen Preisnachlass einräumen, falls der Anteil p der Zwetschgen mit Wurm 10% übersteigt. Um dies zu prüfen, wird folgende Vereinbarung getroffen: Der Lieferung werden 50 Zwetschgen entnommen. Enthalten mehr als sieben davon einen Wurm, so wird angenommen, p sei größer als 10%.

b) Wie testet die Bürgerinitiative?

Bei welchen Ergebnissen sieht sie sich bestätigt?

Realität

Verdienen Frauen im Schnitt weniger als Männer?

Wirkt der neue Corona-Impfstoff besser als der vorherige?

Gibt es einen Zusammenhang zwischen dem CO₂-Ausstoß und der mittleren Temperatur?

Realitätsabgleich ▪ Kontext

Weber (2020)
Krauss et al. (i.V.)

Ergebnisse aus einer Fragebogenstudie mit $N = 50$ Lehramtsstudierenden des Unterrichtsfachs Mathematik (Universität Regensburg)

Item	(fast) nie	selten	manchmal	oft	(fast) immer
Prinzipiell werden Binomialtests bei der Lebensmittelkontrolle in der Realität ... verwendet					

40% unserer Schülerinnen und Schüler glauben uns (leider!) die Verwendung von Hypothesentests in diesen Kontexten!

Realitätsabgleich ▪ Testart

Weber (2020)
Krauss et al. (i.V.)

Schule
<p>➤➤ Binomialtests</p>

Realität
<p>Die häufigsten Tests:</p> <ul style="list-style-type: none"> ➤➤ t-Tests ➤➤ Chi-Quadrat-Tests ➤➤ Tests auf Signifikanz von Korrelationskoeffizienten

- Binomialtests kommen in der **Realität nur sehr selten** vor!
- Binomialtests sind rechnerisch sogar am **schwersten**.

Realitätsabgleich ■ Testprozedere

Weber (2020)
Krauss et al. (i.V.)

Schule

1. Festlegen der Testgröße Z und des Stichprobenumfangs n .
2. Die Nullhypothese wird mathematisch formuliert. Wie die Nullhypothese gewählt wird, ist keine Frage der Mathematik, sondern der konkreten Anwendungssituation. Man wird sie so wählen, dass ihre irrtümliche Ablehnung folgenschwerer ist.
3. Festlegen des Signifikanzniveaus α :
Welcher Wert für α festgelegt wird, hängt davon ab, welche Wahrscheinlichkeit für den Fehler 1. Art man zuzulassen bereit ist.
4. Bestimmen der Entscheidungsregel, d.h. Konstruktion des kritischen Bereichs K (Ablehnungsbereich): Dabei ist es hilfreich, sich zunächst zu überlegen, welcher Rand des Wertebereichs der Testgröße Z sicher im kritischen Bereich liegen wird.

**Vorsicht:
The Null-Ritual**

Realität

2 Klicks in einer Statistik-Software

 $p\text{-Wert} \leq 0.05$

Signifikant

z.B. Das Gehalt von Männern und Frauen unterscheidet sich signifikant.

 $p\text{-Wert} > 0.05$

Nicht signifikant

z.B. Das Gehalt von Frauen unterscheidet sich nicht signifikant vom Gehalt der Männer.

Der unterrichtliche Fokus könnte so mehr auf den **dahinterliegenden Konzepten** liegen und weniger auf der Prozedur.



Erstes Fazit

- Das Modul „Statistik“ wirkt auf den ersten Blick sehr **umfangreich**.
- Es beinhaltet möglicherweise einige **Begriffe**, mit denen Sie sich erst vertraut machen müssen.

- Nicht zuletzt die Corona-Pandemie zeigt: Statistisches Denken ist inzwischen so wichtig wie Lesen und Schreiben
- Das Statistik-Modul behandelt sämtliche statistische Themen sehr **realitätsbezogen** und schafft ein breites Verständnis für statistische Zusammenhänge und Verfahren.
- Das Modul „Statistik“ kann viele Schüler:innen hervorragend auf das **Studium** vorbereiten.

Fachliche Grundlage: Skalenniveaus





Fachbegriffe: Merkmale und Merkmalsausprägungen

Im Rahmen von Erhebungen werden interessierende **Merkmale** (bzw. Variablen) erfasst, wie z.B. Geschlecht, Körpergröße einer Person, durchschnittliches Gehalt.

Die verschiedenen Werte, die ein Merkmal annehmen kann, werden **Merkmalsausprägungen** genannt.

Beispiele:

Merkmal
Körpergröße

Merkmalsausprägungen
z.B. 1,76m oder 1,8m

**Kategoriale Variable
durch Klassenbildung**

[1,10m – 1,20m[; [1,20m – 1,30m[; ...; [2,00m-2,10m[
klein; mittel; groß

Skalenniveaus von Merkmalen



Skala	Erlaubte mathematische Operationen	Charakterisierung	Beispiele
Nominalskala	=, ≠ unterscheiden	Größen <u>unterschiede</u> : keine Bedeutung	Kontonummer

Skalenniveaus von Merkmalen



Skala		Erlaubte mathematische Operationen	Charakterisierung	Beispiele
Nominalskala		$=, \neq$ unterscheiden	Größen <u>unterschiede</u> : keine Bedeutung	Kontonummer
Ordinalskala		$=, \neq <, >$ Zusätzlich ordnen	<u>Abstände</u> können un- terschiedlich groß sein	Bundesligatabelle, Schwierigkeitsgrad beim Klettern

Skalenniveaus von Merkmalen



Skala		Erlaubte mathematische Operationen	Charakterisierung	Beispiele
Nominalskala		$=, \neq$ unterscheiden	Größen <u>unterschiede</u> : keine Bedeutung	Kontonummer
Ordinalskala		$=, \neq <, >$ Zusätzlich ordnen	<u>Abstände</u> können un- terschiedlich groß sein	Bundesligatabelle, Schwierigkeitsgrad beim Klettern
Metrische Skala bzw. Kardinal- skala	Intervall- skala	$=, \neq <, > +, -$ Zusätzlich Abstände vergleichen, Mittelwerte bilden	Kein absoluter Nullpunkt Keine natürliche Einheit	Temperatur in Grad Celsius

Skalenniveaus von Merkmalen



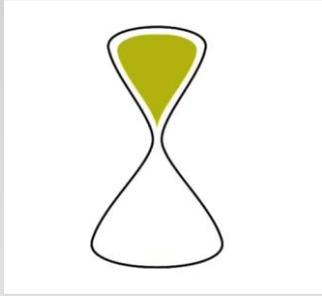
Skala		Erlaubte mathematische Operationen	Charakterisierung	Beispiele
Nominalskala		$=, \neq$ unterscheiden	Größen <u>unterschiede</u> : keine Bedeutung	Kontonummer
Ordinalskala		$=, \neq <, >$ Zusätzlich ordnen	<u>Abstände</u> können un- terschiedlich groß sein	Bundesligatabelle, Schwierigkeitsgrad beim Klettern
Metrische Skala bzw. Kardinal- skala	Intervall- skala	$=, \neq <, > +, -$ Zusätzlich Abstände vergleichen, Mittelwerte bilden	Kein absoluter Nullpunkt Keine natürliche Einheit	Temperatur in Grad Celsius
	Verhältnis- skala	$=, \neq <, > +, - \cdot, :$ Zusätzlich Verhältnisse vergleichen	Absoluter Nullpunkt Keine natürliche Einheit (Größen mit Einheiten)	Preise, Längen, Gewichte, Volumina

Skalenniveaus von Merkmalen



Kategoriale Variablen

Skala		Erlaubte mathematische Operationen	Charakterisierung	Beispiele
Nominalskala		$=, \neq$ unterscheiden	Größen <u>unterschiede</u> : keine Bedeutung	Kontonummer
Ordinalskala		$=, \neq <, >$ Zusätzlich ordnen	<u>Abstände</u> können un- terschiedlich groß sein	Bundesligatabelle, Schwierigkeitsgrad beim Klettern
Metrische Skala bzw. Kardinal- skala	Intervall- skala	$=, \neq <, > +, -$ Zusätzlich Abstände vergleichen, Mittelwerte bilden	Kein absoluter Nullpunkt Keine natürliche Einheit	Temperatur in Grad Celsius
	Verhältnis- skala	$=, \neq <, > +, - \cdot, :$ Zusätzlich Verhältnisse vergleichen	Absoluter Nullpunkt Keine natürliche Einheit (Größen mit Einheiten)	Preise, Längen, Gewichte, Volumina
	Absolut- skala		Absoluter Nullpunkt Natürliche Einheit	Einwohnerzahlen, Anzahlen



Zeit: 7 Minuten

Workshop-Phase

Bearbeiten Sie bitte
Übungsblatt 1



Skalenniveaus von Merkmalen



a) Geben Sie für die nachfolgenden Merkmale jeweils passende (beispielhafte) Merkmalsausprägungen und das zugehörige Skalenniveau an (Nominalskala, Ordinalskala, Intervallskala, Verhältnisskala, Absolutskala).

	Passende Merkmalsausprägungen	Skalenniveau
Haustier		
Jahreszahl		
Fahrzeugklasse		
Anzahl an Geschwistern		
Durchschnittsgehalt		



Mögliche Lösung:

	Passende Merkmalsausprägungen	Skalenniveau
Haustier	Hund, Katze, Maus...	nominal
Jahreszahl	1982; 2023; 125 v. Chr.	intervallskaliert
Fahrzeugklasse	Kleinwagen, Mittelklasse, Oberklasse	ordinal
Anzahl an Geschwistern	0, 1, 2, ...	Absolutskala
Durchschnittsgehalt	3 731 Euro, 2 842 Euro, ...	Verhältnisskala



Mögliche Lösung:

	Passende Merkmalsausprägungen	Skalenniveau
Haustier	Hund, Katze, Maus...	nominal
Jahreszahl	1982; 2023; 125 v. Chr. Steinzeit, Bronzezeit, Eisenzeit...	intervallskaliert
Fahrzeugklasse	Kleinwagen, Mittelklasse, Oberklasse	ordinal
Anzahl an Geschwistern	0, 1, 2, ... Ja/Nein	Absolutskala
Durchschnittsgehalt	3 731 Euro, 2 842 Euro, ... Ohne Einkommen, Geringverdiener, ..., Spitzenverdiener	Verhältnisskala



	Unterschiedshypothese	Zusammenhangshypothese	Unabhängigkeitshypothese
Hypothesentest	t-Test Mittelwertvergleichstest	Korrelation Test, ob sich Korrelationskoeffizient von 0 unterscheidet	Chi-Quadrat-Test
Skalenniveau	Merkmal 1: metrisch Merkmal 2: kategorial (2 Gruppen)	Merkmal 1: metrisch Merkmal 2: metrisch	Merkmal 1: kategorial Merkmal 2: kategorial (je auch mehr als 2 Gruppen möglich!)
Beispiel Forschungsfrage	Unterscheidet sich das Gehalt von Männern und Frauen?	Hängen Intelligenz und Schulleistung zusammen?	Hängt die Genesung (krank vs. gesund) von der Gruppenzugehörigkeit (Placebo vs. Medikament) ab?
Beispiel Forschungshypothese (ungerichtet)	Das Gehalt von Männern und Frauen unterscheidet sich!	Intelligenz und Schulleistung hängen zusammen!	Die Genesung hängt von der Gruppenzugehörigkeit ab?
Beispiel Nullhypothese (ungerichtet)	Das Gehalt von Männern und Frauen unterscheidet sich <u>nicht</u> !	Intelligenz und Schulleistung hängen <u>nicht</u> zusammen!	Die Genesung hängt <u>nicht</u> von der Gruppenzugehörigkeit ab?
Teststatistik (Prüfgröße)	$t = \frac{\bar{x}_1 - \bar{x}_2}{s}$ mit den beiden Mittelwerten im Zähler; s: Standardabweichung	$r_{xy} = \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i$ Bravais-Pearson-Korrelationskoeffizient	χ^2 Vergleich zweier Vierfeldertafeln (real vs. stochastisch unabhängig)
Verteilung	Student-t-Verteilung		Chi-Quadrat-Verteilung
Effektstärke	$Cohens\ d = \frac{\bar{x}_1 - \bar{x}_2}{s}$ 0,2: kleiner Effekt 0,5: mittlerer Effekt 0,8: großer Effekt	$r_{xy} = \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i$ Der Korrelationskoeffizient ist selbst die Effektstärke 0,1: klein; 0,3: mittel; 0,5: groß	$K = \sqrt{\frac{k}{k-1}} \cdot \sqrt{\frac{\chi^2}{n + \chi^2}}$ Korrigierter Kontingenzkoeffizient 0,1: klein; 0,3: mittel; 0,5: groß
CODAP	Differenz der Mittelwerte 	Korrelation von X und Y 	Test auf Unabhängigkeit

Fachliche Grundlage: Verschiedene Arten von Forschungsfragen





Typische Forschungsfragen für Hypothesentests

Typische Hypothesen, die mit Hypothesentests beantwortet werden:

**Unterschieds-
hypothesen**

Beispiel:
Das Einkommen unterscheidet sich je nach Geschlecht (Frauen verdienen weniger als Männer)

⇒ **t-Test**

**Zusammenhangs-
hypothesen**

Beispiel:
Durchschnittstemperaturen hängen mit dem CO₂-Ausstoß zusammen

⇒ **Test auf Signifikanz von Korrelationskoeffizienten**

**Abhängigkeits-
hypothesen**

Beispiel:
Die Genesung hängt von der Gruppenzugehörigkeit (Medikamentengruppe vs. Placebo-Gruppe) ab.

⇒ **Chi-Quadrat-Unabhängigkeitstest**

Hypothesen sind sozusagen die „vermutete Antwort“ auf die gestellte Frage, z.B. „Frauen verdienen im Schnitt weniger als Männer“.



Typische Unterschiedshypothesen

Merkmal 1 (kategorial bzw. dichotom)

Merkmal 2 (metrisch)



Fragestellung	Gruppe 1	Gruppe 2	Zielvariable
Unterscheidet sich das Einkommen von Männern und Frauen?	Frauen	Männer	z.B. durchschnittliches Einkommen
Zeigen die Schüler:innen in Asien bessere Leistungen als deutsche Schüler:innen?	Schüler:innen in Asien	Schüler:innen in Deutschland	z. B. mittlere Leistung in einem Schulleistungstest
Rasen Autofahrer in München 2024 weniger als im Jahr 2020?	Autofahrer in München im Jahr 2020	Autofahrer in München im Jahr 2024	z. B. gemessene durchschnittliche Geschwindigkeiten z. B. durchschnittliche Anzahl an geblitzten Fahrern



Typische Zusammenhangshypothesen

Merkmal 1
(z.B. metrisch)

Merkmal 2
(z.B. metrisch)

Fragestellung	Merkmal 1	Merkmal 2
Hängt der weltweite CO ₂ -Ausstoß mit der durchschnittlichen Temperatur auf der Erde zusammen?	CO ₂ -Ausstoß	Durchschnittstemperatur
Zeigen die Schüler:innen mit höherer Intelligenz auch bessere Schulleistungen?	Intelligenz der Schüler:innen	Schulleistung der Schüler:innen



Typische Abhängigkeitshypothesen

**Merkmal 1
(kategorial)**

**Merkmal 2
(kategorial)**

Fragestellung	Merkmal 1	Merkmal 2
Hängt die Wahl der gekauften Zeitschrift vom Geschlecht ab („Auto, Motor und Sport“ vs. „Bild der Frau“)?	Geschlecht (Männer vs. Frauen)	Zeitschrift (Auto, Motor und Sport vs. Bild der Frau)
Hängt die Wahl des Verkehrsmittels (ÖPNV vs. Auto) von der Parteizugehörigkeit ab?	Gewähltes Verkehrsmittel (ÖPNV vs. Auto)	Parteizugehörigkeit (CDU/CSU, SPD, Grüne, FDP usw.)



Zusammenhänge als übergeordnete Idee

Zusammenhänge stellen in der Statistik allerdings die übergeordnete Idee dar. Daher lassen sich auch alle Unterschieds- und Abhängigkeitshypothesen als Zusammenhangshypothesen formulieren:

Beispiel 1

Unterschiedshypothese: Frauen verdienen im Schnitt weniger als Männer.



Zusammenhangshypothese: Der Durchschnittsverdienst hängt mit dem Geschlecht zusammen.

Beispiel 2

Abhängigkeitshypothese: Die Wahl des Verkehrsmittels hängt von der Parteizugehörigkeit ab.



Zusammenhangshypothese: Das gewählte Verkehrsmittel hängt mit der Parteizugehörigkeit zusammen.



Warum CODAP? Warum nicht R oder SPSS?

Man könnte nun als Statistik-Software ein Profi-Programm nutzen (im Sinne von Wissenschaftspropädeutik), wie z.B.

R oder Python

- syntaxbasiert ✘
- schlank ✓
- Gefahr, dass der Fokus auf der Fehlersuche in Befehlen liegt und weniger über Statistik gesprochen wird. ✘

z.B. SPSS

- mit Buttons 😊 ✓
- Erschlagend viele Möglichkeiten ✘
- Kostenpflichtig ✘
- Gefahr, dass man nur Klick-Pfade auswendig lernt und auch auswendig lernt: „Von den drei angezeigten p-Werte konzentriere ich mich immer auf den rechts unten und ignoriere die anderen beiden – keine Ahnung warum.“ ✘



Warum nicht Fathom oder GeoGebra?

Der Vorteil von Software, die eigens für die Schule entwickelt wurde (Fathom, GeoGebra oder CODAP, ist, dass diese schlanker ist, intuitiver, besser auf Repräsentationswechsel ausgerichtet ist etc.:

Fathom
<ul style="list-style-type: none"> • Didaktisch gut aufbereitet ✓ • eignet sich auch für Sek I ✓ • kostenlos ✓ • nicht für Mac geeignet ✗ • Wird nicht mehr gepflegt und hat somit leider ein Verfallsdatum ✗

GeoGebra
<ul style="list-style-type: none"> • Didaktisch gut aufbereitet ✓ • eignet sich auch für Sek I ✓ • kostenlos ✓ • Der große Nachteil: Bricht bei größeren Datensätzen zusammen ✗



CODAP

- Didaktisch gut aufbereitet ✓
- eignet sich auch für Sek I ✓
- kostenlos ✓
- Webbasiert (!) und auch für MAC geeignet ✓
- Der große Nachteil war bis vor Kurzem: Hat überhaupt keine Tests, t-Test, Chi-Quadrat-Test usw. ✗

Jetzt Dank Tim Erickson möglich! 😊



Zur Benutzung von CODAP für das Vertiefungsmodul finden Sie online in den Materialien auch eine Anleitung!



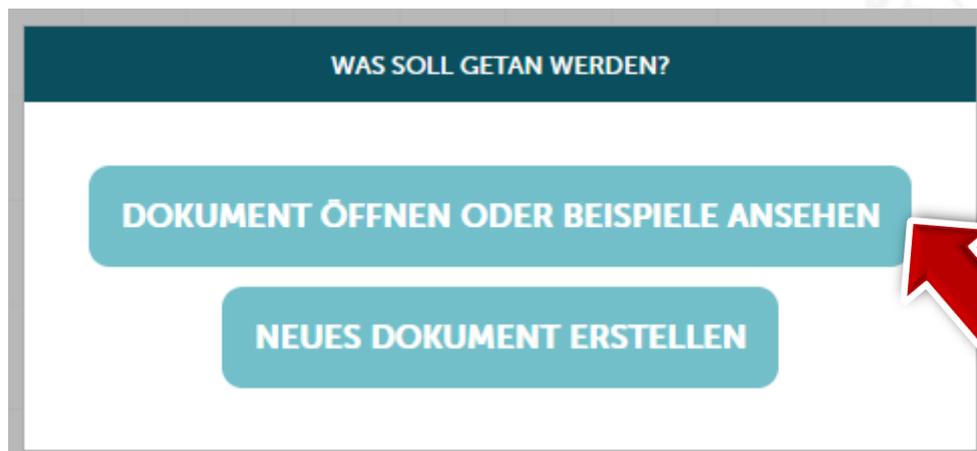
CODAP

Öffnen von CODAP:

<https://codap.concord.org/app/static/dg/de/cert/index.html>

In CODAP gibt es bereits fertige Datensätze, mit denen man arbeiten kann.

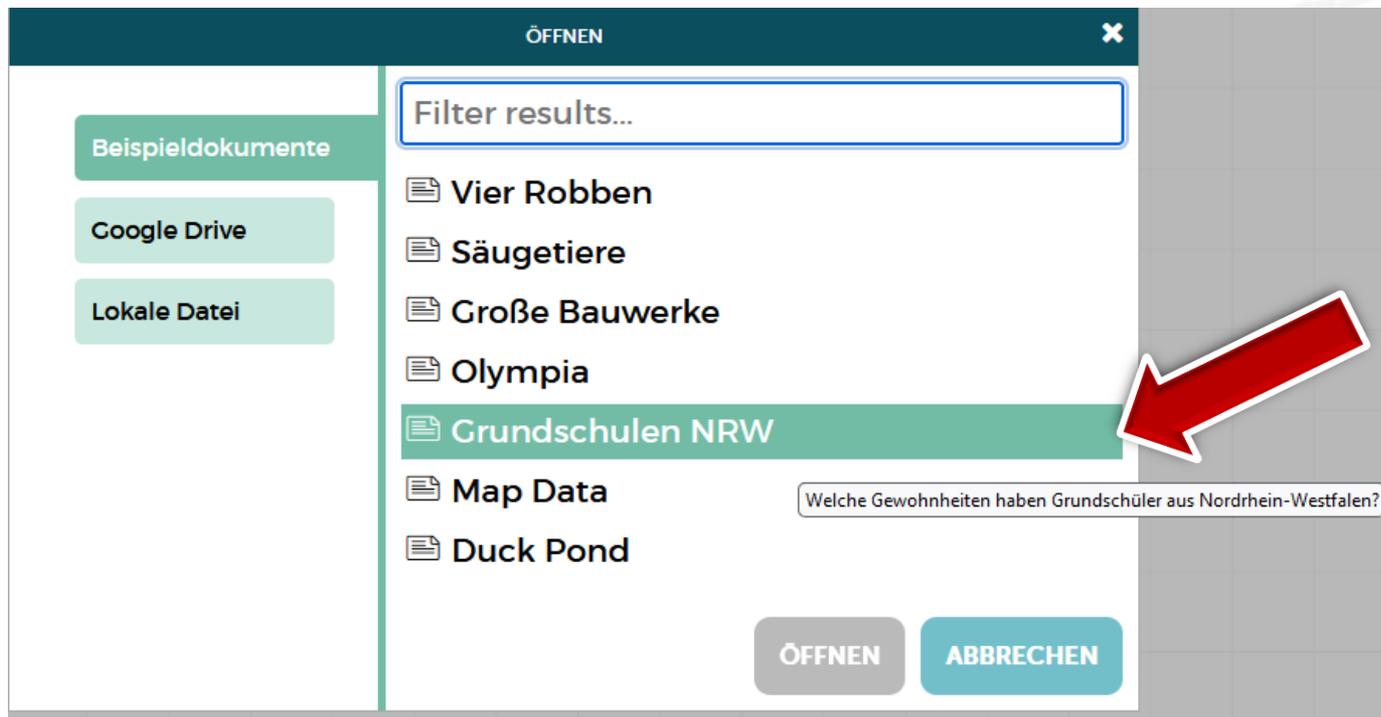
Klicken Sie hierzu auf „Dokument öffnen oder Beispiele ansehen“:





CODAP

Wählen Sie nun z.B. den Datensatz „Grundschulen NRW“ und klicken Sie dann auf „öffnen“:





CODAP

Schon öffnet sich ein umfangreicher Datensatz, an den viele Forschungsfragen gestellt und untersucht werden können.

Grundschulen NRW

Tabellen Graph Karte Regler Rechner Text Plugins

Grundschulen_NRW

Fälle (809 Fälle)

In- dex	Fantasiename	Ort	Großstadt Stadt Dorf	Breiten- grad	Längen- grad	Ge- schlecht	Alter	Klasse	Körper- _Be (cm)	Schuh- größe	Anzahl _amilie	Haus- tier	FallsHaustier Welches	FallsHaus _ proTag	FallsHau _Jahren	Minu- _nuten)	Von Eltern gebracht	Wie z Schul	
1	Bad Girl	Kleve	Stadt	51.79	6.14	Mädchen	8	3		38	2	nein				5	ja	zu Fu	
2	Minecraft	Dortmund	Großstadt	51.57	7.52	Junge	9	4	147	33	1	ja	Wellensittich	1	2	5	ja	Auto	
3	Lukas	Kevelaerl	Stadt	51.59	6.24	Junge	9	4		33	2	ja	sonstiges	1		15	nein	Fahrr	
4	Herry Potter	Kevelaer	Stadt	51.57	6.29	Junge	8	3	132	34	2	nein				2	ja	Auto	
5		Geldern	Stadt	51.52	6.32	Junge	8	3	133	30	4	ja	Hund		13	10	nein	zu Fu	
6	Kaitlen	Walbeck	Dorf	51.5	6.24	Mädchen	8	3	136	33	3	nein					ja	Auto	
7	Fenja	Rheurdt	Dorf	51.47	6.47	Mädchen	8	3		35	mehr ...	nein					6	nein	zu Fu
8	Mia	Geldernl	Stadt	51.52	6.32	Mädchen	8	3		36	2	ja	Hamster	12			15	ja	Auto
9	Terence	Willich	Stadt	51.24	6.54	Junge	8	3	145	41	2	nein					3	ja	Auto
10	Max	Dortmun...	Großstadt	51.53	7.55	Junge	8	3		33	3	ja	Hund	12	3		5	nein	Fahrr
11	Nika	Ciesenkir...	Dorf	51.58	6.49	Mädchen	8	3		34	2	ja	Hund	12	1	20	nein	zu Fu	
12	Alexandra	Rheydt	Dorf	51.52	6.42	Mädchen	8	3	135	35	1	nein					2	ja	Auto
13	Anne	Betrath	Dorf	51.23	6.47	Junge	9	4	131	33	3	nein					4	ja	Auto
14	Hanna	Rheinda...	Dorf	51.18	6.41	Mädchen	8	3	150	38	2	ja	sonstiges		4		ja	Auto	
15	Mike	Holt	Dorf	51.18	6.41	Junge	10	4		36	mehr ...	ja	Meerschwei...	1	3		ja	Auto	

Textfeld

Grundschüler in Nordrhein-Westfalen

In diesem Datensatz befinden sich Daten von Grundschulern aus verschiedenen Grundschulen in Nordrhein-Westfalen, die 2017 einen Fragebogen ausgefüllt haben. Die Teilnahme war freiwillig und ist somit nicht repräsentativ für eine Grundgesamtheit. Trotzdem lassen sich einige interessante Beobachtungen machen.

Zum Start:

- Wie viele Jungen und wie viele Mädchen haben an der Umfrage teilgenommen? Sind es mehr Dritt- oder Viertklässler?
- Ein Kind hat angegeben, 3 Jahre alt zu sein. Kann das sein? Dieses Kind kann über das Augensymbol im Graph ausgeblendet

Fälle



CODAP

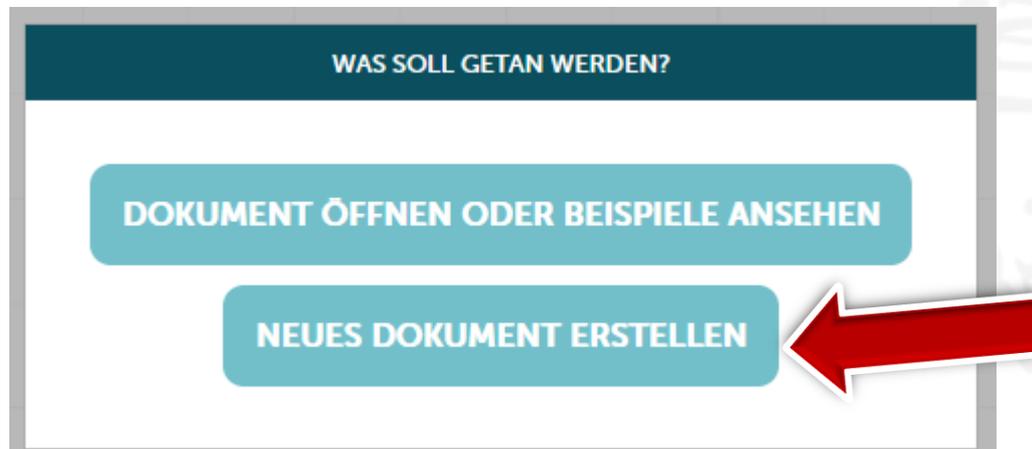
Es lassen sich auch Datensätze importieren.

Vorsicht: Eine *.xls kann nicht direkt in CODAP geöffnet werden. Diese bitte vorher einfach als *.**csv-Datei** abspeichern. So kann man auch Excel-Datensätze einfach in CODAP öffnen.

Wir öffnen hierzu wieder CODAP:

<https://codap.concord.org/app/static/dg/de/cert/index.html>

Nun bitte klicken auf „Neues Dokument erstellen“





CODAP

Ziehen Sie (gedrückte linke Maustaste) den gewünschten Datensatz **Mikrozensus.csv** in CODAP (bitte nur 500 Fälle einfügen):

Tabellen Graph Karte Regler Rechner Text Plugins

Mikrozensus_2010

cases (500 Fälle)

In-dex	Ost-...schland	Erhe-...trument	Zahl der ...aushalt	Er-...bstaetig	Alter	Ge-schlecht	Geburts-jahr	Fa-st
496	Ost	Selbsta...	3	Erwerbs...	29	männlic...	1980	ve
497	Ost	Laptopi...	3	Sonstig...	70	männlic...	1940	ve
498	Ost	Laptopi...	2	Erwerbs...	32	weiblich	1978	le
499	Ost	Selbsta...	4	Sonstig...	1	weiblich	2009	le
500	Ost	Laptopi...	3	Erwerbs...	47	weiblich	1963	ve



CODAP

Der Datensatz ist sehr groß. Wir wollen erst mal nicht alle Zeilen importieren, sondern nur 500 Fälle (Zeilen) zufällig wählen.

A screenshot of the CODAP Importer dialog box. The title bar shows 'v1.0 Importer'. The main text reads: 'Die importierte Datei "Mikrozensus_2010.csv" hat 23.375 Zeilen. Bei mehr als 5.000 Zeilen wird CODAP stark verlangsamt. Eine Stichprobe des Datensatzes kann besser funktionieren, später kann der gesamte Datensatz genutzt werden. Einstellungen zu Stichprobe verkleinern'. There are three radio button options: 'Zufällig auswählen' (selected), 'Jede auswählen', and 'Alle Zeilen importieren'. The 'Zufällig auswählen' option has a dropdown menu set to '500' and the label 'Zeilen'. The 'Jede auswählen' option has a dropdown menu set to '6' and the label 'te Zeile'. At the bottom right, there are two buttons: 'ABBRECHEN' and 'OK'.



CODAP

Der Datensatz ist nun geöffnet. Wir können die Daten nun genauer analysieren. Im pdf „mz_2010_cf_svz“ (**Schlüsselverzeichnis bzw. Skaldokumentation des Mikrozensus**) finden Sie Informationen darüber, was die einzelnen Merkmalsausprägungen bedeuten.

Beispiel: Was bedeutet die Merkmalsausprägung „15“ oder die Merkmalsausprägung „99“ beim „Haushaltseinkommen“?

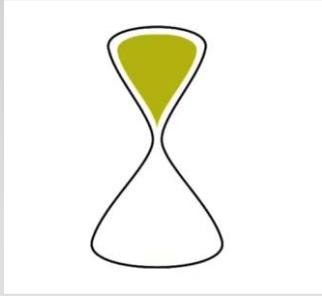
707	Haushaltsnettoeinkommen (nach Korrektur)		
	unter	150 Euro	01
	150 bis unter	300 Euro	02
	300 bis unter	500 Euro	03
	500 bis unter	700 Euro	04
	700 bis unter	900 Euro	05
	900 bis unter	1 100 Euro	06
	1 100 bis unter	1 300 Euro	07
	1 300 bis unter	1 500 Euro	08
	1 500 bis unter	1 700 Euro	09



CODAP

707	Haushaltsnettoeinkommen (nach Korrektur)	
	unter 150 Euro	01
	150 bis unter 300 Euro	02
	300 bis unter 500 Euro	03
	500 bis unter 700 Euro	04
	700 bis unter 900 Euro	05
	900 bis unter 1 100 Euro	06
	1 100 bis unter 1 300 Euro	07
	1 300 bis unter 1 500 Euro	08
	1 500 bis unter 1 700 Euro	09
	1 700 bis unter 2 000 Euro	10
	2 000 bis unter 2 300 Euro	11
	2 300 bis unter 2 600 Euro	12
	2 600 bis unter 2 900 Euro	13
	2 900 bis unter 3 200 Euro	14
	3 200 bis unter 3 600 Euro	15
	3 600 bis unter 4 000 Euro	16
	4 000 bis unter 4 500 Euro	17
	4 500 bis unter 5 000 Euro	18
	5 000 bis unter 5 500 Euro	19
	5 500 bis unter 6 000 Euro	20
	6 000 bis unter 7 500 Euro	21
	7 500 bis unter 10 000 Euro	22
	10 000 bis unter 18 000 Euro	23
	18 000 und mehr Euro	24
	Mindestens ein Haushaltsmitglied ist selbständiger Landwirt in der Haupttätigkeit	50





Workshop-Phase

Bearbeiten Sie bitte
Übungsblatt 2

Zeit: 5 Minuten





Mögliche Lösung:

	Forschungsfrage	Hypothese
Unterschieds- hypothese	Unterscheidet sich die geleistete Arbeitszeit zwischen Ost und West?	Die durchschnittliche Arbeitszeit pro Woche unterscheidet sich im Osten und Westen.
Zusammenhangs- hypothese	Hängen Miete und Gehalt zusammen?	Das Gehalt und die Miete hängen zusammen.
Unabhängigkeits- hypothese	Hängt es vom Geschlecht ab, ob man eher im Osten oder Westen wohnt?	Ja, es hängt vom Geschlecht ab, ob man eher im Osten oder Westen wohnt!



Mögliche Lösung:

<p>Unterschiedshypothese als Zusammenhangshypothese:</p>	<p>Die durchschnittliche wöchentliche Arbeitszeit hängt mit dem Wohnort (Ost vs. West) zusammen.</p>
<p>Unabhängigkeitshypothese als Zusammenhangshypothese:</p>	<p>Es gibt einen Zusammenhang zwischen Wohnort (Ost vs. West) und dem Geschlecht.</p>

PAUSE ☺



Die grundlegende Idee von Signifikanztests



Die Ausgangssituation

Wir möchten beispielsweise wissen:

„Wirkt das neu entwickelte Medikament wirklich besser als ein Placebo?“



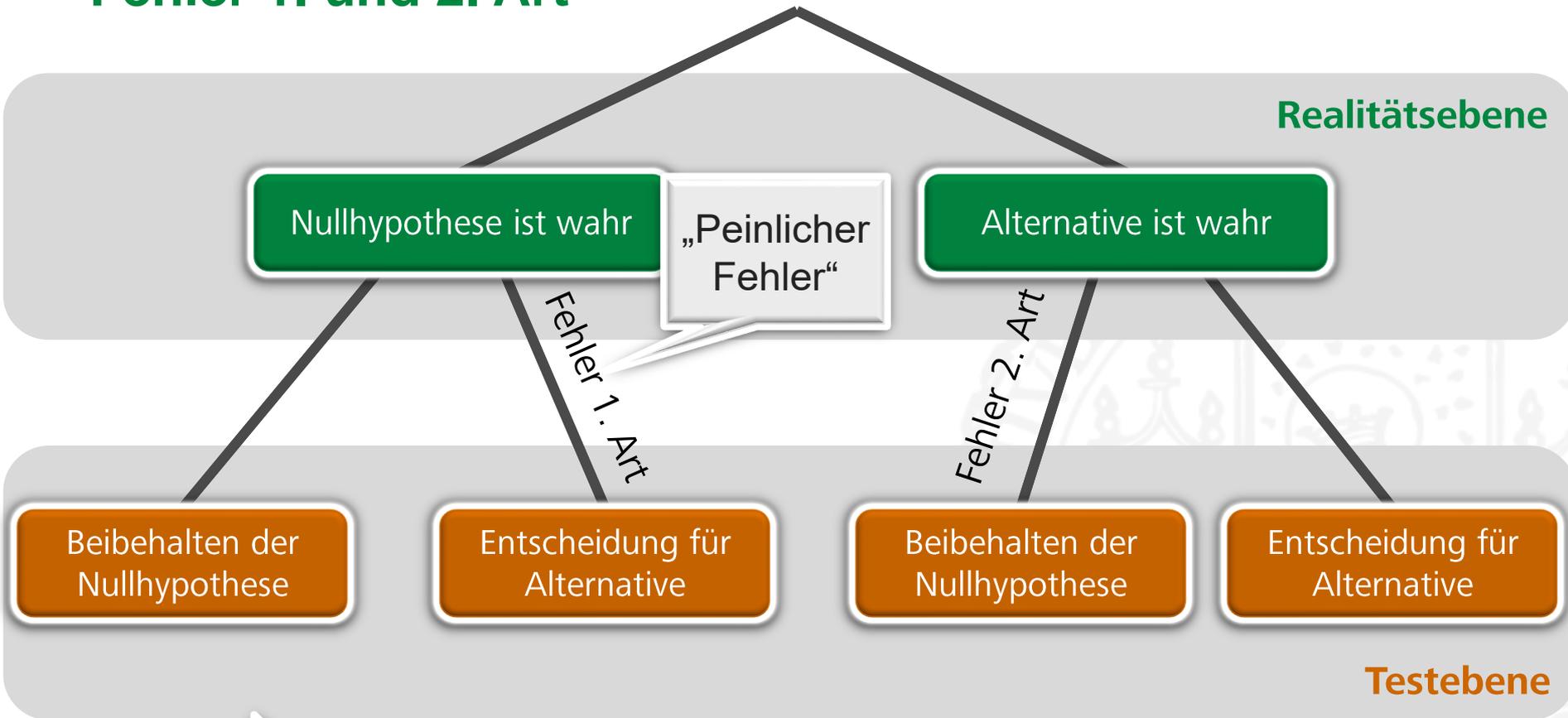
Die grundlegende Idee von Signifikanztests

Stichprobe vs. Grundgesamtheit

Grundgesamtheit

**Stich-
probe**

Fehler 1. und 2. Art



Erst wenn sehr viel gegen die Nullhypothese spricht, wird die Alternative gewählt

Fehler 1. und 2. Art



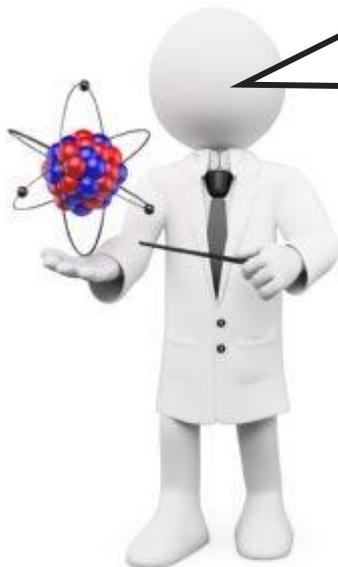
**Type I Fehler
(falsch positiv)**



**Type II Fehler
(falsch negativ)**

Die grundlegende Idee von Signifikanztests

Welcher Fehler ist der „peinliche Fehler“ bei wissenschaftlichen Fragestellungen?



Wir haben einen von Menschen verursachten Klimawandel und müssen sofort handeln! Es gibt einen signifikanten Zusammenhang zwischen dem CO₂-Ausstoß und der Durchschnittstemperatur.

... 2 Wochen später ...

...ups doch nicht... Sorry...

Die grundlegende Idee von Signifikanztests



Typische Nullhypothesen in der Forschung

t-Test
Es gibt <u>keinen</u> Unterschied ...

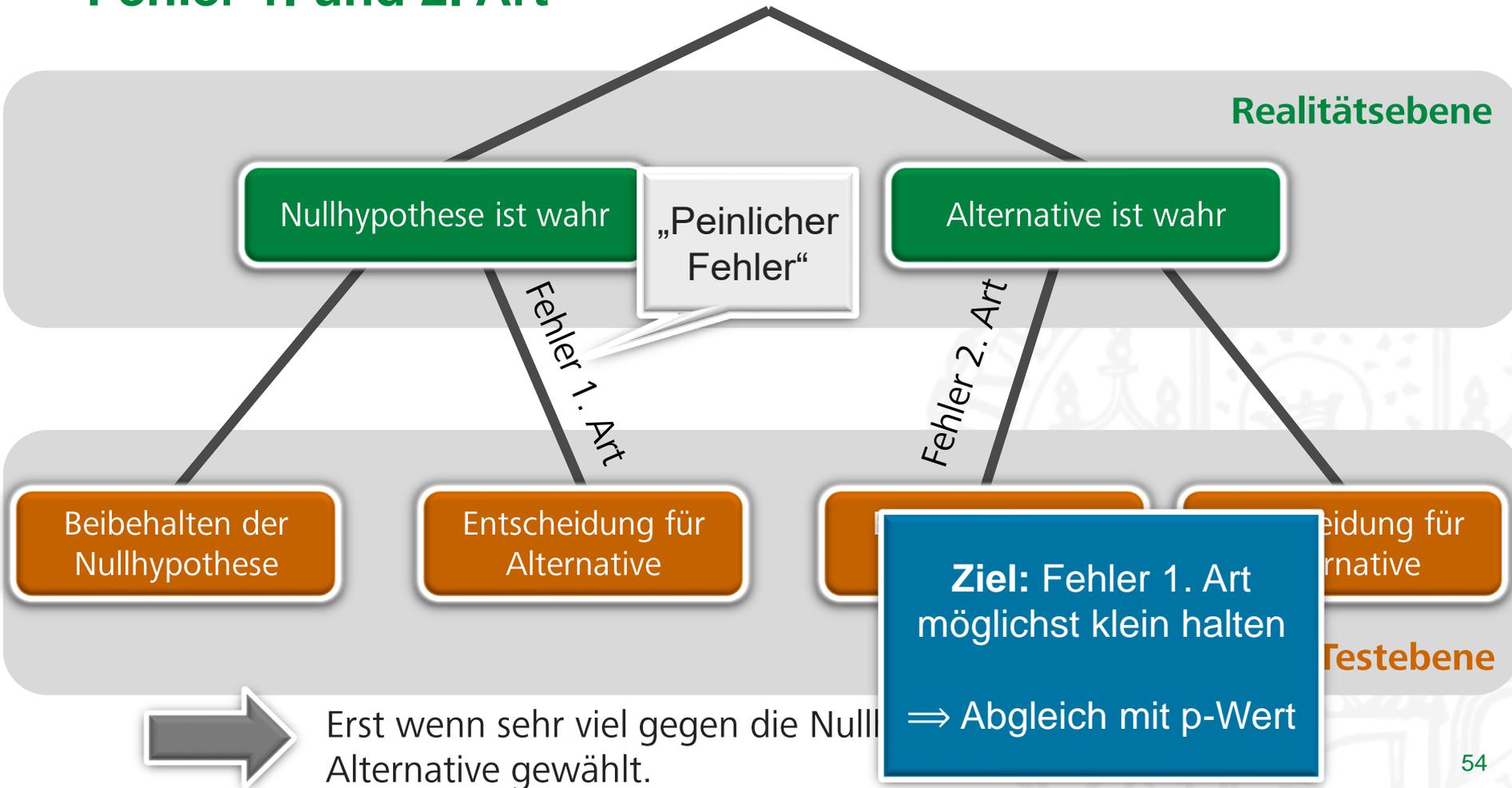
Korrelation
Es gibt <u>keinen</u> Zusammenhang ...

χ^2 -Test
Es gibt <u>keine</u> Abhängigkeit ...

Wir stellen uns also immer eine Welt vor, in der es **KEINEN Unterschied (bzw. Zusammenhang oder Abhängigkeit)** gibt.

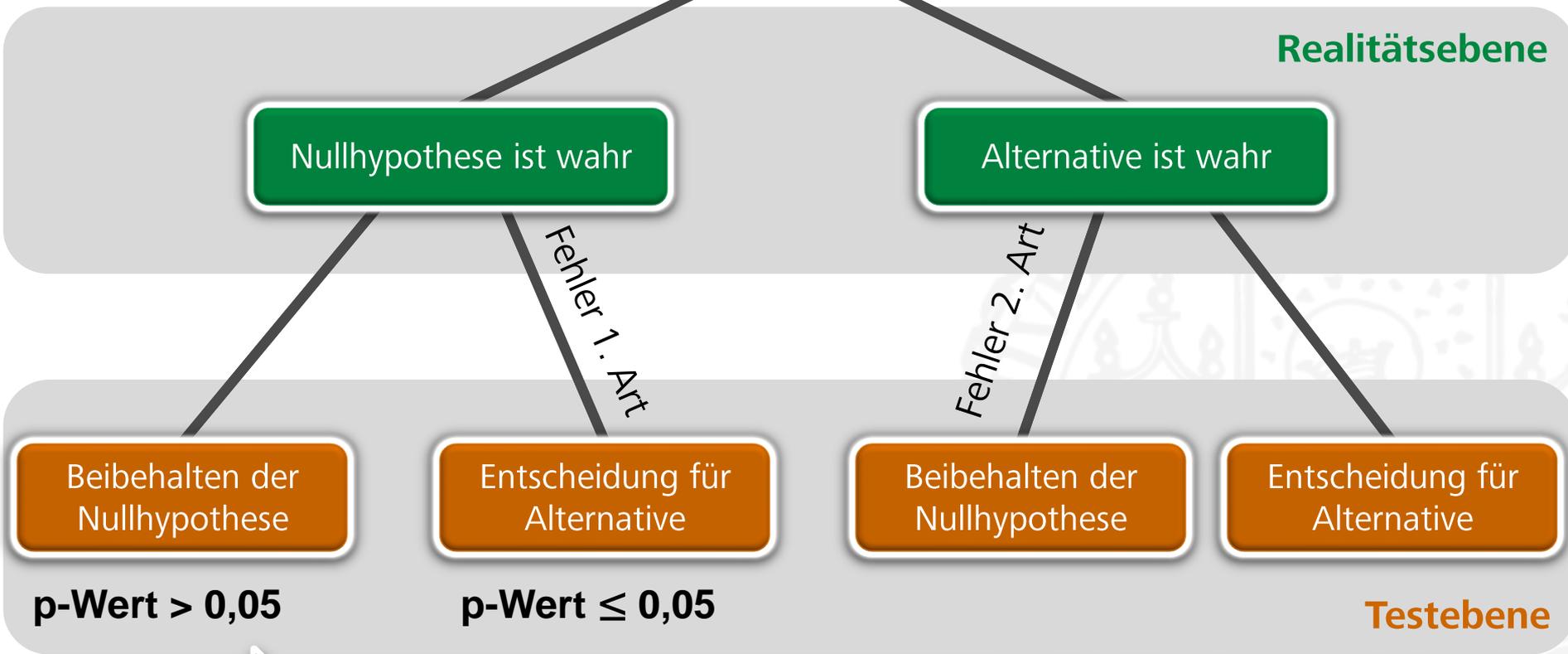
Erst wenn sehr viel dagegen spricht und es sehr unwahrscheinlich ist, solch extreme Unterschiede (Zusammenhänge, oder Abhängigkeiten) zu finden, obwohl es die in der Realität gar nicht gibt, lehnen wir die Nullhypothese ab.

Fehler 1. und 2. Art



Ziel: Fehler 1. Art möglichst klein halten

Auf zum p-Wert



Erst wenn sehr viel gegen die Nullhypothese spricht, wird die Alternative gewählt.

Die grundlegende Idee von Signifikanztests



Das Signifikanzniveau

Da man den Fehler 1. Art möglichst klein halten möchte, begrenzt man diesen Fehler, indem man zuvor ein sogenanntes „Signifikanzniveau“ festlegt.

Häufig wird hierbei ein Signifikanzniveau von 5%, 1% oder 0,1% gewählt.

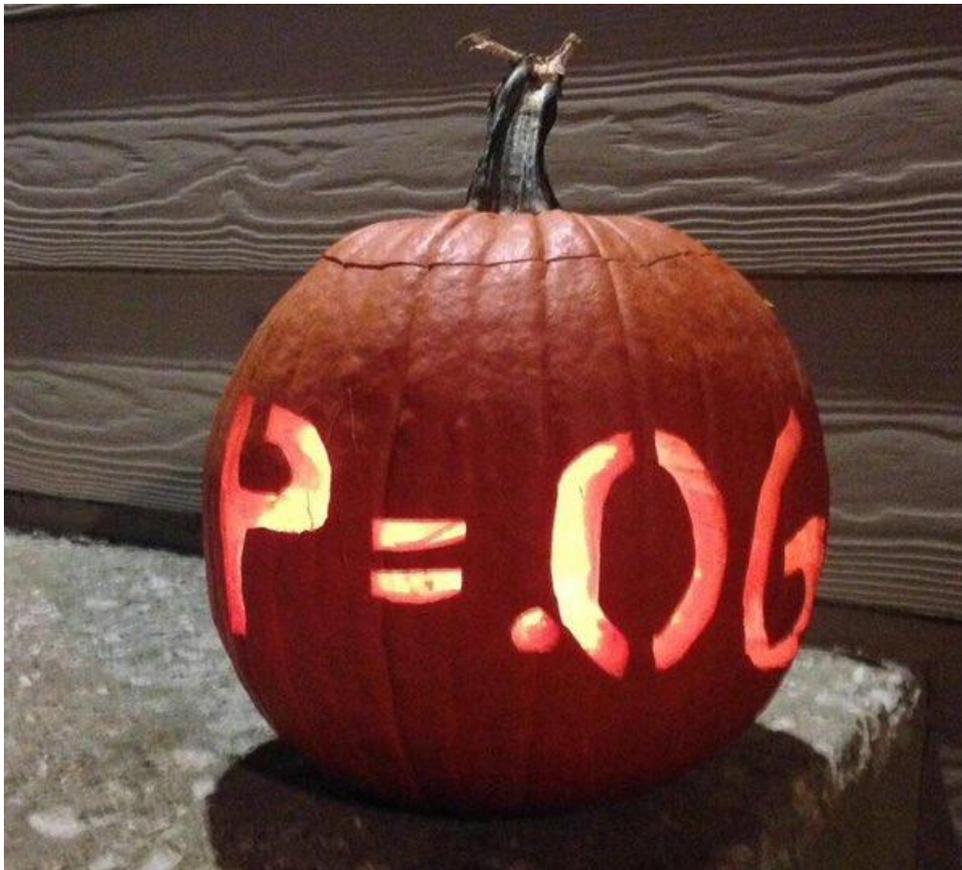
Nun bleibt nur die Frage: Ist der p-Wert kleiner oder größer als diese 5%.

Ist der p-Wert kleiner (also z.B. kleiner als 0,05), lehnt man die Nullhypothese ab.

Der **p-Wert** gibt die Wahrscheinlichkeit an, diese oder noch extremere Daten (z.B. beim t-Test mit einem noch größeren Gruppenunterschied) zu finden, obwohl sich die beiden Gruppen in Wirklichkeit nicht unterscheiden.

Die grundlegende Idee von Signifikanztests

Das „Grauen“ für jeden empirischen Wissenschaftler 😊



Bemerkung: Das stimmt natürlich nicht. Als Wissenschaftler ist man, um echte „Erkenntnis“ bemüht und jagt keine signifikanten p-Werte.

Der χ^2 -Test



Von stochastischer Unabhängigkeit zum χ^2 -Test

15

Ein neues Medikament wird auf seine Wirksamkeit erprobt. Hierzu haben sich Patienten bereit erklärt, an einer Studie teilzunehmen. Ein Teil der Patienten bekommt das neue Medikament, der andere ein Placebo. Kein Patient weiß jedoch, ob er ein Placebo oder das Medikament bekommen hat. In der Tabelle sind die Ergebnisse der Studie zusammengetragen.

	Besserung	Keine Besserung
Medikament	285	150
Placebo	608	320

Betrachtet man das Zufallexperiment „Zufällige Auswahl eines Patienten, der teilgenommen hat“, so werden durch „Besserung“ und „Medikament“ Ereignisse festgelegt. Untersuchen Sie die beiden Ereignisse auf Unabhängigkeit.

	Besserung	Keine Besserung	
Medikament	285 286	150	435
Placebo	608	320	928
	893	470	1363

$$\frac{285}{435} = \frac{19}{29}$$

$$\frac{608}{928} = \frac{19}{29}$$

Die beiden Ereignisse sind stochastisch unabhängig.

Der χ^2 -Test - Auf der Suche nach einer Teststatistik

Reale Daten

	Besserung	Keine Besserung	
Medikament	302	133	435
Placebo	591	337	928
	893	470	1363

Stellen wir uns vor, wir haben diesen Datensatz bei unserer Erhebung erhalten.

Stoch. Unab.

	Besserung	Keine Besserung	
Medikament	285	150	435
Placebo	608	320	928
	893	470	1363

Wie könnte nun ein „Maß“ aussehen, das berechnet, ob der Unterschied zwischen den beiden Tabellen groß oder klein ist?

Ideen?



Der χ^2 -Test - Auf der Suche nach einer Teststatistik

Reale Daten

	Besserung	Keine Besserung	
Medikament	302	133	435
Placebo	591	337	928
	893	470	1363

Welches der beiden folgenden Maße sollten wir bevorzugen?

Maß 1:

$$(302 - 285) + (133 - 150) + (591 - 608) + (337 - 320)$$

Maß 2:

$$(302 - 285)^2 + (133 - 150)^2 + (591 - 608)^2 + (337 - 320)^2$$

Stoch. Unab.

	Besserung	Keine Besserung	
Medikament	285	150	435
Placebo	608	320	928
	893	470	1363

Der χ^2 -Test - Auf der Suche nach einer Teststatistik

Reale Daten

	Besserung	Keine Besserung	
Medikament	302	133	435
Placebo	591	337	928
	893	470	1363

Welches der beiden folgenden Maße sollten wir bevorzugen?

Maß 2:

$$(302 - 285)^2 + (133 - 150)^2 + (591 - 608)^2 + (337 - 320)^2$$

Stoch. Unab.

	Besserung	Keine Besserung	
Medikament	285	150	435
Placebo	608	320	928
	893	470	1363

Maß 3:

$$\frac{(302 - 285)^2}{285} + \frac{(133 - 150)^2}{150} + \frac{(591 - 608)^2}{608} + \frac{(337 - 320)^2}{320}$$

Der χ^2 -Test - Auf der Suche nach einer Teststatistik

Reale Daten

	Besserung	Keine Besserung	
Medikament	302	133	435
Placebo	591	337	928
	893	470	1363

Stoch. Unab.

	Besserung	Keine Besserung	
Medikament	285	150	435
Placebo	608	320	928
	893	470	1363

Teststatistik:

$$\frac{(302 - 285)^2}{285} + \frac{(133 - 150)^2}{150} + \frac{(591 - 608)^2}{608} + \frac{(337 - 320)^2}{320}$$

= 4,32

Fertig! ✓

Der χ^2 -Test - Das Grundprinzip

Gleiche
Randhäufigkeiten

Reale Daten

	Besserung	Keine Besserung	
Medikament	302	133	435
Placebo	591	337	928
	893	470	1363

Stoch. Unab.

	Besserung	Keine Besserung	
Medikament	285	150	435
Placebo	608	320	928
	893	470	1363

Bei einem χ -Quadrat-
Unabhängigkeitstest
werden die Daten zweier
Vierfeldertafeln (oder gerne
auch Mehrfeldertafeln!)
miteinander verglichen:

**Reale Daten vs.
stochastisch
unabhängige Daten**
(bei identischen
Randhäufigkeiten)



CODAP

Ziehen Sie nun (gedrückte linke Maustaste) den gewünschten Datensatz **Placebo-Studie.csv** in CODAP:

The screenshot shows the CODAP application interface. At the top, there is a menu bar with 'Dokument1' and a toolbar with icons for 'Tabellen', 'Graph', 'Karte', 'Regler', 'Rechner', 'Text', and 'Plugins'. The main workspace is a grid with a spreadsheet icon and a '+ Kopieren' button. A file explorer window is open, showing the 'Datensätze' folder. The file list includes:

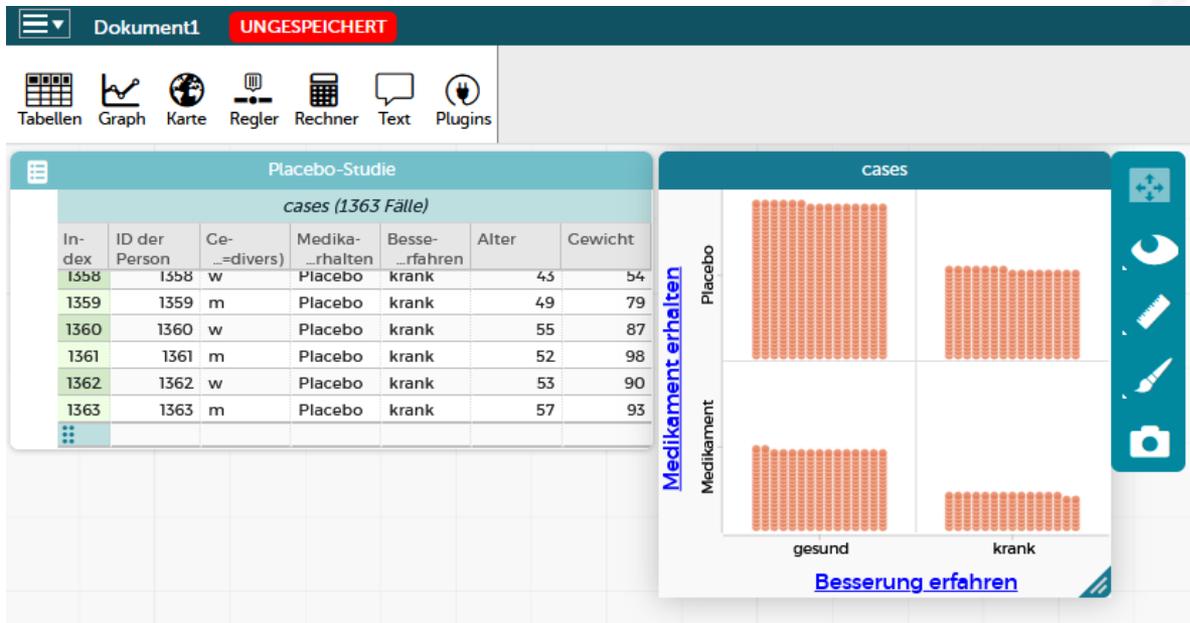
Name	Änderu
Mikrozensus_2010_200Faelle	30.04.20
mz_2010_cf_svz	02.02.20
Placebo-Studie	20.06.20
Placebo-Studie	25.09.20
Placebo-Studie_stochastisch_unabhängig	25.09.20
Simulation_weiß_blaue	26.09.20

The status bar at the bottom of the file explorer indicates '9 Elemente' and '1 Element ausgewählt (38,4 KB)'.



CODAP

Wir können uns mit dem Datensatz vertraut machen und rein deskriptiv untersuchen, ob es einen **Zusammenhang zwischen den Merkmalen „Medikament erhalten“ und „Besserung erfahren“** gibt. Klicken Sie hierzu auf Graph und ziehen Sie die beiden Merkmale auf die beiden Achsen:





CODAP

Durch einen Klick auf das Lineal (Messung) können wir uns auch anzeigen lassen, wie hoch bei den Personen mit Medikament oder mit Placebo jeweils der Anteil der geheilten Personen ist (hierzu bei dieser Anordnung: **Prozent „Zeile“!**)

Dokument1 **UNGESPEICHERT**

Tabellen Graph Karte Regler Rechner Text Plugins

Placebo-Studie

cases (1363 Fälle)

In- dex	ID der Person	Ce- ...=divers)	Medika- ...rhalten	Besse- ...rfahren	Alter	Gewicht
1358	1358	w	Placebo	krank	43	54
1359	1359	m	Placebo	krank	49	79
1360	1360	w	Placebo	krank	55	87
1361	1361	m	Placebo	krank	52	98
1362	1362	w	Placebo	krank	53	90
1363	1363	m	Placebo	krank	57	93

cases

Medikament erhalten

Medikament

gesund krank

Besserung erfahren

Messung

Zeigen...

Anzahl

Prozent

Zeile

Spalte

Zelle



CODAP

Wir führen nun einen Chi-Quadrat-Test durch, um inferenzstatistisch zu prüfen, ob die beiden Merkmale voneinander abhängig sind. Gehen Sie hierbei bei **Plugins** auf das PlugIn **testimate**

Placebo-Studie

cases (1363 Fälle)				
In- dex	ID der Person	Ce- ...=divers)	Medika- ...rhalten	Besse- ...rfahren
1558	1558	w	Placebo	krank
1359	1359	m	Placebo	krank
1360	1360	w	Placebo	krank
1361	1361	m	Placebo	krank
1362	1362	w	Placebo	krank
1363	1363	m	Placebo	krank

cases

Medikament	gesund	krank
cases	64%	36%
Besserung erfahren	69%	31%

Test hypotheses using classical inferential methods.

[Besserung erfahren](#)



CODAP

Ziehen Sie nun die beiden Merkmale „Medikament erhalten“ und „Besserung erfahren“ im neuen Fenster „testimate“ auf „Merkmal hier ablegen“ (die Reihenfolge ist bei Unabhängigkeitstest nicht von Bedeutung!).

Wählen Sie nun bei der **Art des Tests** „Test auf Unabhängigkeit...“

2024i testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig) i

Medikament erhalten

Zielvariable/erstes Merkmal abc 🗑️

Besserung erfahren

Prädiktor/zweites Merkmal abc 🗑️

Test auf Unabhängigkeit von Besserung erfahren und Medikament erhalten

Sind (Besserung erfahren) und (Medikament erhalten) unabhängig?
 $N = 1363$, 2 columns by 2 rows, $\chi^2 = 4,319$, $P = 0,03769$

▶ Test auf Unabhängigkeit, 1-seitig χ^2 Zusammenfassung





CODAP

Durch einen Klick auf den kleinen Pfeil links sehen Sie nun die **beiden Vierfeldertafeln ineinander geschachtelt**: Die Werte, wie sie in der Stichprobe vorliegen und die Werte, wie sie bei stochastischer Unabhängigkeit vorliegen würden:



2024i testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig)

Medikament erhalten
Zielvariable/erstes Merkmal abc

Besserung erfahren
Prädiktor/zweites Merkmal abc

Test auf Unabhängigkeit von Besserung erfahren und Medikament erhalten

Sind (Besserung erfahren) und (Medikament erhalten) unabhängig?
N = 1363, 2 columns by 2 rows, $\chi^2 = 4,319$, P = 0,03769

▼ Test auf Unabhängigkeit, 1-seitig χ^2 Zusammenfassung

beobachtet erwartet	Besserung erfahren =	gesund	krank
Medikament erhalten =	Medikament	302 285	133 150
	Placebo	591 608	337 320

df = 1, $\alpha = 0,05$, $\chi^{2*} = 3,841$

konfiguriere Test auf Unabhängigkeit von (Besserung erfahren) und (Medikament erhalten):
 $\alpha = 0.05$ 1-seitig

▼ Testergebnisse an CODAP übermitteln
 vorliegender Test



CODAP

2024i testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig)

Medikament erhalten
Zielvariable/erstes Merkmal abc

Besserung erfahren
Prädiktor/zweites Merkmal abc

Test auf Unabhängigkeit von Besserung erfahren und Medikament erhalten

Sind (Besserung erfahren) und (Medikament erhalten) unabhängig?
N = 1363, 2 columns by 2 rows, $\chi^2 = 4,319$, P = 0,03769

▼ Test auf Unabhängigkeit, 1-seitig χ^2 Zusammenfassung

beobachtet erwartet	Besserung erfahren =	gesund	krank
Medikament erhalten =	Medikament	302 285	133 150
	Placebo	591 608	337 320

df = 1, $\alpha = 0,05$, $\chi^2_{*} = 3,841$

konfiguriere Test auf Unabhängigkeit von (Besserung erfahren) und (Medikament erhalten):
 $\alpha = 0,05$ 1-seitig

▼ Testergebnisse an CODAP übermitteln
 vorliegender Test

Da der p-Wert kleiner als 0,05 ist, hängen die Merkmale signifikant voneinander ab.



Der χ^2 -Test mit CODAP

2024i testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig)

Medikament erhalten
Zielvariable/erstes Merkmal abc

Besserung erfahren
Prädiktor/zweites Merkmal abc

Test auf Unabhängigkeit von Besserung erfahren und Medikament erhalten

Sind (Besserung erfahren) und (Medikament erhalten) unabhängig?
N = 1363, 2 columns by 2 rows, $\chi^2 = 4,319$, P = 0,03769

▼ Test auf Unabhängigkeit, 1-seitig χ^2 Zusammenfassung

beobachtet erwartet	Besserung erfahren =	gesund	krank
Medikament erhalten =	Medikament	302 285	133 150
	Placebo	591 608	337 320

df = 1, $\alpha = 0,05$, $\chi^{2*} = 3,841$

konfiguriere Test auf Unabhängigkeit von (Besserung erfahren) und (Medikament erhalten):
 $\alpha = 0,05$ 1-seitig

▼ Testergebnisse an CODAP übermitteln
 vorliegender Test

$$\chi^2 = \frac{(302 - 285)^2}{285} + \frac{(133 - 150)^2}{150} + \frac{(285 - 591)^2}{591} + \frac{(337 - 320)^2}{320} = 4,32$$

Ist diese Prüfgröße größer als ein bestimmter **Schwellwert**, wird die Nullhypothese ablehnt.

Die Abweichung von der Unabhängigkeit ist signifikant.



Der χ^2 -Test mit CODAP

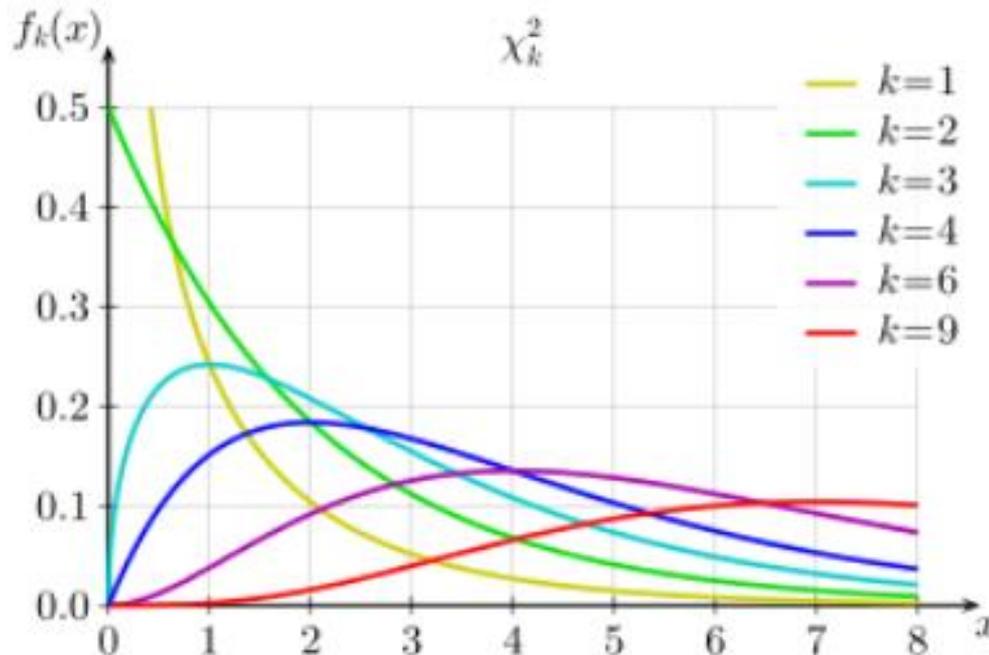
Den Schwellwert kann man über Simulationen erhalten oder in Tabellenbüchern nachschlagen. Hierzu benötigt man die Anzahl der Freiheitsgrade [$df=(n-1)(m-1)$] und das Signifikanzniveau α :

	1- α					
<i>f</i>	0,900	0,950	0,975	0,990	0,995	0,999
1	2,71	3,84	5,02	6,63	7,88	10,83
2	4,61	5,99	7,38	9,21	10,60	13,82
3	6,25	7,81	9,35	11,34	12,84	16,27
4	7,78	9,49	11,14	13,28	14,86	18,47
5	9,24	11,07	12,83	15,09	16,75	20,52
6	10,64	12,59	14,45	16,81	18,55	22,46
7	12,02	14,07	16,01	18,48	20,28	24,32
8	13,36	15,51	17,53	20,09	21,95	26,12
9	14,68	16,92	19,02	21,67	23,59	27,88



Der χ^2 -Test mit CODAP

Den Schwellwert kann man über Simulationen erhalten oder in Tabellenbüchern nachschlagen. Hierzu benötigt man die Anzahl der Freiheitsgrade k und das Signifikanzniveau:





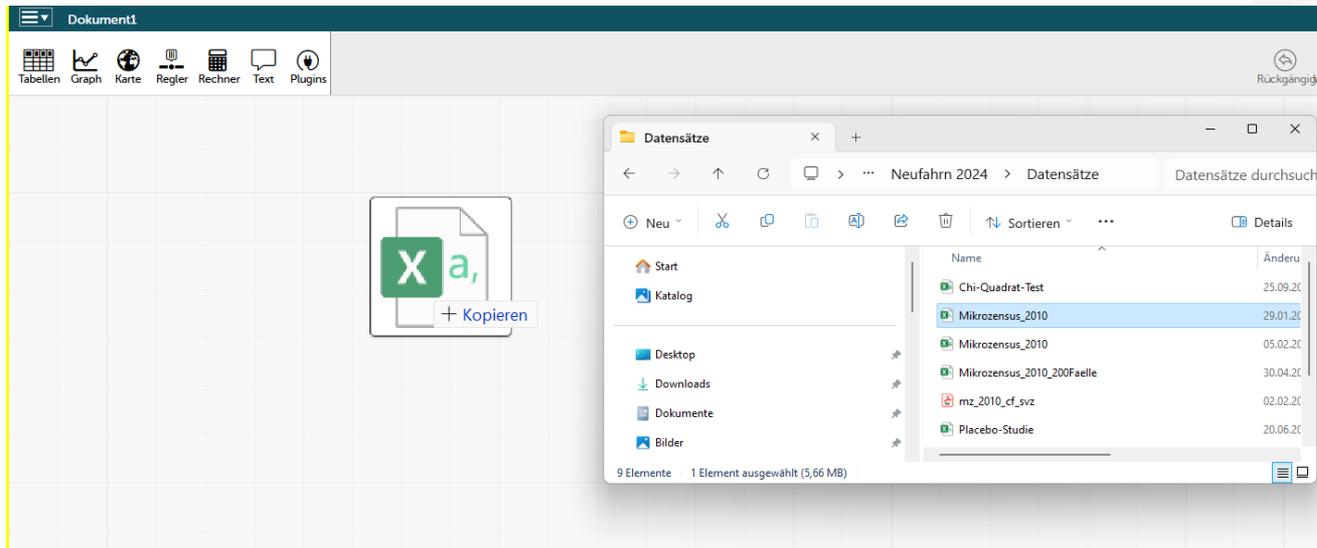
CODAP

Wir wechseln nun den Datensatz und
arbeiten mit dem Datensatz
Mirkozenus_2010.csv



CODAP

Ziehen Sie (gedrückte linke Maustaste) hierzu wieder den gewünschten Datensatz (hier: Mirkozensus_2010.csv) in CODAP:



Vorsicht: Dieser Datensatz ist nun sehr groß. Wir wollen erst mal nicht alle Zeilen importieren, sondern nur 500 Fälle (Zeilen) zufällig wählen.



CODAP

Bei der Durchführung der Tests, müssen Sie gegebenenfalls das Skalenniveau von metrisch auf kategorial umstellen, z.B. bei den Merkmalen „Deutscher Staatsangehöriger“ oder „In_Deutschland_geboren“. Bitte hierzu auf „123“ klicken, damit „abc“ erscheint:

2024i testimate

Datensatz: **Mikrozensus_2010**, 500 Fälle (ist nicht zufällig)

Deutscher_Staatsangehöriger In_Deutschland_geboren

Zielvariable/erstes Merkmal abc

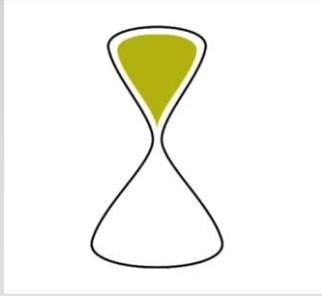
Prädiktor/zweites Merkmal abc

1-Stichproben-Test auf Anteilswert von Deutscher_Staatsangehöriger = 1

Ist der Anteil von (Deutscher_Staatsangehöriger = 1) \neq 0,5 ?

Anteil in der Stichprobe = 0,942, (471 von N = 500)
z = 19,8, P < 0.0001
95% KI = [0,9215, 0,9625]
SE = 0,01045, α = 0,05, z* = 1,96
(nach dem Gauß-Test (z))

A red arrow points to the 'abc' button in the 'Prädiktor/zweites Merkmal' field.



Workshop-Phase

Bearbeiten Sie bitte
Übungsblatt 3

Zeit: 10 Minuten





Mögliche Lösung:

2024i testimate

Datensatz: **Mikrozensus_2010**, 5000 Fälle (ist nicht zufällig) i

Geschlecht

Zielvariable/erstes Merkmal abc 🗑️

Ostdeutschland_Westdeutschland

Prädiktor/zweites Merkmal abc 🗑️

Test auf Unabhängigkeit von Ostdeutschland_Westdeutschland und Geschlecht v

Sind (Ostdeutschland_Westdeutschland) und (Geschlecht) unabhängig?
 $N = 5000$, 2 columns by 2 rows, $\chi^2 = 0,3245$, $P = 0,5689$

▼ Test auf Unabhängigkeit, 1-seitig χ^2 Zusammenfassung

beobachtet erwartet	Ostdeutschland_Westdeutschland =	West	Ost
Geschlecht =	männlich	1937 1929	499 507,2
	weiblich	2022 2030	542 533,8

$df = 1$, $\alpha = 0,05$, $\chi^{2*} = 3,841$

konfiguriere Test auf Unabhängigkeit von (Ostdeutschland_Westdeutschland) und (Geschlecht):
 $\alpha = 0.05$ 1-seitig

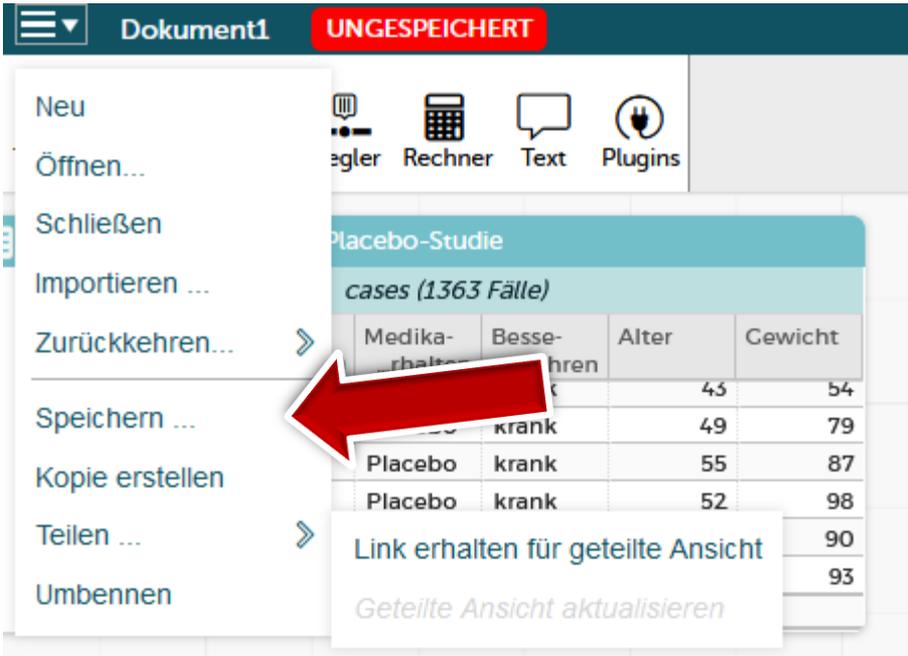
„Die Merkmale Wohnort (West vs. Ost) und Geschlecht (männlich vs. weiblich) hängen **nicht** signifikant voneinander ab.“

Vorsicht, das heißt aber nicht, dass die folgende Aussage richtig ist: „Die Merkmale Wohnort (West vs. Ost) und Geschlecht (männlich vs. weiblich) sind **signifikant voneinander unabhängig**.“



Abspeichern der CODAP-Daten

Sie können Ihr CODAP-Projekt auch abspeichern. Gehen Sie hierzu links oben auf die drei Striche und dann auf „Speichern“:



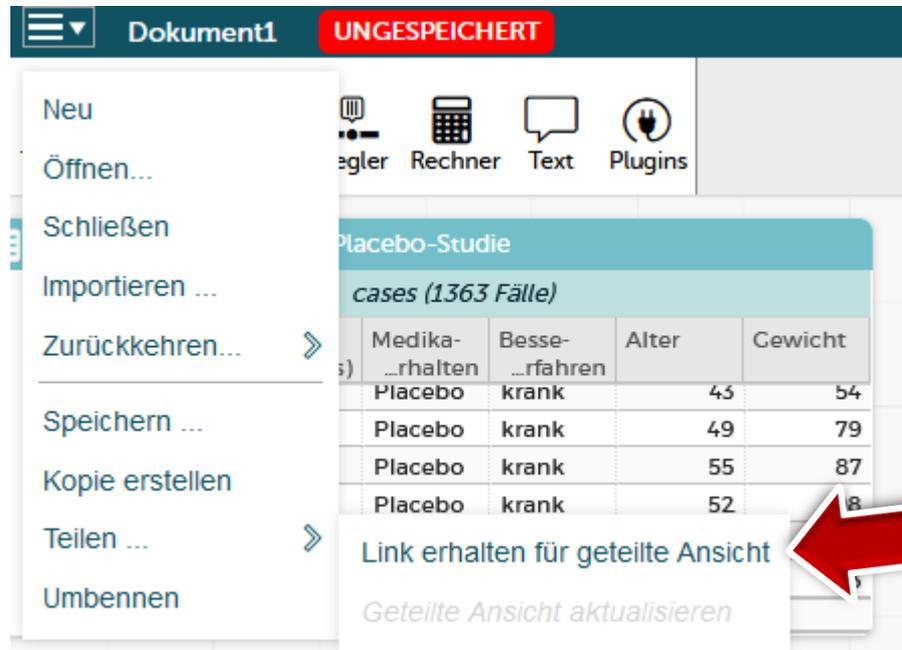
The screenshot shows the CODAP software interface. At the top, there is a dark blue header with a hamburger menu icon on the left, the text 'Dokument1', and a red badge that says 'UNGESPEICHERT'. Below the header is a toolbar with icons for 'Regler', 'Rechner', 'Text', and 'Plugins'. The main area displays a table titled 'Placebo-Studie' with the subtitle 'cases (1363 Fälle)'. The table has four columns: 'Medika-rhalten', 'Besse-ahren', 'Alter', and 'Gewicht'. A red arrow points to the 'Speichern ...' option in the left-hand menu. Below the 'Speichern ...' option, there is a sub-menu with the following items: 'Link erhalten für geteilte Ansicht', 'Geteilte Ansicht aktualisieren', and a numerical value '90'.

Medika-rhalten	Besse-ahren	Alter	Gewicht
		43	54
	krank	49	79
Placebo	krank	55	87
Placebo	krank	52	98
			90
			93



CODAP-Projekt teilen

Sie können Ihr CODAP-Projekt auch über einen Link an Ihre Schüler schicken, z.B. für eine Hausaufgabe. Klicken Sie hierzu auf die drei Striche links oben und dann auf „Teilen“ und „Link erhalten für geteilte Ansicht“:



The screenshot shows the CODAP software interface. At the top, there is a dark blue header with 'Dokument1' and a red 'UNGESPEICHERT' (Unsaved) indicator. Below the header is a toolbar with icons for 'Regler' (Slider), 'Rechner' (Calculator), 'Text', and 'Plugins'. The main workspace displays a table titled 'Placebo-Studie' with the subtitle 'cases (1363 Fälle)'. The table has four columns: 'Medika-...rhalten', 'Besse-...rfahren', 'Alter', and 'Gewicht'. The first four rows of data are visible, all with 'Placebo' in the first column and 'krank' in the second. A red arrow points to the 'Teilen ...' option in the left-hand menu, which has opened a sub-menu with the option 'Link erhalten für geteilte Ansicht'.

Medika-...rhalten	Besse-...rfahren	Alter	Gewicht
Placebo	krank	43	54
Placebo	krank	49	79
Placebo	krank	55	87
Placebo	krank	52	8



CODAP-Projekt teilen

Klicken Sie jetzt noch auf „Teilen aktivieren“:

TEILEN ✕

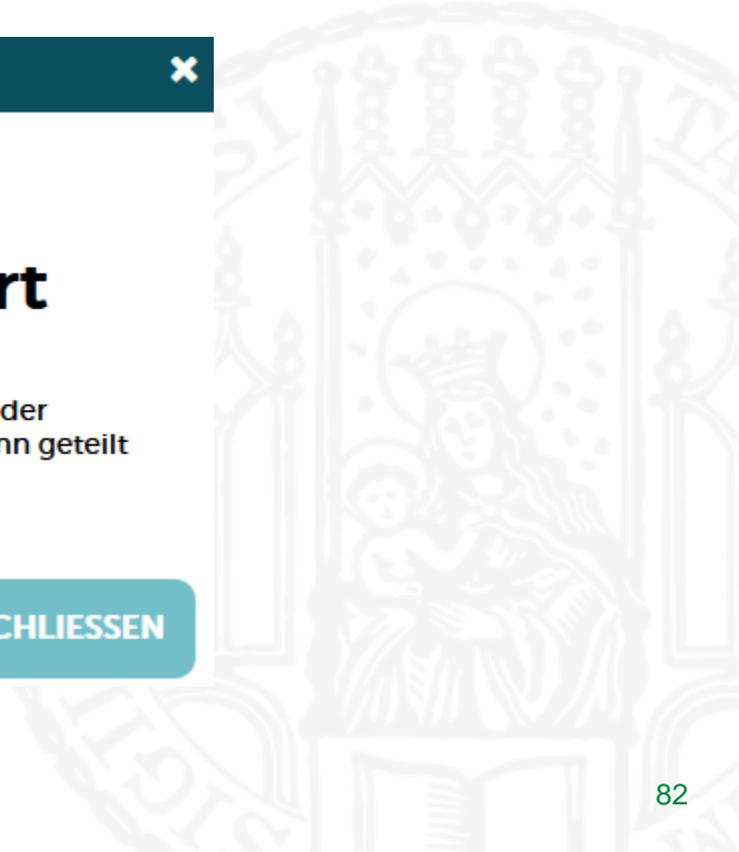
Geteilte Ansicht:deaktiviert

TEILEN AKTIVIEREN

When Teilen aktiviert ist, wird eine Kopie der aktuellen Ansicht erstellt. Diese Kopie kann geteilt werden.



SCHLIESSEN





CODAP-Projekt teilen

Jetzt können Sie die url kopieren und den Link Ihren Schülern schicken:



TEILEN

Geteilte Ansicht:aktiviert [Teilen](#)

[stoppen](#)

GETEILTE ANSICHT AKTUALISIEREN [Vorschau zu geteilter Ansicht](#)

Link | Einbetten | Interactive API

In eine Email oder Nachricht einfügen [Kopieren](#)

`https://codap.concord.org/app/static/dg/de/cert/index.html#shared=https%3A%2F`





Konkrete Arbeitsaufträge mit einbinden

Um die Lernumgebung attraktiver zu gestalten und mit passenden Arbeitsaufträgen zu verknüpfen, können Sie auch Textfelder erstellen. Klicken Sie hierzu oben auf Text:

The screenshot shows the CODAP interface. At the top, there is a dark blue header with a menu icon, the word 'Test', and a red button labeled 'UNGESPEICHERT'. Below this is a toolbar with icons for 'Tabellen', 'Graph', 'Karte', 'Regler', 'Rechner', 'Text', and 'Plugins'. A red arrow points to the 'Text' icon. Below the toolbar is a table titled 'Placebo-Studie' with the subtitle 'cases (1363 Fälle)'. The table has columns for 'Index', 'ID der Person', 'Geschlecht', 'Medikation', 'Beschreibung', and 'Gewicht'. The data rows are as follows:

Index	ID der Person	Geschlecht	Medikation	Beschreibung	Gewicht
1358	1358	w	Placebo	krank	54
1359	1359	m	Placebo	krank	79
1360	1360	w	Placebo	krank	87
1361	1361	m	Placebo	krank	98
1362	1362	w	Placebo	krank	90
1363	1363	m	Placebo	krank	93

Below the table, there is a vertical sidebar with the text 'Medikament erhalten' and 'Placebo'.



Wirkt das neue Medikament?

In einer Studie mit 1363 Versuchspersonen wurde ein neues Medikament gegen ein Placebo getestet. Unten siehst du die Daten, die dabei entstanden sind.

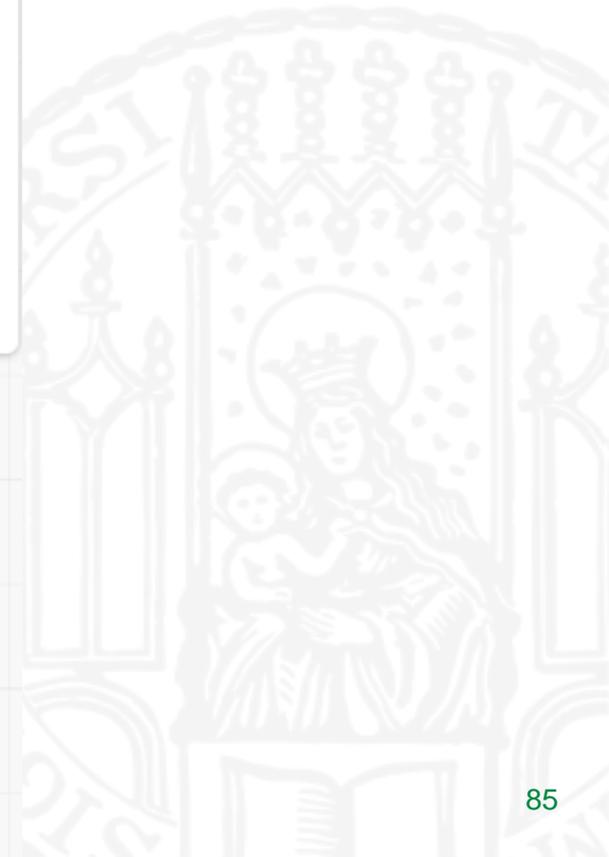


- a) Untersuche mit passenden Diagrammen und Kennwerten, ob das neu entwickelte Medikament besser ist als das Placebo.
- b) Führe einen passenden Hypothesentest durch, um die Wirksamkeit des Medikaments gegenüber des Placebos zu überprüfen.
- c) Welchen Einfluss haben dabei etwaige Hintergrundvariablen (wie z.B. das Alter)?

Placebo-Studie

cases (1363 Fälle)

Index	ID der Person	Geschlecht (=divers)	Medikament	Beschreibung	Alter	Gewicht
1357	1357	m	Placebo	krank	44	87
1358	1358	w	Placebo	krank	43	54
1359	1359	m	Placebo	krank	49	79
1360	1360	w	Placebo	krank	55	87
1361	1361	m	Placebo	krank	52	98
1362	1362	w	Placebo	krank	53	90
1363	1363	m	Placebo	krank	57	93



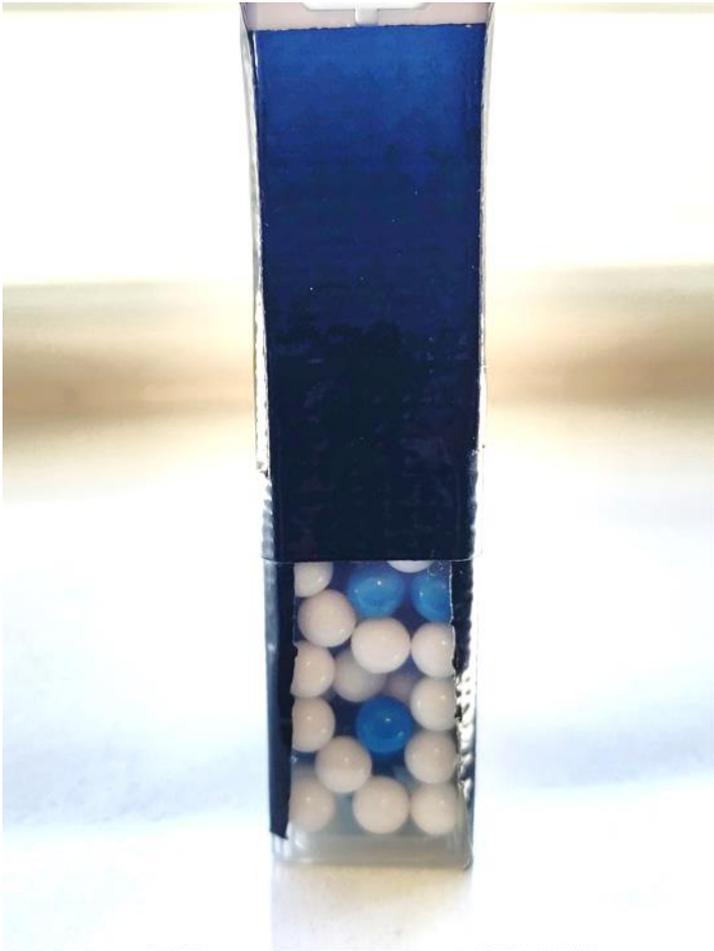
PAUSE 😊



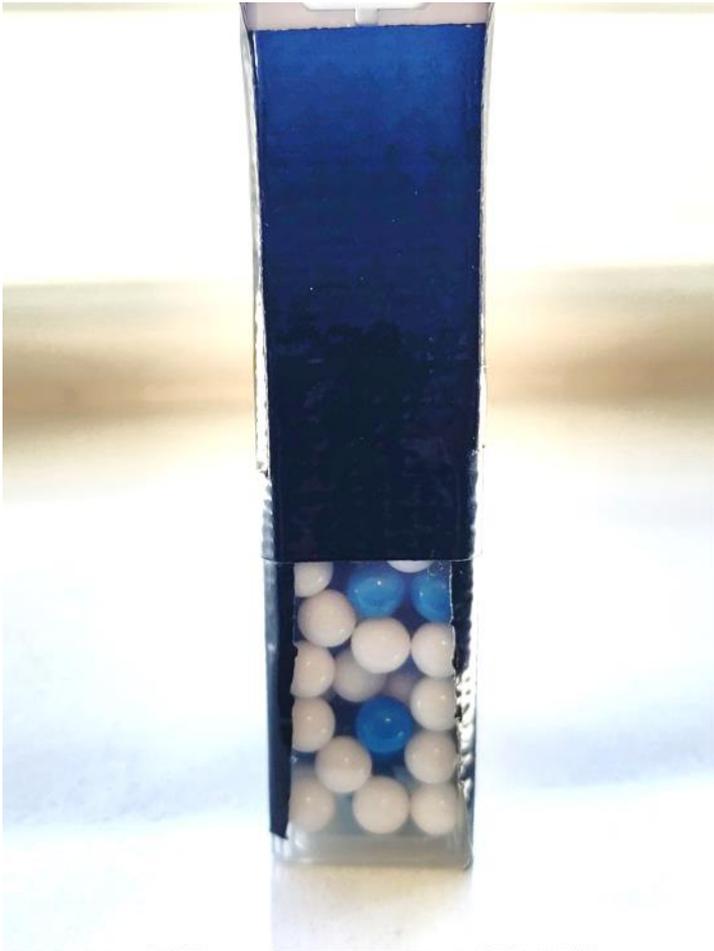
Die Schüttelbox

Schütteln Sie Ihre Box nur einmal.

Was glauben Sie, wie die weißen und blauen Kugeln in der Schüttelbox verteilt sind? Gibt es mehr weiße Kugeln oder mehr blaue? Und um wieviel mehr?



Die Schüttelbox



Schütteln Sie jetzt Ihre Box häufiger und beobachten Sie, welche Verteilungen weiß-blau Sie bei jedem Schütteln erhalten.

Was glauben Sie nun?

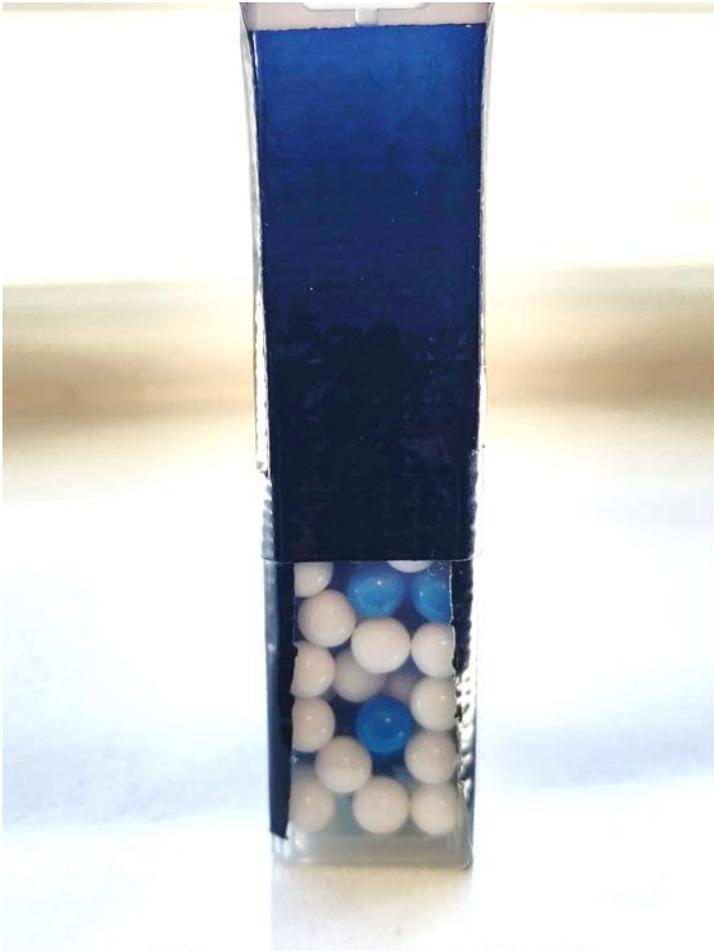
Sind in Ihrer Box mehr blaue oder mehr weiße Kugeln?

Die Schüttelbox

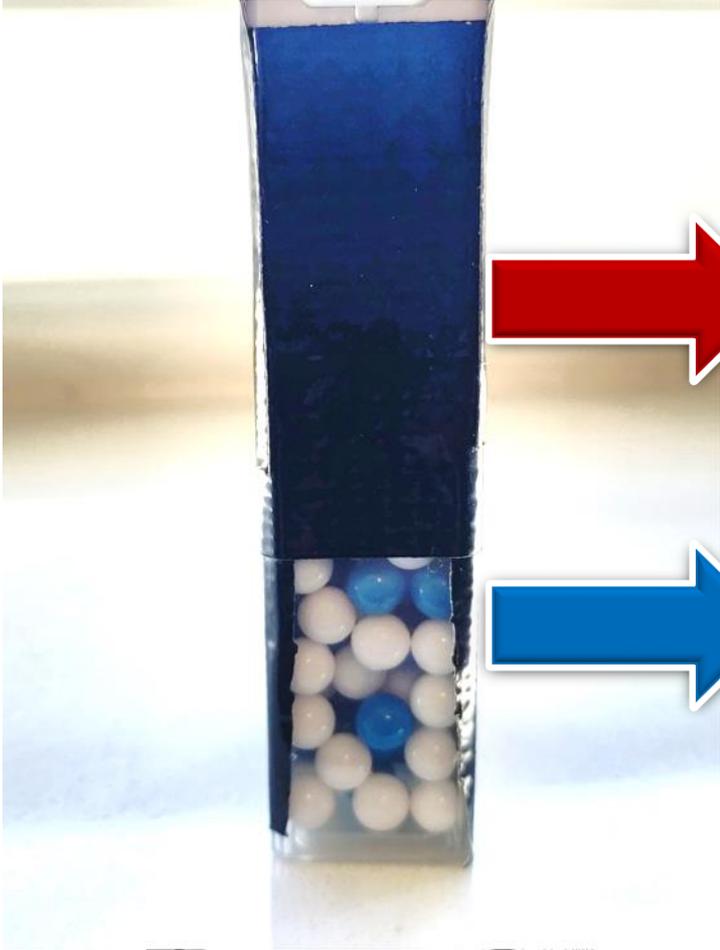
Die Schüttelboxen sind alle mit gleich vielen blauen und weißen Kugeln befüllt.

Wir können daher nun beobachten, welche Verteilung man erhält, wenn man die Box schüttelt. Manchmal sind nur ganz wenige weiße oder ganz wenige blaue Kugeln zu sehen.

Wie wahrscheinlich ist es, dass ich nur 3 blaue Kugeln (oder sogar noch weniger) aber 14 weiße Kugeln in der Stichproben-Scheibe sehe, wenn in der Null-Grundgesamtheits-Box „null Unterschied“ ist, also gleich viele blaue und weiße Kugeln?



p-Wert über Simulationen erklären



**Null-Grundgesamtheits-Box
(also kein Unterschied)**

Stichproben-Fenster

Wie wahrscheinlich ist es, dass ich nur 3 blaue Kugeln (oder sogar noch weniger) aber 14 weiße Kugeln in der Stichproben-Scheibe sehe, wenn in der Null-Grundgesamtheits-Box „null Unterschied“ ist, also gleich viele blaue und weiße Kugeln?

p-Wert über Simulationen mit CODAP erklären

Wie wahrscheinlich ist es, dass ich nur 3 blaue Kugeln (oder sogar noch weniger) aber 14 weiße Kugeln in der Stichprobenscheibe sehe, wenn in der Null-Grundgesamtheits-Box „null Unterschied“ ist, also gleich viele blaue und weiße Kugeln?

Dies können wir mithilfe von CODAP simulieren! Hierzu gibt es den **Binomial Simulator** auf der Seite <https://codap.xyz/>

001b
Binomial Simulator

Binomial Simulator

One Kugel is either **weiß** or **blau**.
 The probability of **weiß** is **0.5**.
 One Ziehung consists of **17** Kugels.
 Press to simulate **100** Ziehungs.

Settings

probability of **weiß** is **0.5**.
 Kugels per Ziehung Not more than 20000.
 Ziehungs per run. Probably between 100 and 400.

Vocabulary

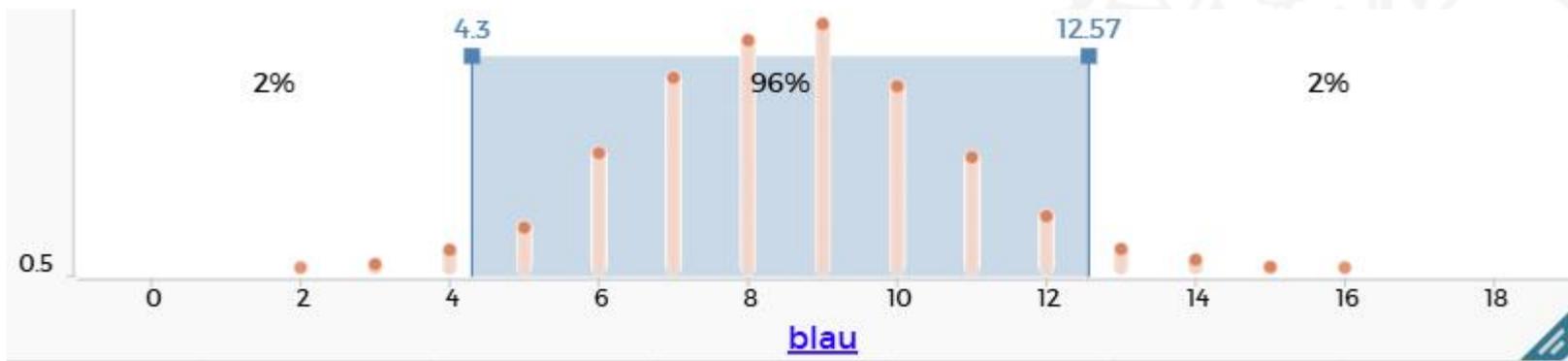
<input type="text" value="Kugel"/>	the basic event
<input type="text" value="weiß"/>	possible result of one Kugel
<input type="text" value="blau"/>	the alternative to weiß
<input type="text" value="Ziehung"/>	what do you call a set of 17 Kugels?

Note: Changing this vocabulary will delete your data on the next run.

p-Wert über Simulationen mit CODAP erklären

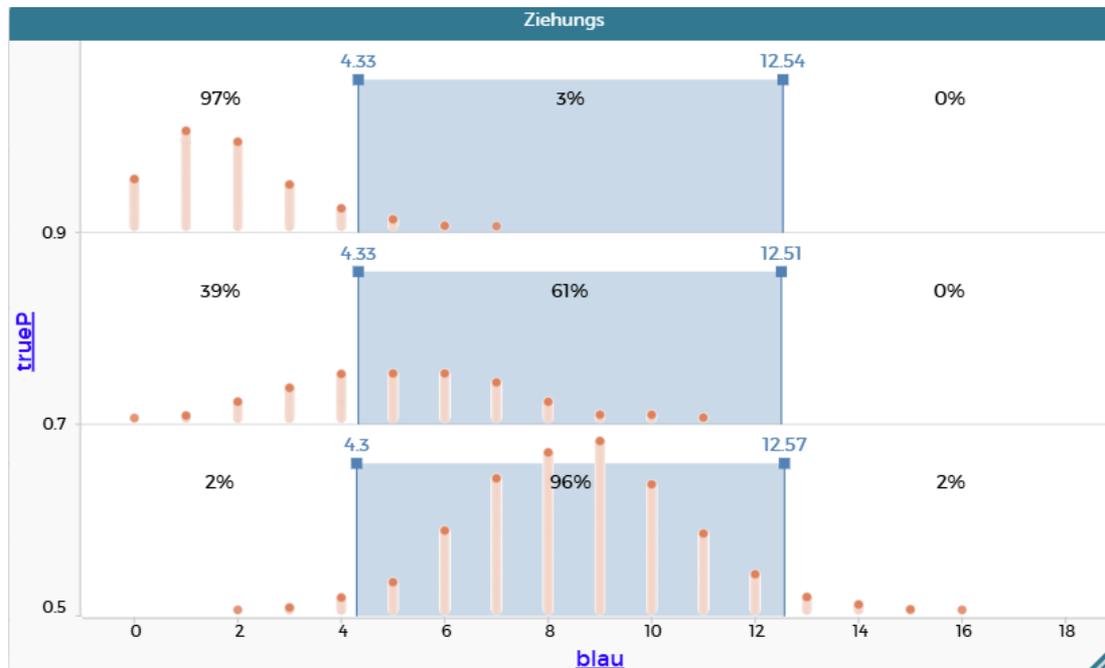
Wir simulieren 100 Ziehungen, in dem wir je 17 Kugel ziehen.
Im ersten Fall sollen es in der Box gleich viele blaue und weiße Kugeln sein.

In einem Graphen können wir uns anzeigen lassen, wie selten mehr als 13 oder weniger als 4 Kugeln blau sind. Das scheint in ca. 4 % der Fälle der Fall zu sein (knapp unter unserem festgelegten Signifikanzniveau von 5%).



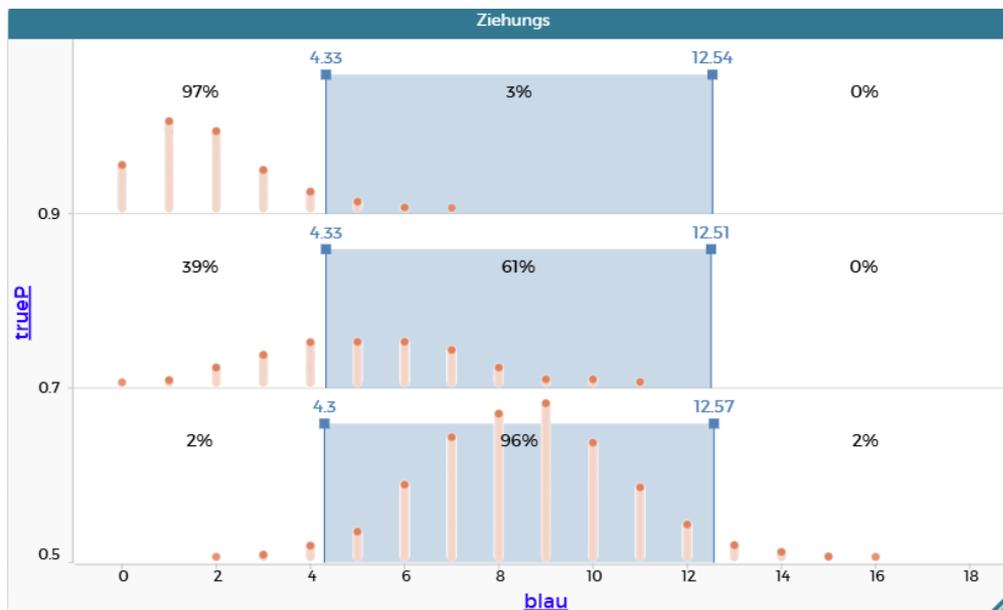
p-Wert über Simulationen mit CODAP erklären

Um den Fehler 2. Art nun mit ins Spiel zu bringen und uns zu überlegen, wie groß ein Effekt sein müsste, damit man ihn mit nur 17 Kugeln erfassen kann, simulieren wir nun aber auch Schüttelboxen mit einer Wahrscheinlichkeit von 70% weiß oder sogar 90% weiß.



p-Wert über Simulationen mit CODAP erklären

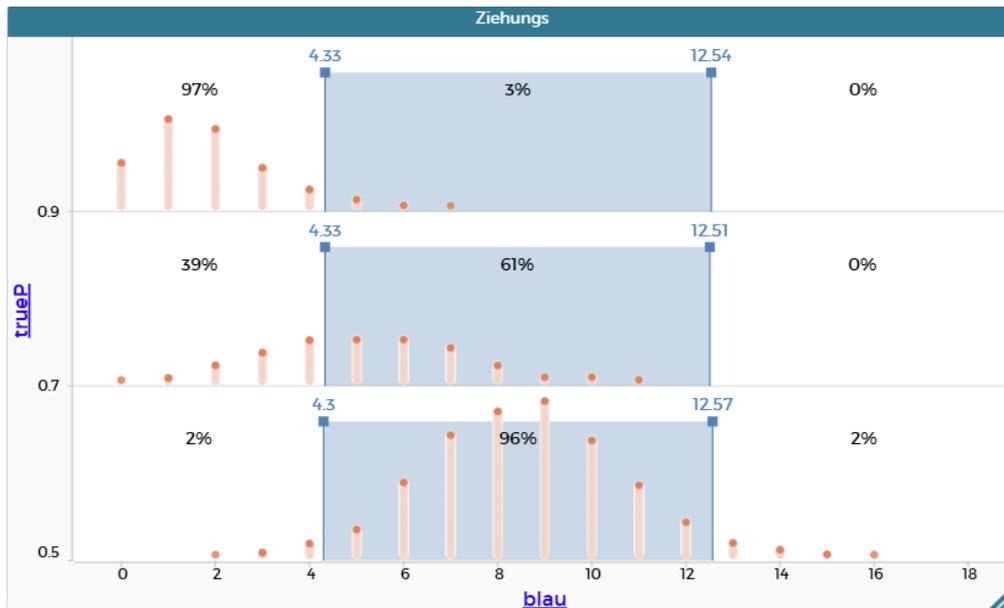
Wenn in der Realität 70% der Kugeln in der Grundgesamtheit weiß sind und ich bei 4 oder weniger blauen Kugeln die Nullhypothese ablehne und davon ausgehe, dass in der Box nicht gleich viele weiße und blaue Kugeln sind, würde ich in 61% einen Fehler 2. Art begehen und die Nullhypothese beibehalten.



Einen solchen Effekt:
Verteilung ist 70:30 statt
50:50 kann ich also mit so
wenigen Kugeln (nämlich
17) nicht gut erfassen!
Die Stichprobe ist zu
klein!

p-Wert über Simulationen mit CODAP erklären

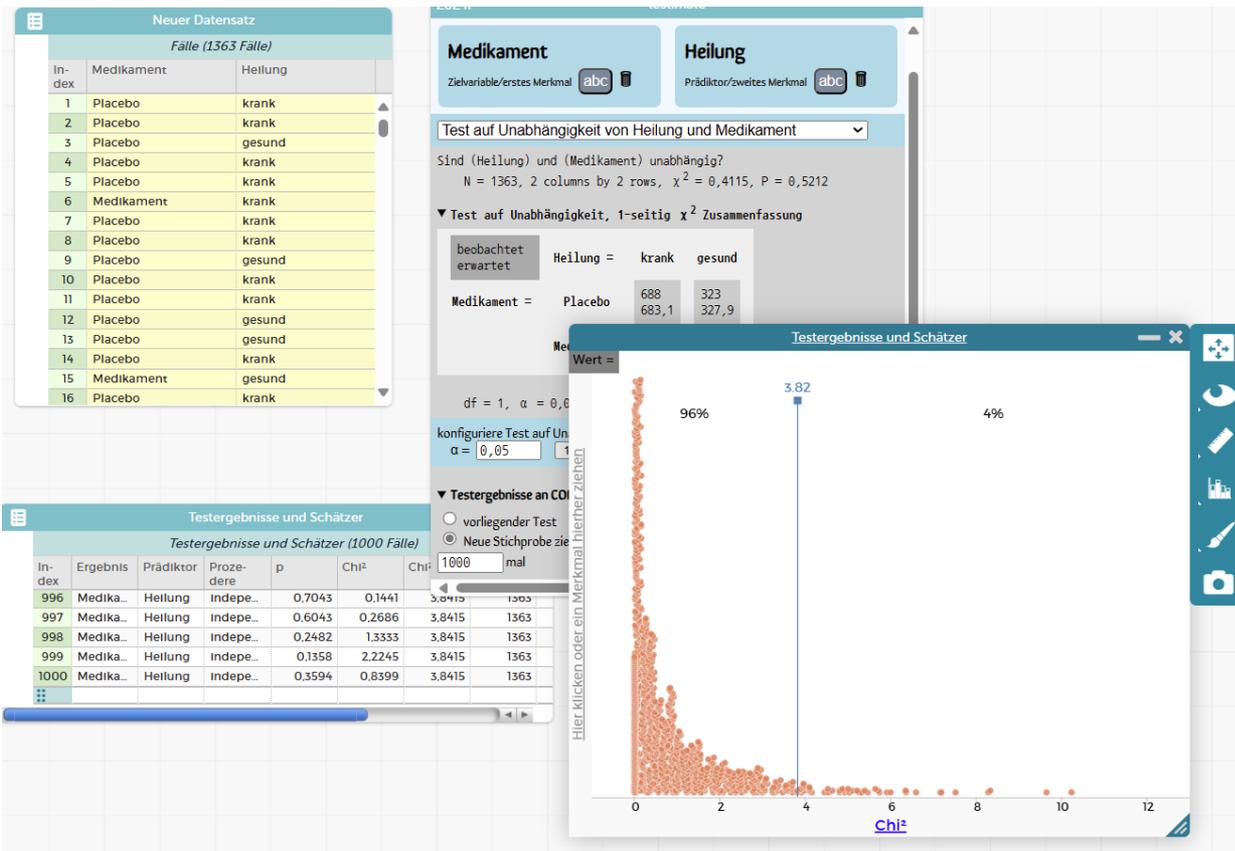
Wenn aber in der Realität 90% der Kugeln in der Grundgesamtheit weiß sind und ich bei 4 oder weniger blauen Kugeln die Nullhypothese ablehne und davon ausgehe, dass in der Box nicht gleich viele weiße und blaue Kugeln sind, würde ich nur in 3% einen Fehler 2. Art begehen und die Nullhypothese beibehalten.



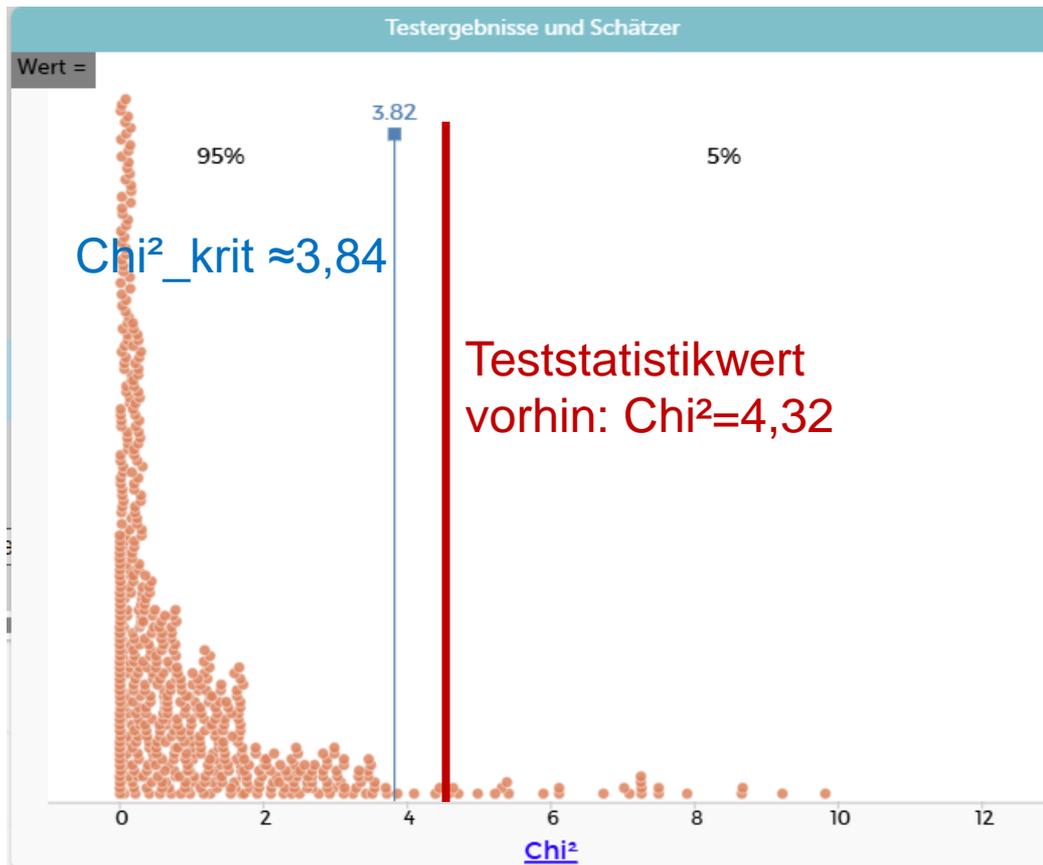
17 Kugeln würden also tatsächlich ausreichen, um einen solch großen Effekt (90:10 statt 50:50) nachweisen zu können.

Bei realen Experimenten wäre dann aber wichtig, diese 17 (Personen, Objekte, Merkmalsträger) gut auszuwählen (repräsentativ).

Zusammenhang Teststatistikwert, p-Wert, Annahme- und Ablehnungsbereich, Signifikanzniveau



Zusammenhang Teststatistikwert, p-Wert, Annahme- und Ablehnungsbereich, Signifikanzniveau



Wenn ich aus einer Welt Stichproben ziehe, in der Medikamenteneinnahme und Heilung nicht voneinander abhängen, hätte ich in nur ca. 5% der Fälle einen Teststatistikwert, der bei 3,84 liegt oder sogar noch größer ist. Da unser berechnete Teststatistikwert sogar noch größer ist, sprechen wir davon, dass die Merkmale Medikamenteneinnahme und Heilung signifikant voneinander abhängen.

Zusammenhang Teststatistikwert, p-Wert, Annahme- und Ablehnungsbereich, Signifikanzniveau

Wir wollten ja den Fehler 1. Art möglichst klein halten \Rightarrow Signifikanzniveau $\alpha = 0,05$.

Der p-Wert gibt die folgende Wahrscheinlichkeit an: Wie wahrscheinlich ist es, diese Abhängigkeit oder sogar noch eine extremere Abhängigkeit in den Daten zu erhalten, obwohl ich die Daten aus einer Welt gezogen habe, in der stochastische Unabhängigkeit zwischen den Merkmalen herrscht.

Gedankenexperiment: Wir ziehen 10.000 Stichproben aus einer erdachten Null-Grundgesamtheit, in der stochastische Unabhängigkeit herrscht. Dann erhalten wir ja 10.000 verschiedene Teststatistikwerte χ^2 .

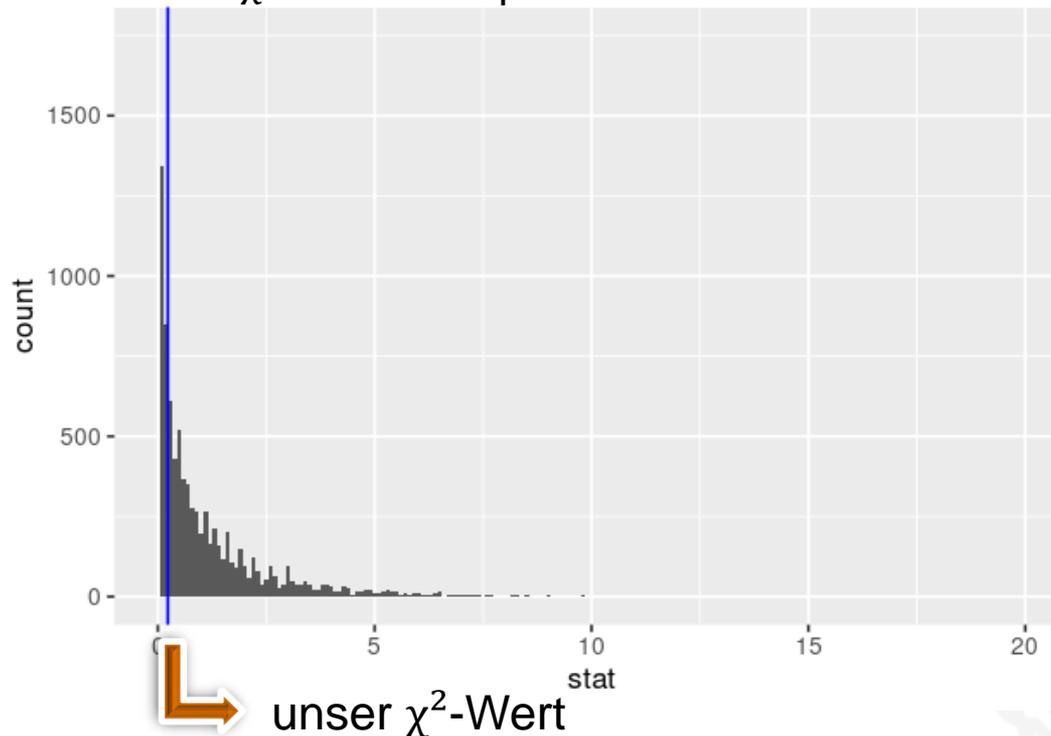
Nun vergleicht man: Wo liegen diese 10.000 Teststatistikwerte? Und wo liegt im Vergleich dazu mein Teststatistikwerte χ^2 , den ich gerade zuvor berechnet hatte (aus einer Stichprobe, bei der ich ja nicht wissen kann, ob sie aus einer Null-Grundgesamtheit stammt).

Grundsätzliche Frage: Ist es plausibel bei all den Werten, dass mein Teststatistikwert aus dieser Grundgesamtheit stammt, in der keine Abhängigkeit herrscht?

Zusammenhang Teststatistikwert, p-Wert, Annahme und Ablehnungsbereich, Signifikanzniveau

Die Verteilung, die wir erhalten, heißt χ^2 -Verteilung

So verteilen sich die χ^2 -Werte beispielsweise in einer stochastisch unabhängigen Welt:

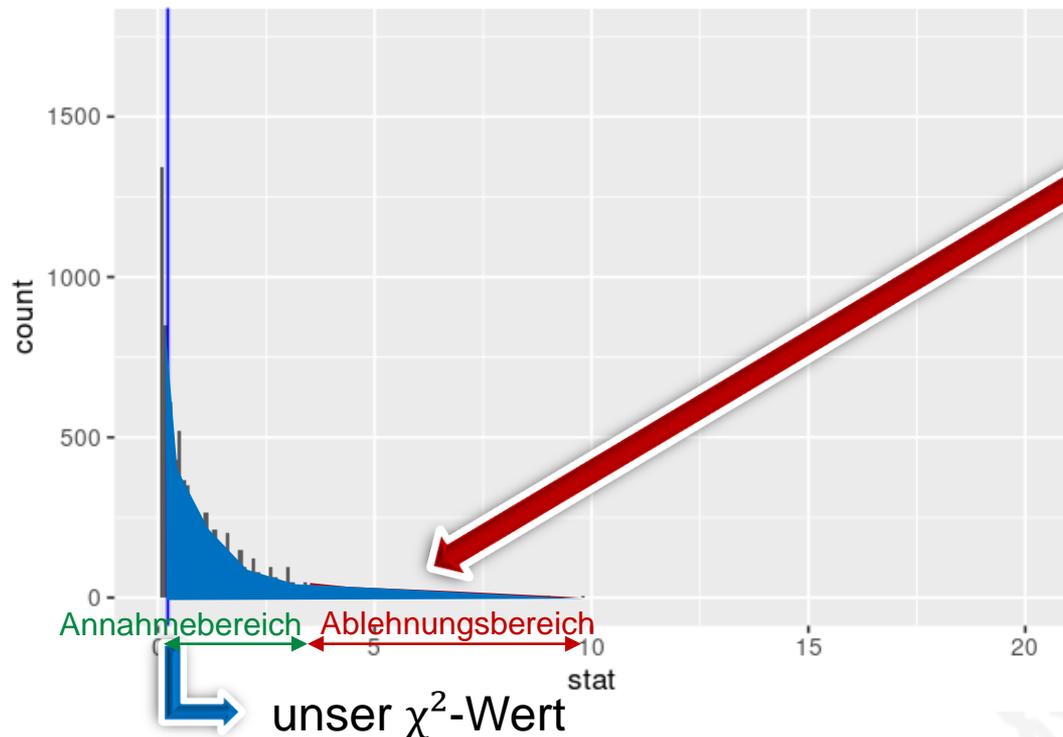


In diesem Beispiel wäre es also sehr plausibel, dass der χ^2 -Wert aus einer „Welt“ stammt, in der stochastische Unabhängigkeit herrscht.

Hier: p-Wert 90%
Nullhypothese kann nicht
abgelehnt werden.

Zusammenhang Teststatistikwert, p-Wert, Annahme und Ablehnungsbereich, Signifikanzniveau

Die χ^2 -Verteilung ist eigentlich eine stetige Verteilung. Der Flächeninhalt unter der Kurve ist „1“, also die Wahrscheinlichkeit, irgendeinen der möglichen Teststatistikwerte zu erhalten.



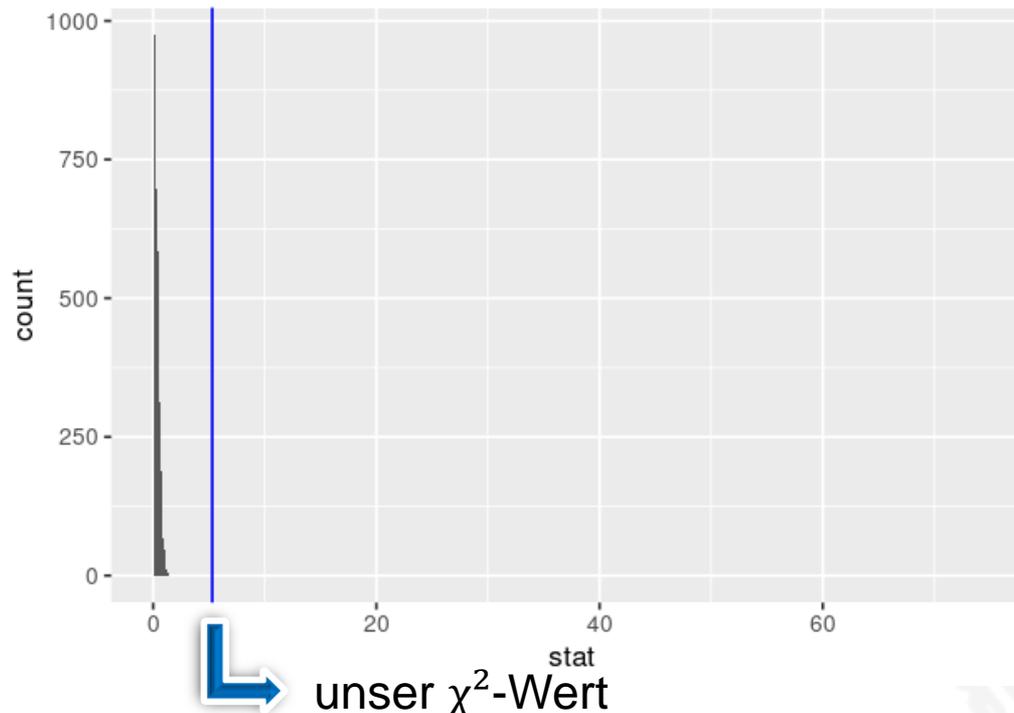
Das **Signifikanzniveau 5%** entspricht dem rot eingezeichneten Flächeninhalt. Diesen Fehler wollen wir maximal haben.

Der **p-Wert** entspricht der Wahrscheinlichkeit, die zu unserem χ^2 -Wert gehört.

Hier: p-Wert < 0,01%
Nullhypothese wird.

Zusammenhang Teststatistikwert, p-Wert, Annahme und Ablehnungsbereich, Signifikanzniveau

In diesem Beispiel wäre es hingegen äußerst unplausibel, dass der χ^2 -Wert aus einer „Welt“ stammt, in der stochastische Unabhängigkeit herrscht.



Der t-Test





Der t-Test für Unterschiedshypothesen

Merkmal 1 (kategorial
bzw. dichotom)

Merkmal 2 (metrisch)



Fragestellung	Gruppe 1	Gruppe 2	Zielvariable
Unterscheidet sich das Einkommen von Männern und Frauen?	Frauen	Männer	z.B. durchschnittliches Einkommen
Zeigen die Schüler:innen in Asien bessere Leistungen als deutsche Schüler:innen?	Schüler:innen in Asien	Schüler:innen in Deutschland	z. B. mittlere Leistung in einem Schulleistungstest
Rasen Autofahrer in der Oberpfalz 2022 weniger als im Jahr 2021?	Autofahrer in der Oberpfalz im Jahr 2021	Autofahrer in der Oberpfalz im Jahr 2022	z. B. gemessene durchschnittliche Geschwindigkeiten z. B. durchschnittliche Anzahl an geblitzten Fahrern

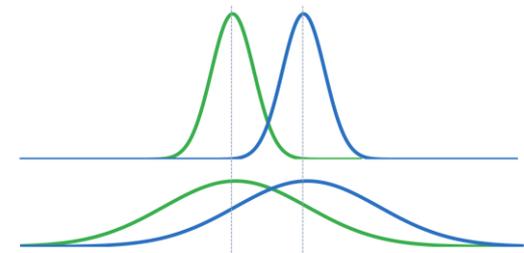


Nebenbemerkung: Gerichtete vs. ungerichtete Hypothesen

Fragestellung	
Unterscheidet sich das Einkommen von Männern und Frauen?	<p>Hierbei handelt es sich um eine ungerichtete Hypothese! Forschungshypothese: $\mu_{\text{Frauen}} \neq \mu_{\text{Männer}}$</p>
Zeigen die Schüler:innen in Asien bessere Leistungen als deutsche Schüler:innen?	<p>Hierbei handelt es sich um eine gerichtete Hypothese! Forschungshypothese: $\mu_{\text{Asien}} > \mu_{\text{Deutschland}}$ und $\mu_{2022} < \mu_{2021}$</p>
Rasen Autofahrer in der Oberpfalz 2022 weniger als im Jahr 2021?	

Wir betrachten hier durchgängig ungerichtete Hypothesen. Da sich bei ungerichteten Hypothesen der Fehler 1. Art auf zwei Seiten der Verteilung der Teststatistikwerte (bei „kein Unterschied“) aufteilt, könnte man gerichtete Hypothesen betrachten, indem man das Signifikanzniveau verdoppelt.

„Unterscheiden sich die Mittelwerte in zwei verschiedenen Gruppen?“ oder „Trampeltier oder Dromedar“?





Zurück zum t-Test ▪ Mit CODAP

Zurück zur Placebo-Studie. Wir möchten mehr über unsere Stichprobe erfahren und überprüfen, ob sich das Durchschnittsalter in der Placebo-Gruppe und der Medikamenten-Gruppe unterschieden hatte.

Merkmal 1: Alter – metrisch

Merkmal 2: Placebo- vs. Medikamenten-Gruppe – kategorial

Öffnen Sie hierzu wieder den Placebo-Daten-Satz in CODAP.



Zurück zum t-Test ▪ Mit CODAP

Ziehen Sie nun das „Alter“ auf das erste Merkmal und „Medikament erhalten“ auf das zweite Merkmal.

Achten Sie wieder darauf, ob richtig eingestellt ist, ob es sich um ein **metrisches** („123“) oder **kategoriales** („abc“) Merkmal handelt:

2024i testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig) i

Alter

Zielvariable/erstes Merkmal 123

Medikament erhalten

Prädiktor/zweites Merkmal abc





Zurück zum t-Test ▪ Mit CODAP

Wählen Sie als Testart aus „Differenz der Mittelwerte.... gruppiert durch...“

2024i testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig) i

Alter

Zielvariable/erstes Merkmal 123 🗑️

Medikament erhalten

Prädiktor/zweites Merkmal abc 🗑️

Differenz der Mittelwerte Alter gruppiert durch Medikament erhalten ▼

Ist die Differenz arithmetisches Mittel (Alter) : Gruppe Medikament - Gruppe Pla

N = 1363, t = -16, P < 0.0001
diff = -9,95, 95% KI = [-11,17, -8,731]

► Differenz der Mittelwerte, 2-Stichprobenfall t Verfahren

Test auf Mittelwert(Alter):

[Medikament] - [Placebo] ≠ 0 ⬇

conf = 95 ⬇ %





Zurück zum t-Test ▪ Mit CODAP

Mit Klick auf den kleinen Pfeil links sehen wir auch die deskriptiven Ergebnisse. Im Mittel sind die Probanden in der Placebo-Gruppe etwa **10 Jahre älter** als die Personen in der Medikamentengruppe.

Der Unterschied ist **signifikant**, da der zugehörige p-Wert unter 0,05 liegt.

2024i
testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig) ?

Alter

Zielvariable/erstes Merkmal 123 🗑️

Medikament erhalten

Prädiktor/zweites Merkmal abc 🗑️

Differenz der Mittelwerte Alter gruppiert durch Medikament erhalten ▾

Ist die Differenz arithmetisches Mittel (Alter) ... Medikament - Gruppe Pla

N = 1363, t = -16, P < 0.0001
diff = -9,95, 95% KI = [-11,17, -8,731]

▼ Differenz der Mittelwerte, 2-Stichprobenfall t Verfahren

Medikament erhalten	N	arithmetisches Mittel(Alter)	s	SE
Medikament	435	30,73	9,313	0,4465
Placebo	928	40,67	11,26	0,3695
pooled	1363	diff = -9,95	10,68	0,6203





Zurück zum t-Test ▪ Mit CODAP

Diese **Verzerrung in der Stichprobe** könnte einen Effekt auf unseren zuvor berechneten Chi-Quadrat-Test haben.

Nämlich z.B. dann, wenn das Medikament bei jüngeren Patienten besser wirkt als bei älteren Patienten.

Das prüfen wir jetzt!

2024i
testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig) ?

Alter

Zielvariable/erstes Merkmal 123 🗑️

Medikament erhalten

Prädiktor/zweites Merkmal abc 🗑️

Differenz der Mittelwerte Alter gruppiert durch Medikament erhalten ▾

Ist die Differenz arithmetisches Mittel (Alter) - Gruppe Placebo

N = 1363, t = -16, P < 0.0001
diff = -9,95, 95% KI = [-11,17, -8,731]

▼ Differenz der Mittelwerte, 2-Stichprobenfall t Verfahren

Medikament erhalten	N	arithmetisches Mittel(Alter)	s	SE
Medikament	435	30,73	9,313	0,4465
Placebo	928	40,67	11,26	0,3695
pooled	1363	diff = -9,95	10,68	0,6203





Zurück zum t-Test ▪ Mit CODAP

Berechnen wir hierzu doch einen weiteren t-Test, um zu untersuchen, ob sich das Durchschnittsalter bei den geheilten vs. nicht-geheilten ebenfalls signifikant unterscheidet.

2024i testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig) i

Alter

Zielvariable/erstes Merkmal 123

Besserung erfahren

Prädiktor/zweites Merkmal abc

Differenz der Mittelwerte Alter gruppiert durch Besserung erfahren ▾

Ist die Differenz arithmetisches Mittel (Alter) : Gruppe gesund - Gruppe krank =

N = 1363, t = -9,17, P < 0.0001
diff = -5,9, 95% KI = [-7,167, -4,642]

▼ Differenz der Mittelwerte, 2-Stichprobenfall t Verfahren

Besserung erfahren	N	arithmetisches Mittel(Alter)	s	SE
gesund	893	35,46	11,86	0,3968
krank	470	41,37	10,15	0,468
pooled	1363	diff = -5,9	11,3	0,6438

df = 1360, α = 0,05, t* = 1,96



Zurück zum t-Test ▪ Mit CODAP

Tatsächlich: Das Durchschnittsalter ist in der Gruppe der geheilten Personen signifikant niedriger als in der Gruppe der kranken Personen.

Derartige unterrichtliche Aktivitäten sind besonders reichhaltig, da wir nun auch über den Einfluss von Hintergrundvariablen nachdenken!

2024i testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig) i

Alter

Zielvariable/erstes Merkmal 123 🗑️

Besserung erfahren

Prädiktor/zweites Merkmal abc 🗑️

Differenz der Mittelwerte Alter gruppiert durch Besserung erfahren ▾

Ist die Differenz arithmetisches Mittel (Alter) : Gruppe gesund - Gruppe krank =

N = 1363, t = -9,17, P < 0.0001
diff = -5,9, 95% KI = [-7,167, -4,642]

▼ Differenz der Mittelwerte, 2-Stichprobenfall t Verfahren

Besserung erfahren	N	arithmetisches Mittel(Alter)	s	SE
gesund	893	35,46	11,86	0,3968
krank	470	41,37	10,15	0,468
pooled	1363	diff = -5,9	11,3	0,6438

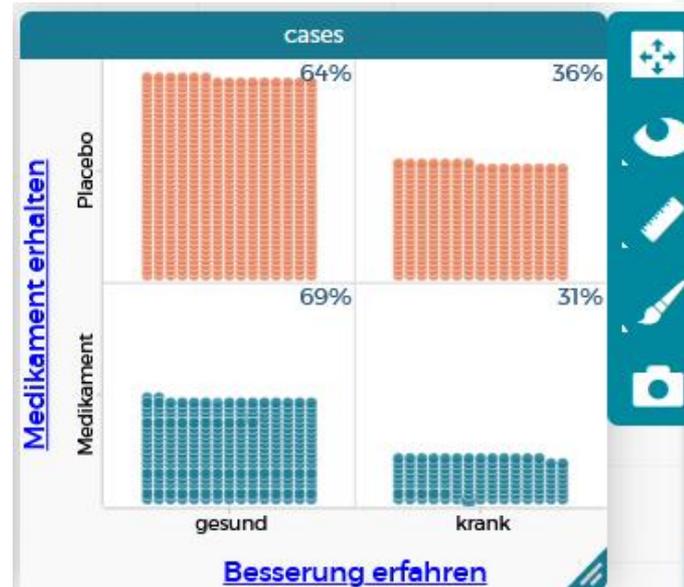
df = 1360, α = 0,05, t* = 1,96



Der t-Test ■ Mit CODAP

Filtern wir nun nur die Fälle, in denen die Person tatsächlich das Medikament erhalten hat (und kein Placebo) und wiederholen wir den Test.

Sie können die Fälle einfach auswählen, in dem Sie diese im Diagramm markieren (linke Maustaste gedrückt halten):





Der t-Test ■ Mit CODAP

Auch im Datensatz selbst sind diese Fälle nun alle markiert. Und dort (**Vorsicht:** bitte nicht im Diagramm, sondern im Datensatz!) alle Fälle ablegen (im Sinne von „wegelegen“ „beiseitelegen“), die nicht markiert sind, indem Sie zunächst auf das Symbol mit dem Auge klicken und dann auf „Nicht ausgewählte Fälle ablegen“):

The screenshot shows the CODAP interface for a dataset titled "Placebo-Studie" containing 1363 cases. The table displays columns for Index, ID der Person, Geschlecht, Medikation, Besserungserwartung, Alter, and Gewicht. A context menu is open over the table, showing options to manage selected and unselected cases. A red arrow points to the "Nicht ausgewählte Fälle ablegen" option.

Index	ID der Person	Geschlecht	Medikation	Besserungserwartung	Alter	Gewicht
419	419	w	Medika...	krank	40	70
420	420	w	Medika...	krank	41	55
421	421	m	Medika...	krank	28	70
422	422	m	Medika...	krank	33	76
423	423	w	Medika...	krank	35	80
424	424	w	Medika...	krank	38	76
425	425	m	Medika...	krank	42	9

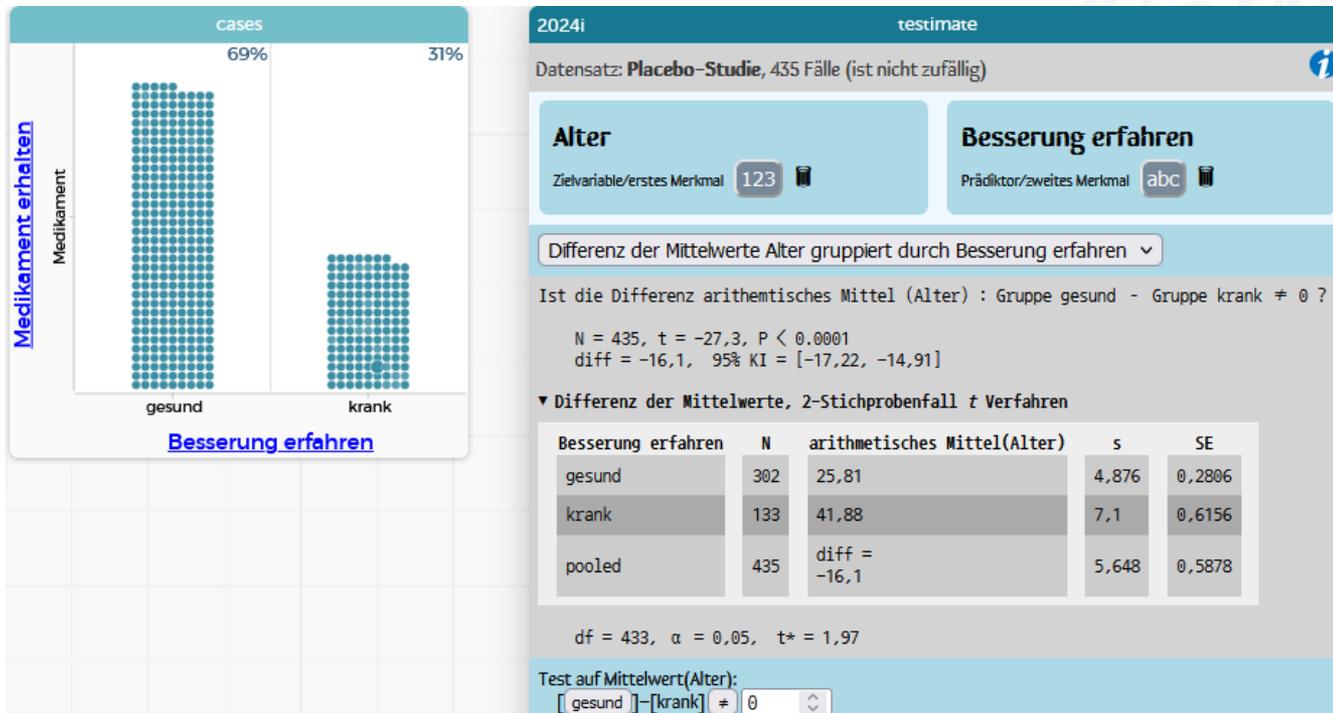
Context menu options:

- Ausgewählte Fälle ablegen
- Nicht ausgewählte Fälle ablegen
- Abgelegte Fälle Ordnen / Wiederherstellen
- Versteckte Merkmale anzeigen



Der t-Test ■ Mit CODAP

Sie merken nun, dass das Diagramm auch nur noch diese Fälle anzeigt, in denen die Personen das Medikament erhalten haben. Und auch der t-Test wird nun auf Grundlage dieser Stichprobe durchgeführt.





Der t-Test ■ Mit CODAP

2024i testimate

Datensatz: **Placebo-Studie**, 435 Fälle (ist nicht zufällig) ?

Alter

Zielvariable/erstes Merkmal **123**

Besserung erfahren

Prädiktor/zweites Merkmal **abc**

Differenz der Mittelwerte Alter gruppiert durch Besserung erfahren ▼

Ist die Differenz arithmetisches Mittel (Alter) : Gruppe gesund - Gruppe krank $\neq 0$?

N = 435, t = -27,3, P < 0.0001
diff = -16,1, 95% KI = [-17,22, -14,91]

▼ Differenz der Mittelwerte, 2-Stichprobenfall t Verfahren

Besserung erfahren	N	arithmetisches Mittel(Alter)	s	SE
gesund	302	25,81	4,876	0,2806
krank	133	41,88	7,1	0,6156
pooled	435	diff = -16,1	5,648	0,5878

df = 433, $\alpha = 0,05$, $t^* = 1,97$

Test auf Mittelwert(Alter):
 - \neq
 conf = %

▼ Testergebnisse an CODAP übermitteln

vorliegender Test

Auch wenn wir nur Personen betrachten, die das Medikament einnehmen, zeigt sich ein signifikanter Effekt des Alters.

Das Durchschnittsalter ist in der Gruppe der geheilten Personen signifikant niedriger als in der Gruppe der kranken Personen.



Der t-Test ■ Mit CODAP

Möchte man nun wieder alle Daten aus dem Datensatz in die Analysen einbeziehen, klicken Sie wieder auf das Augensymbol und dann auf „Abgelegte Fälle wiederherstellen“:

Placebo-Studie

cases (435 Fälle, 928 ablegen)

In- dex	ID der Person	Ge- ...=divers)	Medika- ...rhalten	Besse- rung er-	Alter	Gewicht
424	441	m	Medika...	gesund	44	90
425	442	m	Medika...	gesund	43	101
426	443	w	Medika...	gesund	49	79
427	444	w	Medika...	gesund	55	60
428	445	m	Medika...	gesund	52	70
429	446	m	Medika...	gesund	53	70
430	447	w	Medika...	gesund	57	61
431	448	w	Medika...	gesund	24	5

Ausgewählte Fälle ablegen

Nicht ausgewählte Fälle ablegen

[Abgelegte Fälle 928 wiederherstellen](#)

Versteckte Merkmale zeigen



Der t-Test ■ Mit CODAP

Wählen wir nun etwa gleich viele alte und junge Personen in den beiden Gruppen Medikament vs. Placebo, so sieht man: Der signifikante Effekt ist sogar in der anderen Richtung: Das Placebo scheint besser zu „wirken“ als das Medikament:

ent?

es Medikament gegen ein Placebo sind.

b das neu entwickelte Medikament besser

amkeit des Medikaments gegenüber des

(wie z.B. das Alter)?

cases

Medikament	Placebo	89%	11%
	Medikament	16%	84%
		gesund	krank

Besserung erfahren

2024i **testimate**

Datensatz: **Placebo-Studie**, 274 Fälle (ist nicht zufällig)

Medikament erhalten

Zielvariable/erstes Merkmal abc

Besserung erfahren

Prädiktor/zweites Merkmal abc

Test auf Unabhängigkeit von Besserung erfahren und Medikament erhalten

Sind (Besserung erfahren) und (Medikament erhalten) unabhängig?
N = 274, 2 columns by 2 rows, $\chi^2 = 146,7$, $P < 0.0001$

▼ Test auf Unabhängigkeit, 1-seitig χ^2 Zusammenfassung

beobachtet erwartet	Besserung erfahren =	gesund	krank
Medikament erhalten =	Medikament	23 73,03	122 71,97
	Placebo	115 64,97	14 64,03

df = 1, $\alpha = 0,05$, $\chi^2_{*} = 3,841$

konfiguriere Test auf Unabhängigkeit von (Besserung erfahren) und (Medikament erhalten):
 $\alpha = 0.05$ 1-seitig

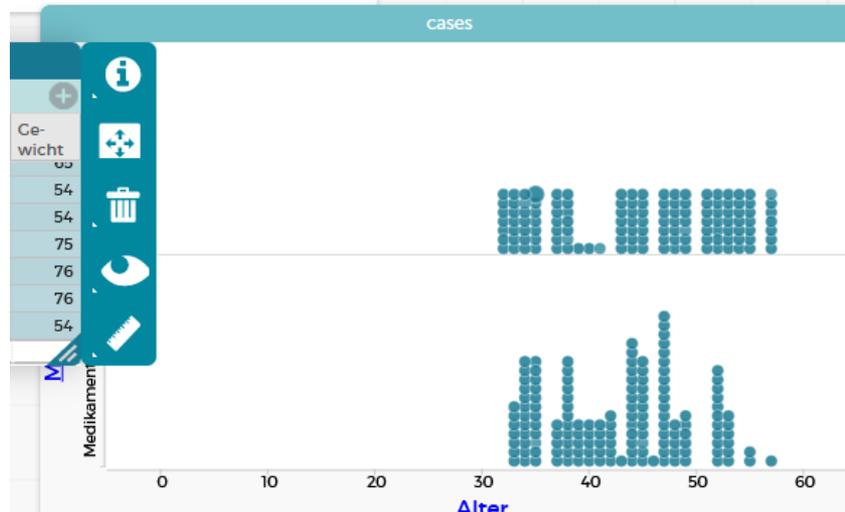
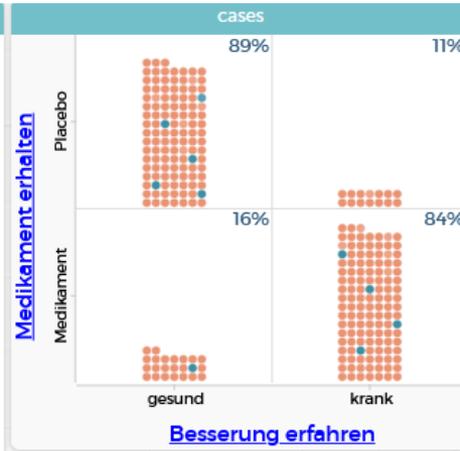
▼ Testergebnisse an CODAP übermitteln
vorliegender Test



Der t-Test ■ Mit CODAP

ent?
ies Medikament gegen ein Placebo sind.

b das neu entwickelte Medikament besser
 amkeit des Medikaments gegenüber des
 (wie z.B. das Alter)?



2024i estimate

Datensatz: **Placebo-Studie**, 274 Fälle (ist nicht zufällig)

Medikament erhalten
 Zielvariable/erstes Merkmal abc

Besserung erfahren
 Prädiktor/zweites Merkmal abc

Test auf Unabhängigkeit von Besserung erfahren und Medikament erhalten

Sind (Besserung erfahren) und (Medikament erhalten) unabhängig?
 N = 274, 2 columns by 2 rows, $\chi^2 = 146,7$, $P < 0.0001$

▼ Test auf Unabhängigkeit, 1-seitig χ^2 Zusammenfassung

beobachtet erwartet	Besserung erfahren =	gesund	krank
Medikament erhalten =	Medikament	23 73,03	122 71,97
	Placebo	115 64,97	14 64,03

$df = 1, \alpha = 0,05, \chi^{2*} = 3,841$

konfiguriere Test auf Unabhängigkeit von (Besserung erfahren) und (Medikament erhalten):
 $\alpha = 0.05$ 1-seitig

▼ Testergebnisse an CODAP übermitteln
 vorliegender Test



Der t-Test ■ Mit CODAP

Ähnliches Bild, wenn ich nur Personen zwischen ca. 35 und ca. 50 Jahren auswähle

Medikament erhalten

Zielvariable/erstes Merkmal abc

Besserung erfahren

Prädiktor/zweites Merkmal abc

Test auf Unabhängigkeit von Besserung erfahren und Medikament erhalten

Sind (Besserung erfahren) und (Medikament erhalten) unabhängig?
 N = 401, 2 columns by 2 rows, $\chi^2 = 97,5$, $P < 0.0001$

▼ Test auf Unabhängigkeit, 1-seitig χ^2 Zusammenfassung

beobachtet erwartet	Besserung erfahren =	gesund	krank
Medikament erhalten =	Placebo	195 153,1	112 153,9
	Medikament	5 46,88	89 47,12

df = 1, $\alpha = 0,05$, $\chi^{2*} = 3,841$

konfiguriere Test auf Unabhängigkeit von (Besserung erfahren) und (Medikament erhalten):
 $\alpha =$

▼ Testergebnisse an CODAP übermitteln

vorliegender Test

Der t-Test ▪ Mit CODAP

- Das Entdecken möglicher **weiterer** Unterschiede, Abhängigkeiten oder Zusammenhänge und Herausarbeiten des Einflusses möglicher weiterer Variablen sollte gefördert werden.
- Befähigen Sie Ihre Schüler:innen dazu, **möglichst viele Fragen an den Datensatz zu stellen** und auch weiterführende Fragestellungen zu bereits erzielten Erkenntnissen.





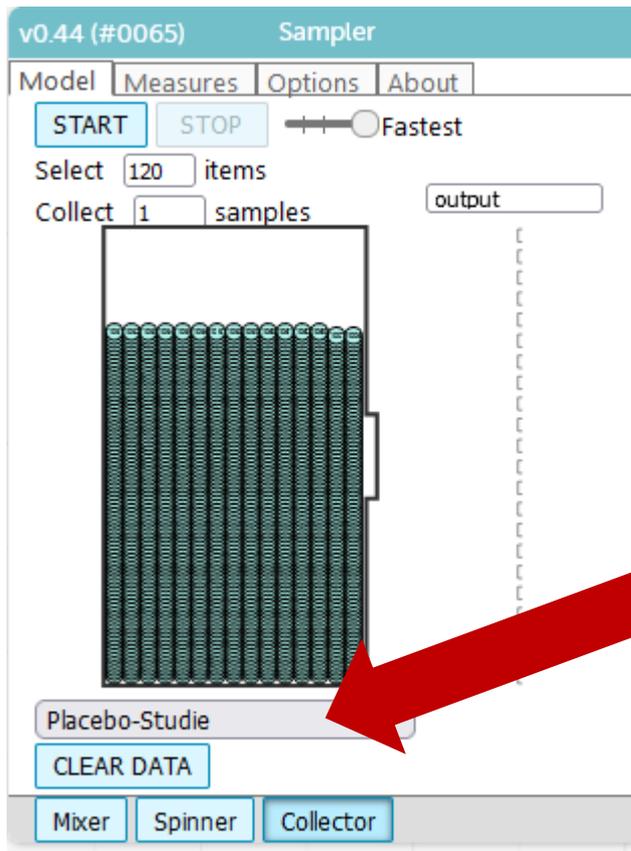
Signifikanztests reagieren sensibel auf die Stichprobengröße

Wir ziehen daher Stichproben mithilfe des Plugins „Sampler“ aus unserem Datensatz und beobachten, wie sich die p-Werte verändern:

The screenshot shows a software interface with a top bar containing a menu icon, the word "Test", and a red button labeled "UNGESPEICHERT". Below the top bar is a toolbar with icons for "Tabellen", "Graph", "Karte", "Regler", "Rechner", "Text", and "Plugins". The main content area displays a text block starting with "In einer Studie mit 1363 Versuchspersonen wurde ein neues Medikament gegen ein Placebo getestet. Unten siehst du die Daten, die dabei entstanden sind." Below the text is an image of various pills and capsules. A red arrow points to the "Sampler" plugin in a dropdown menu, which also includes "Mischmasch", "Testimate", "Draw Tool", "Microdata Portal", "NOAA Weather", "Choosy", "Transformers", "Sonify", and "Story Builder". A tooltip for the "Sampler" plugin reads "Create random datasets from scratch or existing datasets." The text block also contains two sub-questions: "a) Untersuche mit passenden Diagrammen und Kennlinien, ob das neu entwickelte Medikament besser ist als das Placebo." and "b) Führe einen passenden Hypothesentest durch, um die Wirksamkeit des Medikaments gegenüber des Placebo zu untersuchen."



Signifikanztests reagieren sensibel auf die Stichprobengröße



Klicken Sie unten auf Collector und wählen Sie den Datensatz Placebo-Studie. Wir wollen nun 1 Stichprobe mit 120 Fällen ziehen. Geben Sie diese beiden Zahlen oben ein und klicken Sie auf Start.



Signifikanztests reagieren sensibel auf die Stichprobengröße

Nun entsteht ein neuer Datensatz, der sich nur auf die 120 zufällig gezogenen Fälle bezieht:

samples (1 Fälle)		items (120 Fälle)					
sample	output	ID der Person	Ge...=divers)	Medika...rhalten	Besse...rfahren	Alter	Ge-wicht
18		1000	m	Placebo	krank	44	90
		818	m	Placebo	gesund	49	75
		1165	m	Placebo	krank	44	90
		183	m	Medika...	gesund	24	80
		116	w	Medika...	gesund	26	76
		812	w	Placebo	gesund	27	65
		353	m	Medika...	krank	42	99
		1024	w	Placebo	gesund	25	65
		203	m	Medika...	gesund	25	80
		1043	w	Placebo	gesund	49	65
		980	w	Placebo	gesund	45	76
		396	w	Medika...	krank	41	60
		1192	w	Placebo	krank	55	83
		188	m	Medika...	gesund	23	67



Signifikanztests reagieren sensibel auf die Stichprobengröße

Mit diesem Datensatz kann nun erneut ein Test durchgeführt werden. Öffnen Sie hierzu ein neues „Testimate“ und ziehen Sie nun aus dem neuen kleineren Datensatz die beiden entsprechenden Merkmale auf „Testimate“ und wählen Sie den zugehörigen Test aus:

2024i
testimate

Datensatz: **experiments/samples/items**, 120 Fälle (ist nicht zufällig) i

Medikament erhalten

Zielvariable/erstes Merkmal abc

Besserung erfahren

Prädiktor/zweites Merkmal abc

Test auf Unabhängigkeit von Besserung erfahren und Medikament erhalten

Sind (Besserung erfahren) und (Medikament erhalten) unabhängig?
 N = 120, 2 columns by 2 rows, $\chi^2 = 0,1875$, P = 0,665

▼ Test auf Unabhängigkeit, 1-seitig χ^2 Zusammenfassung

beobachtet erwartet	Besserung erfahren =	krank	gesund
Medikament erhalten =	Medikament	15 16,1	27 25,9
	Placebo	31 29,9	47 48,1

df = 1, $\alpha = 0,05$, $\chi^{2*} = 3,841$



Signifikanztests reagieren sensibel auf die Stichprobengröße

Klicken Sie unten auf „senden“, um die Testergebnisse an CODAP zu übermitteln:

Testergebnisse und Schätzer								
Testergebnisse und Schätzer (2 Fälle)								
Index	Ergebnis	Prädiktor	Prozedere	p	Chi ²	Chi ² krit	N	df
1	Medika...	Besseru...	indepe...	0,665	0,1875	3,8415	120	
2	Medika...	Besseru...	indepe...	0,1709	1,8747	3,8415	120	



Signifikanztests reagieren sensibel auf die Stichprobengröße

Indem man bei „experiments/samples/items“ alle Daten löscht und immer neue Stichproben generiert, ändern sich die Testergebnisse, die man alle sammeln kann:

Testergebnisse und Schätzer										
Testergebnisse und Schätzer (7 Fälle)										
Index	Ergebnis	Prädiktor	Prozedere	p	Chi ²	Chi ² krit	N	df	alpha	Sei
1	Medika...	Besseru...	indepe...	0,665	0,1875	3,8415	120	1	0,05	
2	Medika...	Besseru...	indepe...	0,1709	1,8747	3,8415	120	1	0,05	
3	Medika...	Besseru...	indepe...	0,1174	2,4519	3,8415	120	1	0,05	
4	Medika...	Besseru...	indepe...	0,9612	0,0024	3,8415	120	1	0,05	
5	Medika...	Besseru...	indepe...	0,0025	9,1464	3,8415	240	1	0,05	
6	Medika...	Besseru...	indepe...	0,0041	8,2587	3,8415	240	1	0,05	
7	Medika...	Besseru...	indepe...	0,4998	0,4554	3,8415	240	1	0,05	

1113 r

630 r

482 r



Signifikanztests reagieren sensibel auf die Stichprobengröße

Testergebnisse und Schätzer										
Testergebnisse und Schätzer (7 Fälle)										
Index	Ergebnis	Prädiktor	Prozedere	p	Chi ²	Chi ² krit	N	df	alpha	Se
1	Medika...	Besseru...	indepe...	0,665	0,1875	3,8415	120	1	0,05	
2	Medika...	Besseru...	indepe...	0,1709	1,8747	3,8415	120	1	0,05	
3	Medika...	Besseru...	indepe...	0,1174	2,4519	3,8415	120	1	0,05	
4	Medika...	Besseru...	indepe...	0,9612	0,0024	3,8415	120	1	0,05	
5	Medika...	Besseru...	indepe...	0,0025	9,1464	3,8415	240	1	0,05	
6	Medika...	Besseru...	indepe...	0,0041	8,2587	3,8415	240	1	0,05	
7	Medika...	Besseru...	indepe...	0,4998	0,4554	3,8415	240	1	0,05	

Bei den 4 Stichproben mit 120 Personen habe ich in meiner Simulation NIE ein signifikantes Ergebnis erhalten (auf dem 5%-Niveau). Bei einer Stichprobengröße von 240 Personen oder gar 600 Personen wird es aber wahrscheinlicher, ein signifikantes Ergebnis zu erhalten.



Signifikanztests reagieren sensibel auf die Stichprobengröße



Bei hinreichend **großen Stichproben** werden auch **winzig kleine Effekte** signifikant. Das bietet eine Möglichkeit der Manipulation!

Denn „signifikante Effekte“ werden oft mit „großen Effekten“ verwechselt.



Abhängigkeit der Signifikanztests von der Stichprobengröße

Chi-Quadrat-Test ▪ **N = 100**

	Besserung	Keine Besserung	
Medikament	23	17	40
Placebo	27	33	60
	50	50	100

$p = 0,22$

Keine signifikante Abhängigkeit

Chi-Quadrat-Test ▪ **N = 1.000**

	Besserung	Keine Besserung	
Medikament	230	170	400
Placebo	270	330	600
	500	500	1.000

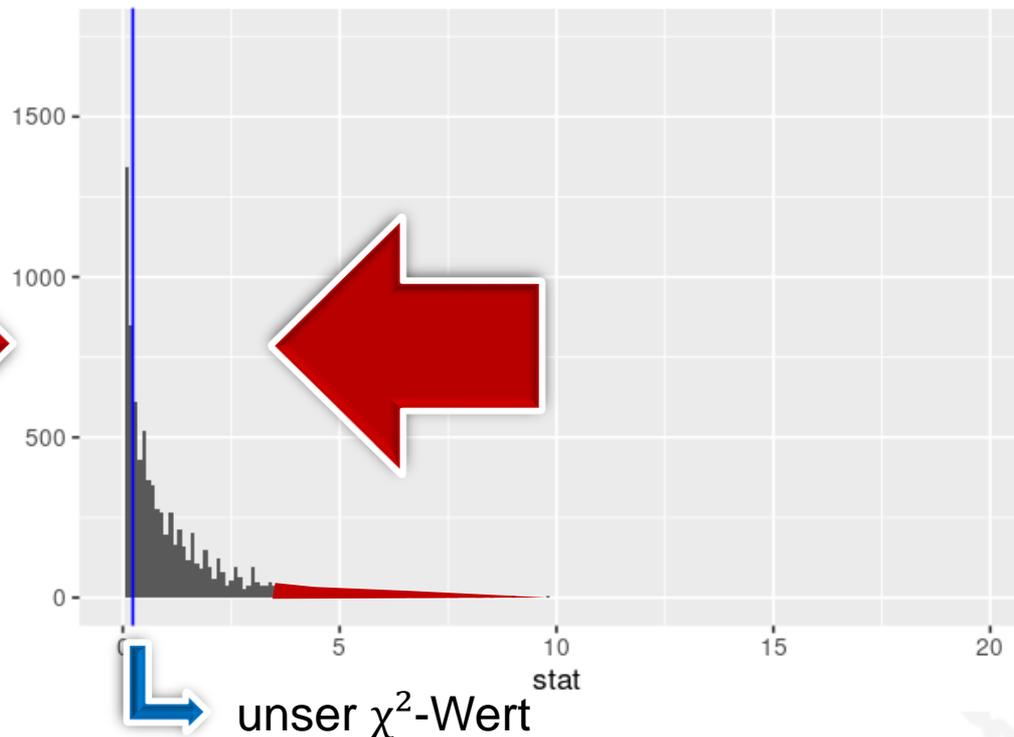
$p = 0,0001$

Signifikante Abhängigkeit

Daher ist es auch nicht verwunderlich, dass im Mikrozensus-Datensatz mit sehr vielen Fällen auch viele Tests signifikante Ergebnisse liefern!

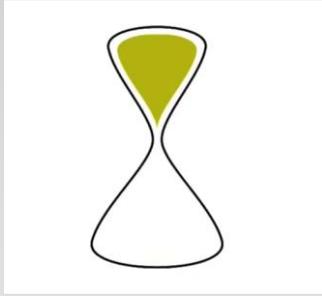
Warum ist der p-Wert von der Stichprobengröße abhängig?

Mit größeren Stichproben wird der wahre Wert in der Grundgesamtheit immer besser geschätzt! Die Stichprobenvarianz sinkt.



Die Verteilung der Teststatistikwerte wird also schmaler, so dass der p-Wert immer kleiner wird, weil der Teststatistikwert immer mehr in den „unplausiblen“ Bereich rutscht.

(Eigentlich verändert sich nicht der Teststatistikwert, sondern die Kurve dahinter)



Workshop-Phase

Bearbeiten Sie bitte
Übungsblatt 4

Zeit: 10 Minuten





Mögliche Lösung:

2024i testimate

Datensatz: **Mikrozensus_2010**, 5000 Fälle (ist nicht zufällig) i

Wohnfläche

Zielvariable/erstes Merkmal 123

Ostdeutschland_Westdeutschland

Prädiktor/zweites Merkmal abc

Differenz der Mittelwerte Wohnfläche gruppiert durch Ostdeutschland_Westdeutschland ▼

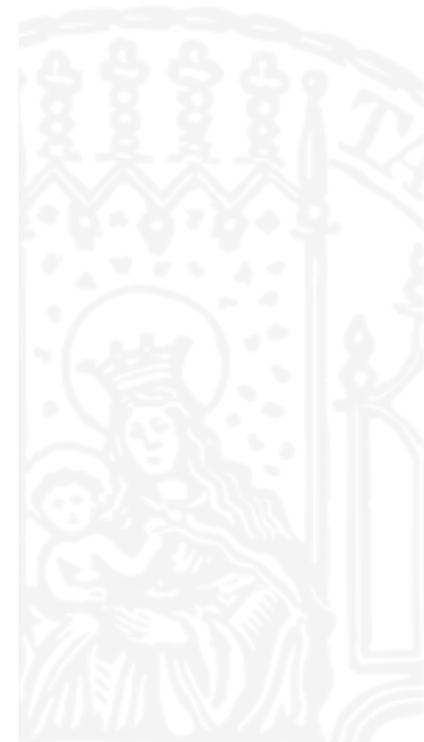
Ist die Differenz arithmetisches Mittel (Wohnfläche) : Gruppe West - Gruppe Ost $\neq 0$?

N = 4945, t = 12,8, P < 0.0001
diff = 20,7, 95% KI = [17,49, 23,83]

▼ Differenz der Mittelwerte, 2-Stichprobenfall t Verfahren

Ostdeutschland_Westdeutschland	N	arithmetisches Mittel(Wohnfläche)	s	SE
West	3914	106,7	48,34	0,7726
Ost	1031	86,04	36,91	1,15
pooled	4945	diff = 20,7	46,19	1,617

df = 4940, $\alpha = 0,05$, t* = 1,96





Mögliche Lösung:

2024i testimate

Datensatz: **Mikrozensus_2010**, 244 Fälle (ist nicht zufällig)

Wohnfläche
Zielvariable/erstes Merkmal 123

Ostdeutschland_Westdeutschland
Prädiktor/zweites Merkmal abc

Differenz der Mittelwerte Wohnfläche gruppiert durch Ostdeutschland_Westdeutschland

Ist die Differenz arithmetisches Mittel (Wohnfläche) : Gruppe West - Gruppe Ost \neq 0 ?

N = 244, t = 3,09, P = 0,002234
diff = 20,7, 95% KI = [7,511, 33,92]

▼ Differenz der Mittelwerte, 2-Stichprobenfall t Verfahren

Ostdeutschland_Westdeutschland	N	arithmetisches Mittel(Wohnfläche)	s	SE
West	192	112	44,73	3,228
Ost	52	91,33	35,07	4,864
pooled	244	diff = 20,7	42,88	6,703

df = 242, α = 0,05, t* = 1,97

Test auf Mittelwert(Wohnfläche):
[West] - [Ost] \neq 0
conf = 95 %

▼ Testergebnisse an CODAP übermitteln
 vorliegender Test



Mögliche Lösung:

Je größer die Stichprobe, desto wahrscheinlicher werden auch kleine Effekte signifikant.

Bei sehr großen Stichproben werden die p-Werte „automatisch“ kleiner.

Daher ist es ratsam, nur eine Stichprobe aus dem großen Datensatz zu betrachten.

Daher gilt bei Signifikanztests auch nicht: Je größer die Stichprobe, desto besser! Statistiker berechnen hier zuvor (mithilfe sogenannter Power-Analysen), wie groß die Stichprobe maximal sein darf, um Effekte bestimmter Größen zeigen zu können.

PAUSE ☺



Nochmal genauer: Der p-Wert



Was bedeutet ein signifikantes Testergebnis?

In der Realität bestimmt man mithilfe einer geeigneten Software und dem passenden Testverfahren den **p-Wert** und prüft, ob dieser kleiner als eine vorgegebene Grenze ist (meist 5%; = 😊) oder größer als die vorgegebene Grenze ist (😞).

Der **p-Wert** gibt die Wahrscheinlichkeit an, diese oder noch extremere Daten (z.B. beim t-Test mit einem noch größeren Gruppenunterschied) zu finden, obwohl sich die beiden Gruppen in Wirklichkeit nicht unterscheiden.

$P(\text{Dieser oder noch extremerer Teststatistikwert} \mid \text{Nullhypothese richtig}) \leq 5\%$

**Stärkere Fokussierung auf den p-Wert vorteilhaft
(Realitätsbezug und Wissenschaftspropädeutik)!**
(Haller & Krauss, 2002; Ufer, 2022)

Wiederholung: Unterschiedliche Arten von Forschungsfragen und Tests

Unterschieds- hypothesen

Beispiel:
Das Einkommen unterscheidet sich je nach Geschlecht

⇒ **t-Test**

Zusammenhangs- hypothesen

Beispiel:
Durchschnitts-
temperaturen hängen
mit dem CO₂-Ausstoß
zusammen

⇒ **Test auf Signifikanz von
Korrelationskoeffizienten**

Abhängigkeits- hypothesen

Beispiel:
Die Genesungsrate hängt von der Gruppenzugehörigkeit (Medikamentengruppe vs. Placebogruppe) ab.

⇒ **Chi-Quadrat-
Unabhängigkeitstest**

Ein p-Wert von höchstens 5% zeigt dabei ein „signifikantes“ (bzw. „überzufälliges“) Ergebnis an.

Der p-Wert gibt die Wahrscheinlichkeit an, diese oder noch extremere Daten (z.B. beim t-Test: mit einem noch größeren Unterschied) zu finden, obwohl sich die beiden Gruppen in Wirklichkeit nicht unterscheiden.

Was bedeutet ein signifikantes Testergebnis nicht?

$P(\text{Dieser oder noch extremerer Teststatistikwert} \mid \text{Nullhypothese richtig}) \leq 5\%$

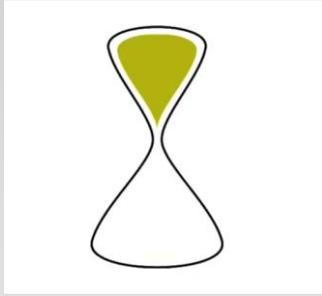
Der p-Wert gibt uns jedoch keine Auskunft darüber, wie wahrscheinlich die Nullhypothese ist, also $P(\text{Nullhypothese richtig})$. Sie gibt uns auch keine Auskunft darüber, wie wahrscheinlich die Nullhypothese ist, unter der Bedingung, dass wir die Daten so erhalten, wie wir sie erhalten haben $P(\text{Nullhypothese richtig} \mid \text{Daten})$.

Formel von Bayes

$$P(\text{Daten} \mid H_0) = \frac{P(H_0 \mid \text{Daten}) \cdot P(\text{Daten})}{P(H_0)}$$



Es gibt innerhalb der Statistik einen eigenen Zweig, die sogenannte „Bayessche Statistik“, die sich mit einem solchen Ansatz beschäftigt.



Workshop-Phase

Bearbeiten Sie
bitte Übung 5

Zeit: 7 Minuten



Der **p-Wert** gibt die Wahrscheinlichkeit an, diese oder noch extremere Daten (z.B. beim t-Test mit einem noch größeren Gruppenunterschied) zu finden, obwohl sich die beiden Gruppen in Wirklichkeit nicht unterscheiden.

Was bedeutet ein signifikantes Testergebnis ($p = 0,05$)?

Wahr oder falsch?

1. Es ist eindeutig bewiesen, dass die Nullhypothese (die beiden Gruppen unterscheiden sich nicht) falsch ist.
2. Die Wahrscheinlichkeit des Zutreffens der Nullhypothese ist gefunden worden.
3. Es ist eindeutig bewiesen, dass die Alternativhypothese (dass sich die beiden Gruppen unterscheiden) wahr ist.
4. Man kann nun die Wahrscheinlichkeit ableiten, dass die Alternativhypothese richtig ist.
5. Entscheidet man sich nun, die Nullhypothese zu verwerfen, dann weiß man jetzt die Wahrscheinlichkeit, dass diese Entscheidung falsch sein könnte.
6. Der experimentelle Befund ist reliabel in dem Sinne, dass man in 95% der Fälle ein signifikantes Ergebnis bekäme, wenn man das Experiment sehr oft wiederholen würde



Ist die folgende Aussage wahr oder falsch?

1. Es ist eindeutig bewiesen, dass die Nullhypothese (die beiden Gruppen unterscheiden sich nicht) falsch ist.



Zur Erinnerung:

Der **p-Wert** gibt die Wahrscheinlichkeit an, diese oder noch extremere Daten (z.B. beim t-Test mit einem noch größeren Gruppenunterschied) zu finden, obwohl sich die beiden Gruppen in Wirklichkeit nicht unterscheiden.

Die Interpretation ist falsch.

- Mit empirischen Studien und Signifikanztests gibt es keine „eindeutigen Beweise“.
- Außerdem wissen wir nichts über die Wahrscheinlichkeit, dass die Nullhypothese eintritt oder nicht.

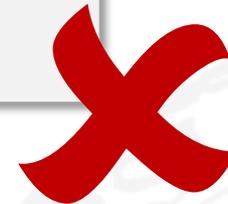
Der p-Wert

(Haller & Krauss, 2002; Oakes, 1986)



Ist die folgende Aussage wahr oder falsch?

2. Die Wahrscheinlichkeit des Zutreffens der Nullhypothese ist gefunden worden.



Zur Erinnerung:



Der **p-Wert** gibt die Wahrscheinlichkeit an, diese oder noch extremere Daten (z.B. beim t-Test mit einem noch größeren Gruppenunterschied) zu finden, obwohl sich die beiden Gruppen in Wirklichkeit nicht unterscheiden.

Die Interpretation ist falsch.

- Wir wissen nichts über die Wahrscheinlichkeit, dass die Nullhypothese eintritt oder nicht. (Die Nullhypothese steht in Bedingung.)



Ist die folgende Aussage wahr oder falsch?

3. Es ist eindeutig bewiesen, dass die Alternativhypothese (dass sich die beiden Gruppen unterscheiden) wahr ist.



Zur Erinnerung:



Der **p-Wert** gibt die Wahrscheinlichkeit an, diese oder noch extremere Daten (z.B. beim t-Test mit einem noch größeren Gruppenunterschied) zu finden, obwohl sich die beiden Gruppen in Wirklichkeit nicht unterscheiden.

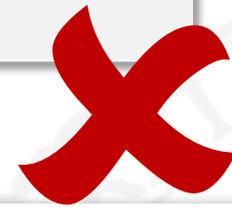
Die Interpretation ist falsch.

- Mit empirischen Studien und Signifikanztests gibt es keine „eindeutigen Beweise“.
- Außerdem wissen wir nichts über die Wahrscheinlichkeit, dass die Alternativhypothese eintritt oder nicht.



Ist die folgende Aussage wahr oder falsch?

4. Man kann nun die Wahrscheinlichkeit ableiten, dass die Alternativhypothese richtig ist.



Zur Erinnerung:



Der **p-Wert** gibt die Wahrscheinlichkeit an, diese oder noch extremere Daten (z.B. beim t-Test mit einem noch größeren Gruppenunterschied) zu finden, obwohl sich die beiden Gruppen in Wirklichkeit nicht unterscheiden.

Die Interpretation ist falsch.

- Wir wissen nichts über die Wahrscheinlichkeit, dass die Alternativhypothese eintritt oder nicht – auch nicht unter der Bedingung, dass wir einen bestimmten Datensatz vorliegen haben.

Der p-Wert

(Haller & Krauss, 2002; Oakes, 1986)



Ist die folgende Aussage wahr oder falsch?

5. Entscheidet man sich nun, die Nullhypothese zu verwerfen, dann weiß man jetzt die Wahrscheinlichkeit, dass diese Entscheidung falsch sein könnte.



Zur Erinnerung:



Der **p-Wert** gibt die Wahrscheinlichkeit an, diese oder noch extremere Daten (z.B. beim t-Test mit einem noch größeren Gruppenunterschied) zu finden, obwohl sich die beiden Gruppen in Wirklichkeit nicht unterscheiden.

Die Interpretation ist falsch.

- Diese Interpretation mag auf den ersten Blick wie der Fehler erster Art wirken, beschreibt aber die zugehörige invertierte bedingte Wahrscheinlichkeit, nämlich lediglich: $P(\text{Nullhypothese wahr} | \text{Entscheidung gegen Nullhypothese})$. Und über die Wahrscheinlichkeit der Nullhypothese können wir keine Aussage treffen.



Ist die folgende Aussage wahr oder falsch?

6. Der experimentelle Befund ist reliabel in dem Sinne, dass man in 95% der Fälle ein signifikantes Ergebnis bekäme, wenn man das Experiment sehr oft wiederholen würde.



Zur Erinnerung:

Der **p-Wert** gibt die Wahrscheinlichkeit an, diese oder noch extremere Daten (z.B. beim t-Test mit einem noch größeren Gruppenunterschied) zu finden, obwohl sich die beiden Gruppen in Wirklichkeit nicht unterscheiden.

Die Interpretation ist falsch.

- Hier wird aus $p=5\%$ eine Aussage über die Replizierbarkeit eines signifikanten Ergebnisses abgeleitet. Ein signifikantes Ergebnis macht aber eine Aussage über $P(\text{Daten}|H_0)$. Jede Aussage, die eine Paraphrasierung des p-Wertes beinhaltet, muss also in irgendeiner Form Stellung zu H_0 beziehen (z.B. „unter der Bedingung, dass H_0 wahr ist.“). Ein solcher Bezug fehlt aber in obiger Aussage.



Befunde zu Verwirrungen mit p-Werten

(HALLER & KRAUSS, 2002)

% angekreuzt

Deutschland 2000 Psychologische
Fachbereiche deutscher Universitäten

USA 1986
(Oakes)

Äußerungen (gekürzt)	Statistik Dozenten (N = 30)	Psychologie Dozenten (N = 39)	Psychologie Studenten (N = 44)	Psychologen
1) H_0 ist widerlegt	10%	15%	34%	1%
2) Wsk der H_0	17%	26%	32%	36%
3) H_1 ist bewiesen	10%	13%	20%	6%
4) Wsk der H_1	33%	33%	59%	66%
5) Wsk des Fehlers 1. Art	73%	67%	68%	86%
6) Wsk einer Replikation	37%	49%	41%	60%
<i>% mindestens einen Fehler</i>	80%	90%	100%	97%
<i># durchschnittlich angekreuzt</i>	1,9	2,0	2,5	



Weitere typische Fehlvorstellung ▪ Nr. 1

(FUNDAMENTE DER MATHEMATIK, 2024)

Mit einer Software kann man nun analysieren, wie gut die beobachteten Daten zu den unter H_0 erwarteten Daten passen, indem man den p-Wert bestimmen lässt: Der p-Wert gibt die Wahrscheinlichkeit an, dass die beobachteten Daten unter der Bedingung, dass die Nullhypothese zutrifft (dass also X und Y unabhängig sind), tatsächlich auftreten. Je kleiner also der p-Wert ist, desto unwahrscheinlicher ist es, dass H_0 wahr ist. Wie bei Signifikanztests wird H_0 abgelehnt, falls der p-Wert unter einem vorher festgelegten Signifikanzniveau (z.B. 5%) liegt, das bedeutet aber nicht automatisch, dass H_0 nicht zutrifft.



Nein, das stimmt nicht! Wir können keine Aussage über die Wahrscheinlichkeit der H_0 treffen.

Schlimmer noch: Wenn die H_0 zutrifft, dann sind die p-Werte sogar gleichverteilt!



Weitere typische Fehlvorstellung ■ Nr. 1

(FUNDAMENTE DER MATHEMATIK, 2024)

Gleichverteilte p-Werte bei gültiger H0:

Neuer Datensatz		
Fälle (301 Fälle)		
In-dex	Medikament	Heilung
296	Medikament	gesund
297	Placebo	krank
298	Medikament	krank
299	Medikament	gesund
300	Placebo	gesund
301	Placebo	gesund

2024i testimate

Datensatz: **Neuer Datensatz**, 301 Fälle (ist nicht zufällig)

Medikament

Zielvariable/erstes Merkmal abc

Heilung

Prädiktor/zweites Merkmal abc

Test auf Unabhängigkeit von Heilung und Medikament

Sind (Heilung) und (Medikament) unabhängig?
 $N = 301$, 2 columns by 2 rows, $\chi^2 = 0,2643$, $P = 0,6072$

► Test auf Unabhängigkeit, 1-seitig χ^2 Zusammenfassung

konfiguriere Test auf Unabhängigkeit von (Heilung) und (Medikament):
 $\alpha = 0,05$ 1-seitig

▼ Testergebnisse an CODAP übermitteln

vorliegender Test

Neue Stichprobe ziehen

100 mal

Testergebnisse und Schätzer



Weitere typische Fehlvorstellung ▪ Nr. 2

(FUNDAMENTE DER MATHEMATIK, 2024)

Mit einer Software kann man nun analysieren, wie gut die beobachteten Daten zu den unter H_0 erwarteten Daten passen, indem man den p-Wert bestimmen lässt: Der p-Wert gibt die Wahrscheinlichkeit an, dass die beobachteten Daten unter der Bedingung, dass die Nullhypothese zutrifft (dass also X und Y unabhängig sind), tatsächlich auftreten. Je kleiner also der p-Wert ist, desto unwahrscheinlicher ist es, dass H_0 wahr ist. Wie bei Signifikanztests wird H_0 abgelehnt, falls der p-Wert unter einem vorher festgelegten Signifikanzniveau (z.B. 5%) liegt, das bedeutet aber nicht automatisch, dass H_0 nicht zutrifft.



Nein, das stimmt nicht! Die Chi-Quadrat-Verteilung ist stetig. Die Wahrscheinlichkeit $P(X=x)$ ist daher immer 0.
Man kann hier nur interpretierbare Wahrscheinlichkeitsaussagen über Intervalle treffen.



Weitere typische Fehlvorstellung ■ Nr. 2

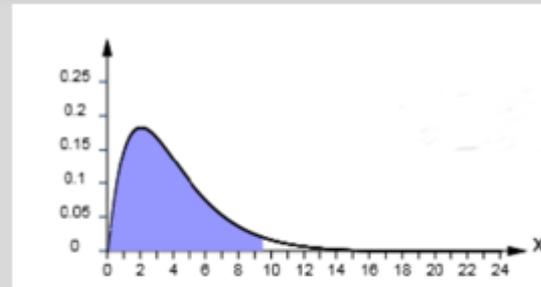
(FUNDAMENTE DER MATHEMATIK, 2024)

Dichtefunktion und Verteilungsfunktion einer stetigen Verteilung

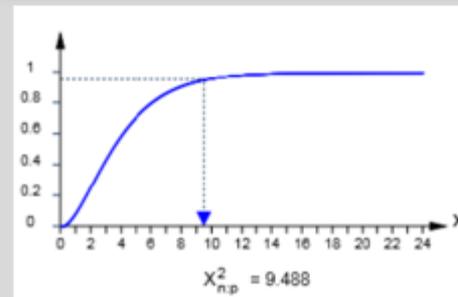
Die Wahrscheinlichkeit, dass X zwischen 0 und 9,5 liegt, lässt sich über den Flächeninhalt berechnen.

Daher die Formulierung beim p-Wert „dieser Wert, oder noch extremere Werte“, weil ohne den Zusatz die Wahrscheinlichkeit „0“ wäre.

Dichtefunktion



Verteilungsfunktion





Daher Vorsicht: Signifikanztests und p-Werte sind umstritten

- Signifikanztests und p-Werte werden in einem großen Teil quantitativer empirischer Studien angewandt.
- Signifikanztests werden jedoch in empirischen Wissenschaften auch äußerst kritisch gesehen, weil es viele Fehlinterpretationen bezüglich des p-Wertes gibt.

Kritikpunkt 1

Das Konzept der Signifikanz ist auch für „Profis“ nur schwer verständlich.

Kritikpunkt 2

Signifikanz ist von der Stichprobengröße abhängig.

Kritikpunkt 3

Oft werden sie mechanisch angewendet, obwohl andere Methoden geeigneter wären.



Zu Kritikpunkt 1: Auch für Profis schwer verständlich

Significance Tests Die Hard

The Amazing Persistence of a Probabilistic Misconception

Ruma Falk and Charles W. Greenbaum
THE HEBREW UNIVERSITY OF JERUSALEM

ABSTRACT. We present a critique showing the flawed logical structure of statistical significance tests. We then attempt to analyze why, in spite of this faulty reasoning, the use of significance tests persists. We identify the illusion of probabilistic proof by contradiction as a central stumbling block, because it is based on a misleading generalization of reasoning from logic to inference under uncertainty. We present new data from a student sample and examples from the psychological literature showing



Zu Kritikpunkt 2: Abhängigkeit von der Stichprobengröße

Chi-Quadrat-Test ▪ **N = 100**

	Besserung	Keine Besserung	
Medikament	23	17	40
Placebo	27	33	60
	50	50	100

$p = 0,22$

Keine signifikante Abhängigkeit

Chi-Quadrat-Test ▪ **N = 1.000**

	Besserung	Keine Besserung	
Medikament	230	170	400
Placebo	270	330	600
	500	500	1.000

$p = 0,0001$

Signifikante Abhängigkeit

Daher ist es auch nicht verwunderlich, dass im Mikrozensus-Datensatz mit sehr vielen Fällen auch viele Tests signifikante Ergebnisse liefern!



Zu Kritikpunkt 3: Alternative Methoden

Alternative Methoden zu Signifikanztests wären beispielsweise:

Effektstärken

Effektstärken beschreiben das Ausmaß (z.B. klein, mittel, groß) eines empirischen Effekts und sind unabhängig von der Stichprobengröße.

Konfidenzintervalle

Ein Konfidenzintervall zum Konfidenzniveau 95% gibt an: Mit 95%-iger Wahrscheinlichkeit überdeckt das angegebene Intervall den geschätzten Parameter (teils mit ähnlichen Fehlvorstellungen verbunden wie p-Werte).

Bayes-Statistik

Neben den Signifikanztests aus der klassischen Statistik, gibt es auch die sogenannte „Bayessche Statistik“, die den umgekehrten Weg geht.



Nicht mehr im Lehrplan,
aber aufschlussreich!

Der χ^2 -Test ■ Effektstärke

Eine Alternative zu Signifikanztests bieten sogenannte Effektstärken.

Aus dem berechneten **Teststatistikwert** χ^2 kann man eine Effektstärke – den **korrigierten Kontingenzkoeffizienten** – mit Hilfe dieser Formel berechnen:

$$C_{\text{korr}} = \sqrt{\frac{k}{k-1}} \cdot C = \sqrt{\frac{k}{k-1}} \cdot \sqrt{\frac{\chi^2}{n + \chi^2}}$$

n: Stichprobengröße, k: Minimum (Anzahl Zeilen; Anzahl Spalten)



Nicht mehr im Lehrplan,
aber aufschlussreich!

Der χ^2 -Test ▪ Effektstärke

Interpretation des **korrigierten Kontingenzkoeffizienten**:

$$C_{\text{korrr}} = \sqrt{\frac{k}{k-1}} \cdot C = \sqrt{\frac{k}{k-1}} \cdot \sqrt{\frac{\chi^2}{n + \chi^2}}$$

Der Wert liegt zwischen 0 und 1.

Werte nahe an 0 bedeuten einen kleinen bzw. keinen Effekt. Je näher der Wert an 1 geht, desto größer ist der Effekt.

Häufig vorgeschlagene Interpretation:

0,1 **kleiner Effekt**
0,3 **mittlerer Effekt**
0,5 **großer Effekt**



Nicht mehr im Lehrplan,
aber aufschlussreich!

Der χ^2 -Test ▪ Effektstärke

Beispiel für einen **korrigierten Kontingenzkoeffizienten**:

$$C_{\text{korr}} = \sqrt{\frac{k}{k-1}} \cdot C = \sqrt{\frac{k}{k-1}} \cdot \sqrt{\frac{\chi^2}{n + \chi^2}}$$

	Besserung	Keine Besserung	
Medikament	302	133	435
Placebo	591	337	928
	893	470	1363

$$\chi^2 = \frac{(302 - 285)^2}{285} + \frac{(133 - 150)^2}{150} + \frac{(591 - 608)^2}{608} + \frac{(337 - 320)^2}{320} = 4,32$$

In unserem Beispiel der Medikamentenstudie gilt: $n=1363$, $k=2$, $\text{Chi}^2 = 4,32$

Also $C_{\text{korr}} = 0,08$. Wir haben also nur einen (wenn überhaupt) sehr kleinen Effekt.



Zwischenfazit

Das sollten Sie im Unterricht unbedingt herausarbeiten

- Unterrichten Sie auch die **Kontroversen** von Signifikanztests
- Unterrichten Sie auch die Abhängigkeit der **Stichprobengröße**
- Kontrastieren Sie die Vorgehensweise mit der **Bayes-Formel** und stellen Sie Signifikanztest und Bayes-Statistik gegenüber.
- Unterrichten Sie die **Alternativen** (z.B. Effektstärken)

Das sollten Sie im Unterricht vermeiden

- **Kochrezeptartiges Durchführen** des Tests mit Berechnung des Teststatistikwerts und des Ablehnungsbereichs. **Besser:** Das **grundlegende Konzept** (und auch die Fehlvorstellungen hinter den p-Werten) unterrichten.
- Fokussierung auf **ausgedachte Beispiele**. **Besser:** authentische Fragestellungen aus der realen Welt

Korrelation und Regression



Korrelation

Das wichtigste Zusammenhangsmaß zweier Merkmale ist die Korrelation. Dazu wird für zwei metrische Merkmale der **(Bravais-Pearson-) Korrelationskoeffizient r** berechnet:

$$r_{xy} := \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i \quad \text{mit standardisierten } x_i = \frac{X_i - \bar{X}}{s_x} \text{ und } y_i \text{ (} X_i: \text{Merkmalsausprägung, } \bar{X}: \text{arithmetisches Mittel; } s_x: \text{Standardabweichung von X).}$$

Der Korrelationskoeffizient r kann als Effektstärke interpretiert werden.

Oft nimmt man dabei folgende Faustregeln an:

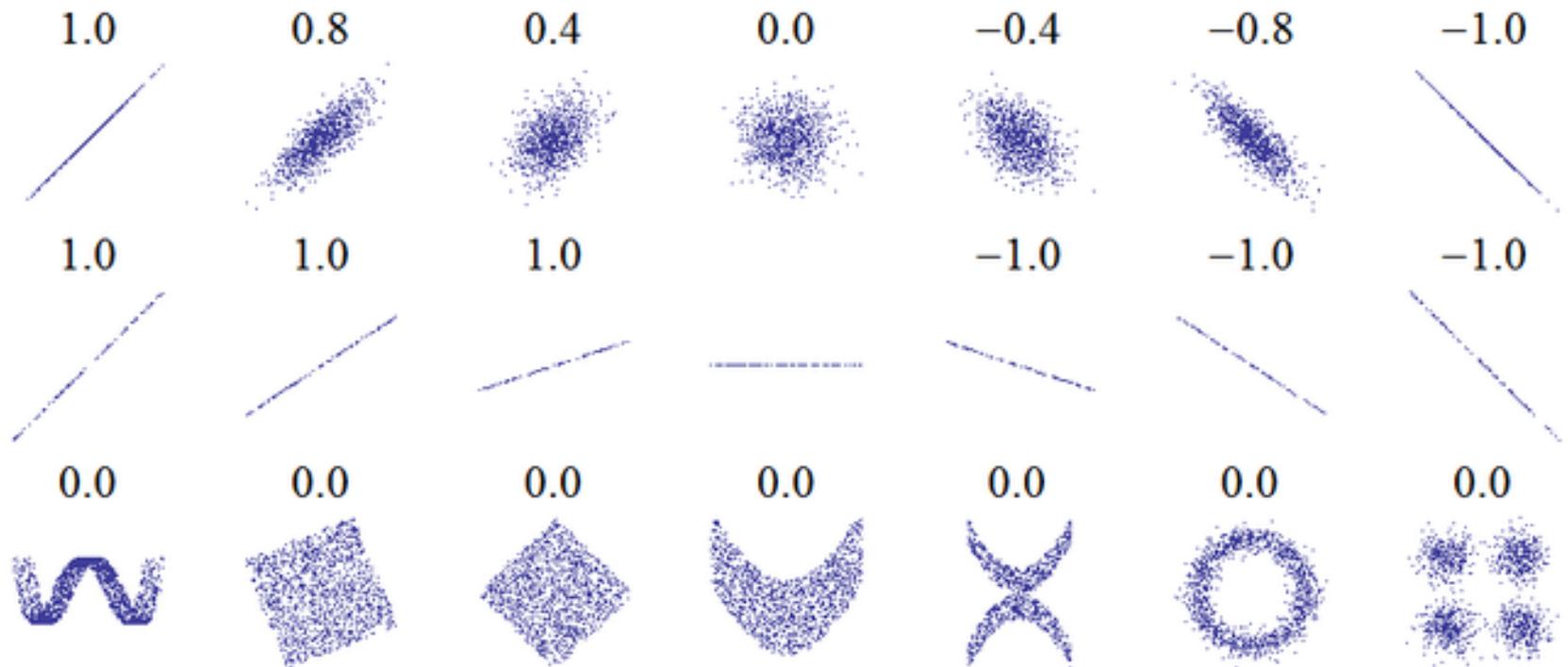
- $r = (\pm) 0,1$: kleiner Effekt
- $r = (\pm) 0,3$: mittlerer Effekt
- $r = (\pm) 0,5$: großer Effekt

r kann dabei Werte zwischen -1 und 1 annehmen.

Mit einem Signifikanztest kann überprüft werden, ob sich r signifikant von 0 unterscheidet.

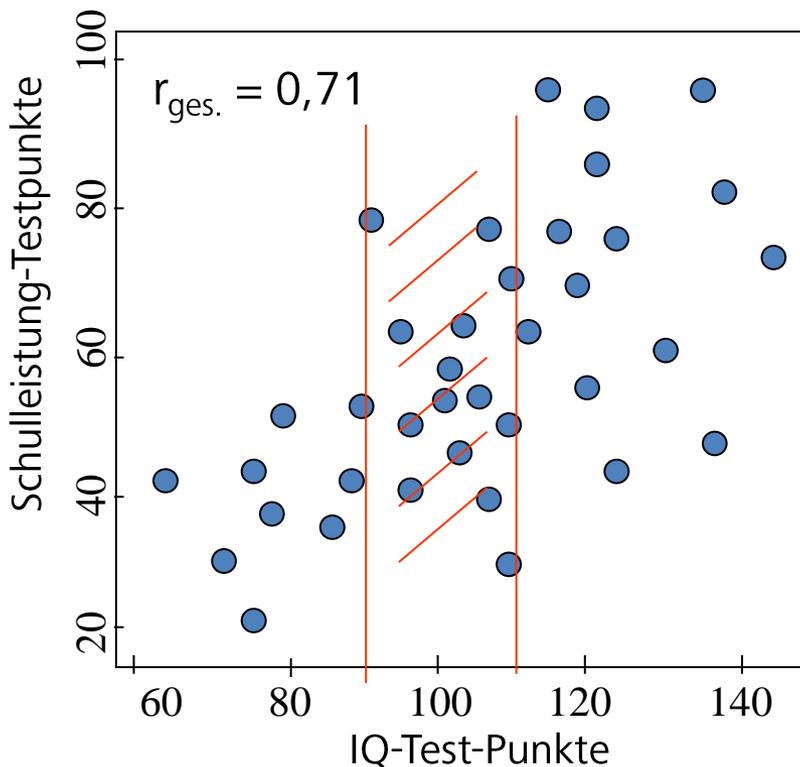
Korrelation

Maß für den linearen Zusammenhang – einige Beispiele für verschiedene Korrelationskoeffizienten (Wikipedia):



Korrelation

Durch eine **ungünstige Selektion der Stichprobe** kann die Korrelation größer oder kleiner ausfallen, als sie in der Realität zugrundliegt.



Werden Datenpunkte „in der Mitte“ unterschlagen, so wird der Korrelationskoeffizient **größer**.

Umgekehrt wird der Korrelationskoeffizient **kleiner**, wenn Datenpunkte am Rand unbeachtet bleiben.



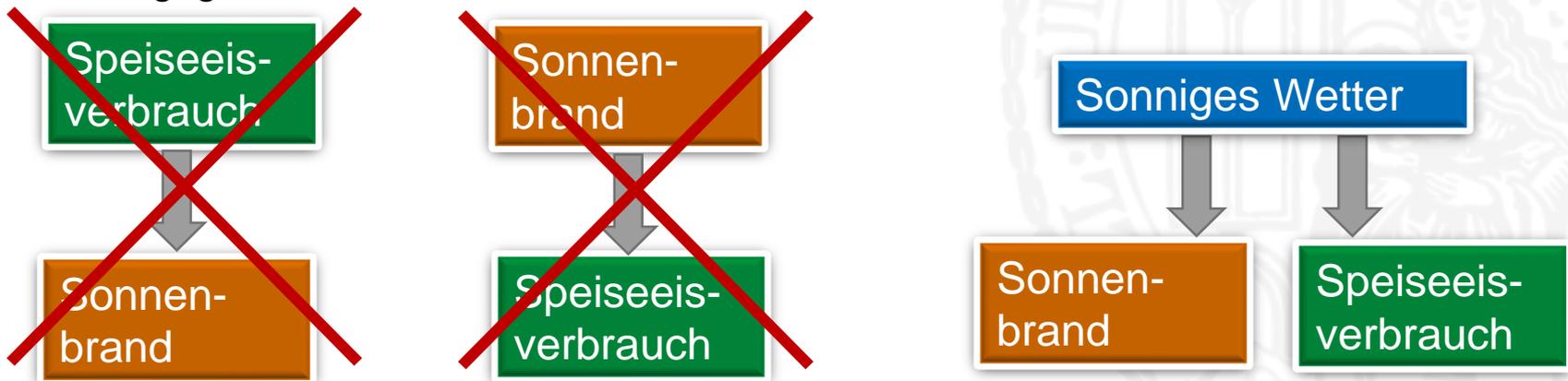
Korrelation vs. Kausalität

Ein Spezialfall des Zusammenhangs zweier Merkmale ist die **Kausalität** (→ eines der Merkmale ist die Ursache für das andere Merkmal); diese lässt sich aber nur schwer durch Rechnung belegen, sondern wird meist theoretisch begründet. Kausalität kann nur bei Korrelation vorliegen.

Aus Korrelation folgt noch keine Kausalität.

Beispiel: Es gibt einen Zusammenhang zwischen **Speiseeisverbrauch** und der Anzahl von **Sonnenbränden**.

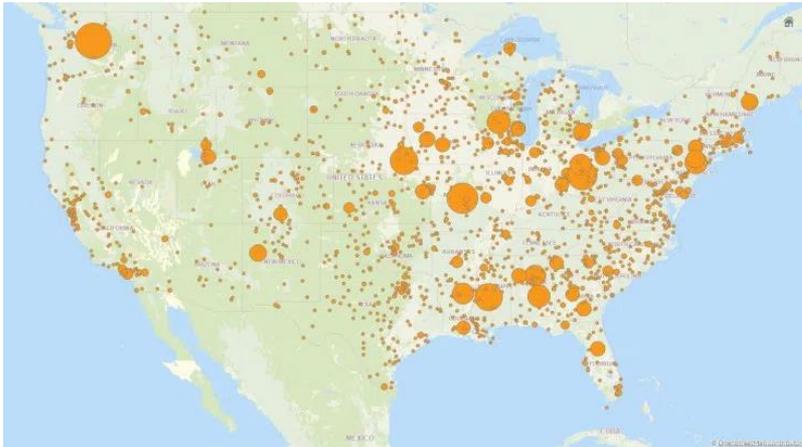
Jedoch ist der Speiseeisverbrauch nicht die Ursache von Sonnenbrände. Und auch Sonnenbrände führen nicht kausal zu einem höheren Eiskonsum. Eine dritte Variable ist hingegen die Ursache für die beiden anderen.



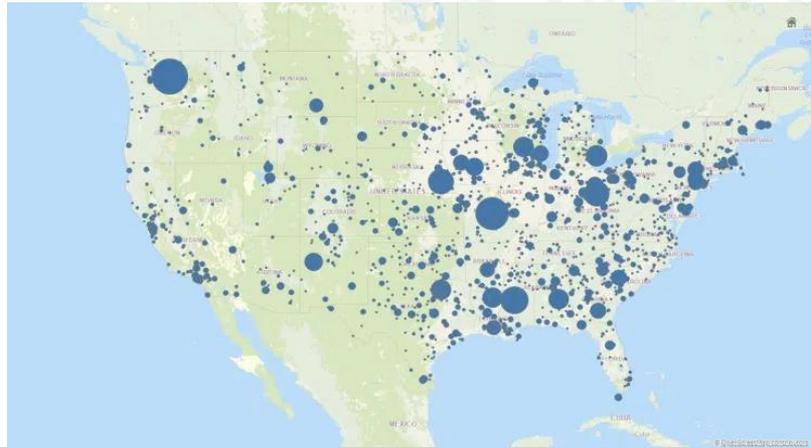
Korrelation vs. Kausalität

Heinen (2021)

Corona-Fälle 2020



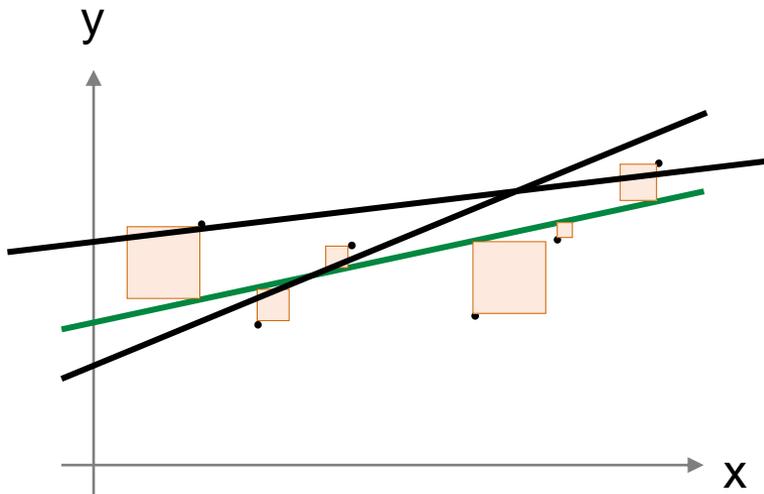
G5-Masten in den USA



Hintergrundvariable: Bevölkerungsdichte

Regression

- Mithilfe einer **Regressionsanalyse** kann (unter bestimmten Voraussetzungen) die Beziehung einer abhängigen (bzw. erklärten Variable) und einer (oder mehrerer) unabhängiger (bzw. erklärender) Variablen modelliert werden.
- Wir konzentrieren uns hier auf den Spezialfall der **linearen Einfachregression** (heißt: linearer Zusammenhang und nur eine erklärende Variable).
- Häufig wird hier die **Methode der kleinsten Quadrate** angewandt:



Idee: Welche der vielen denkbaren Geraden beschreibt den linearen Zusammenhang zwischen x und y am besten?

⇒ Wähle **die Gerade**, für die die **Summe der quadratischen Abweichungen minimal** ist.

Regression

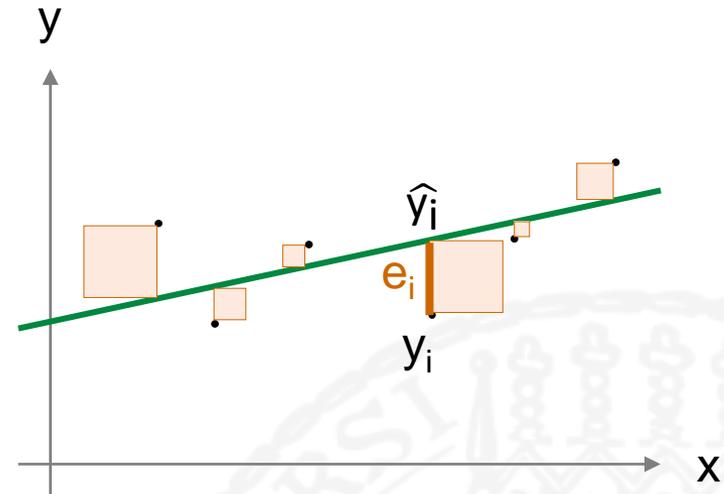
Ein erster Überblick

Regression

$$e_i = y_i - \hat{y}_i$$

y_i : tatsächlich gemessener Wert der Person i
mit der Ausprägung x_i

\hat{y}_i : durch das Modell prognostizierter Wert für
Person i



Die **Gleichung der Regressionsgeraden** lautet dann

$$\hat{y}_i = a + b \cdot x_i$$

$$\text{mit } a = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$\text{und } b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

Wir interessieren uns bei Regressionsanalysen für die **Steigung b** , denn mithilfe dieses Parameters können wir vorhersagen, um welchen Wert sich y ändert, wenn sich x ändert.

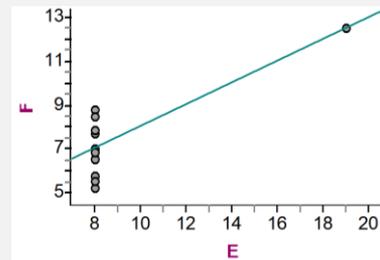
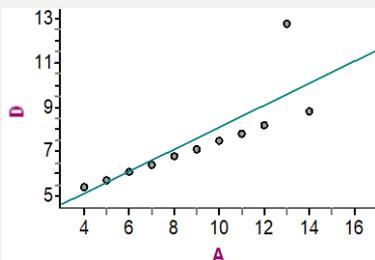
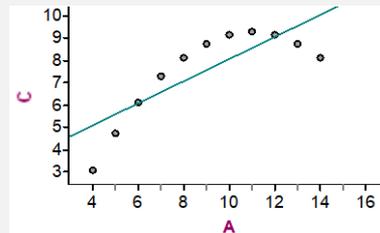
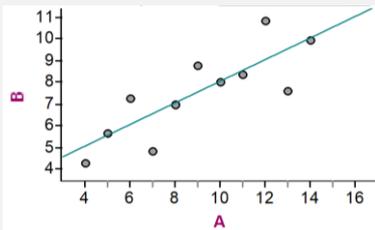
Beispiel: Um wieviel besser ist die erwartete Mathematikleistung von Schüler:innen, die im Schnitt 10 IQ-Punkte mehr haben?

Regression ■ Typische Fehlvorstellungen

1. Nicht-Linearität

Eine lineare Einfachregression analysiert nur **lineare** Zusammenhänge. Zwei Variablen können aber auch einen **hohen nicht-linearen Zusammenhang** aufweisen.

Gleiche Korrelation $r = 0,67$;
gleiche lineare Regressionsgerade $y = 0,5x + 3$



2. Kausalität

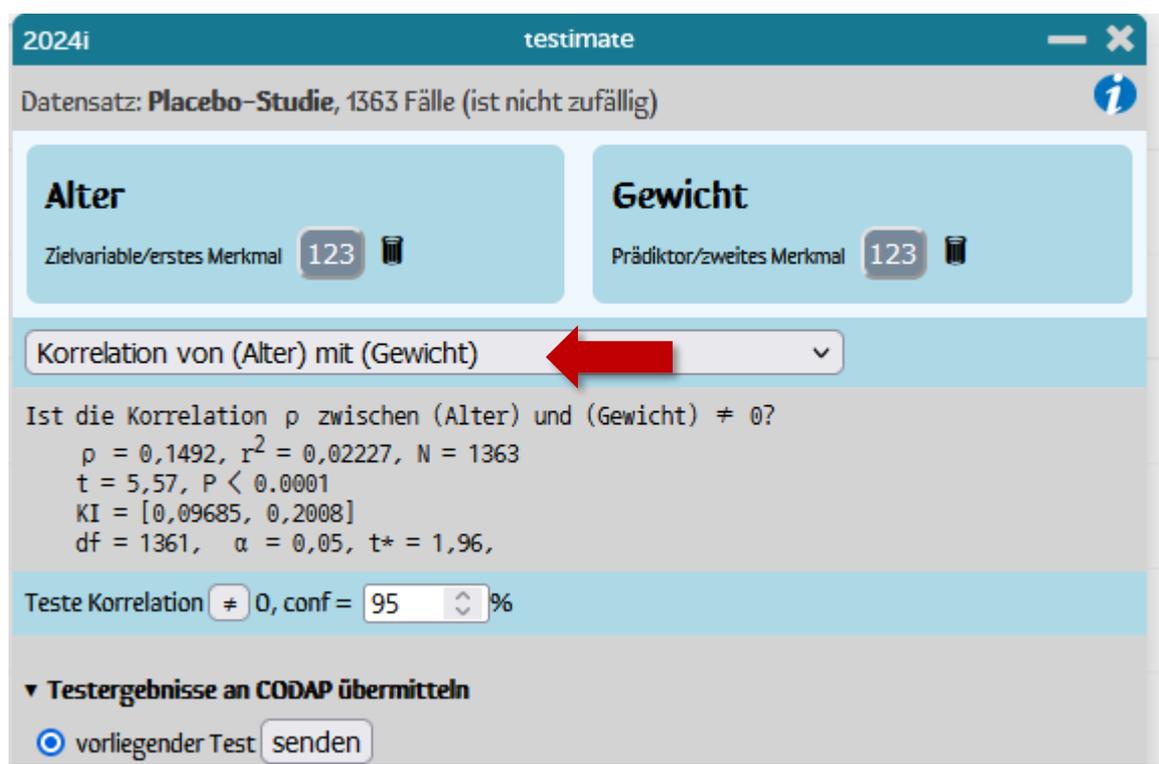
Wie schon bei der Korrelation gilt auch für die Regression: Ob der gefundene Zusammenhang auch tatsächlich **kausal** interpretiert werden darf, muss zunächst kritisch geprüft werden.





Korrelation ▪ Mit CODAP

Betrachten Sie nun wieder den Datensatz **Placebo-Studie**.



2024i testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig)

Alter
Zielvariable/erstes Merkmal 123

Gewicht
Prädiktor/zweites Merkmal 123

Korrelation von (Alter) mit (Gewicht) ←

Ist die Korrelation ρ zwischen (Alter) und (Gewicht) $\neq 0$?

$\rho = 0,1492$, $r^2 = 0,02227$, $N = 1363$
 $t = 5,57$, $P < 0,0001$
 $KI = [0,09685, 0,2008]$
 $df = 1361$, $\alpha = 0,05$, $t^* = 1,96$,

Teste Korrelation $\neq 0$, conf = 95 %

▼ Testergebnisse an CODAP übermitteln
 vorliegender Test

Ziehen Sie dann
 „Alter“ auf das 1. Merkmal
 und „Gewicht“ auf das 2.
 Merkmal
 (oder umgekehrt)

Und wählen Sie als Testart
 „Korrelation von ...“



Korrelation ■ Mit Fathom

Der **Korrelationskoeffizient r bzw. ρ (als Effektstärke)** beträgt etwa 0,15 (genau: 0,1492). Es zeigt sich also ein kleiner positiver Zusammenhang zwischen Alter und Gewicht.

2024i estimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig)

Gewicht
Zielvariable/erstes Merkmal 123

Alter
Prädiktor/zweites Merkmal 123

Lineare Regression von (Gewicht) als eine Funktion von (Alter)

Wie hängt (Gewicht) von (Alter) ab?
LSRL: Gewicht = 0,1513 (Alter) - NaN
N = 1363, $\rho = 0,1492$, $r^2 = 0,02227$

▼ Regression Zusammenfassung

Steigung	0,1513	95% KI = [0,09801, 0,2047]
y-Achsenabschnitt	69,18	95% KI = [67,08, 71,27]

Test auf Steigung $\neq 0$
t = 5,57, P < 0,0001
df = 1361, $\alpha = 0,05$, t* = 1,96,

Es kann auch mithilfe eines **Signifikanztests** überprüft werden, ob sich der Korrelationskoeffizient signifikant von 0 unterscheidet. Da dieser p-Wert kleiner als 0,05 ist (sogar kleiner als 0,0001), kann man von einem signifikanten (aber eben kleinen!) Zusammenhang zwischen den beiden Merkmalen ausgehen.



Regression ▪ Mit CODAP

Betrachten Sie nun wieder den Datensatz **Placebo-Studie**.
Wir möchten nun eine Regressionsanalyse durchführen.
Wählen Sie hierzu „**Lineare Regression von ... auf...**“:

2024i
testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig) i

Gewicht

Zielvariable/erstes Merkmal 123 🗑️

Alter

Prädiktor/zweites Merkmal 123 🗑️

Lineare Regression von (Gewicht) als eine Funktion von (Alter) ▾

Wie hängt (Gewicht) von (Alter) ab?
 LSRL: $\text{Gewicht} = 0,1513 (\text{Alter}) - \text{NaN}$
 $N = 1363, \rho = 0,1492, r^2 = 0,02227$

↑

▼ **Regression Zusammenfassung**

Steigung	0,1513	95% KI = [0,09801, 0,2047]
y-Achsenabschnitt	69,18	95% KI = [67,08, 71,27]

Test auf Steigung $\neq 0$





Regression ■ Mit CODAP

Wählen Sie nun „**Gewicht**“ als **abhängige Variable** und „**Alter**“ als **unabhängige Variable**. Die **Richtung** ist bei der Regression (im Gegensatz zur Korrelation) entscheidend! Sie vermuten, dass das Alter tendenziell einen Einfluss auf das Gewicht hat, aber nicht, dass das Gewicht einen Einfluss auf das Alter hat. **ABER**: Vorsicht mit Kausaldeutungen bei Regressionen!

2024i testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig)

Gewicht
Zielvariable/erstes Merkmal 123

Alter
Prädiktor/zweites Merkmal 123

Lineare Regression von (Gewicht) als eine Funktion von (Alter) ▾

Wie hängt (Gewicht) von (Alter) ab?
LSRL: $\text{Gewicht} = 0,1513 (\text{Alter}) - \text{NaN}$
 $N = 1363, \rho = 0,1492, r^2 = 0,02227$

▼ Regression Zusammenfassung

Steigung	0,1513	95% KI = [0,09801, 0,2047]
y-Achsenabschnitt	69,18	95% KI = [67,08, 71,27]



Regression ■ Mit CODAP

2024i testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig) i

Gewicht

Zielvariable/erstes Merkmal 123 🗑️

Alter

Prädiktor/zweites Merkmal 123 🗑️

Lineare Regression von (Gewicht) als eine Funktion von (Alter) ▾

Wie hängt (Gewicht) von (Alter) ab?
LSRL: $\text{Gewicht} = 0,1513 (\text{Alter}) - \text{NaN}$
N = 1363, $\rho = 0,1492$, $r^2 = 0,02227$

▼ Regression Zusammenfassung

Steigung	0,1513	95% KI = [0,09801, 0,2047]
y-Achsenabschnitt	69,18	95% KI = [67,08, 71,27]

Test auf Steigung $\neq 0$
t = 5,57, P < 0,0001
df = 1361, $\alpha = 0,05$, t* = 1,96,

Ich bin leider zu klein für
mein Gewicht. Aber ich
esse und esse...aber
werde einfach nicht größer.



Die Regressionsgleichung lautet

$$\text{Gewicht} = 0,1513 \cdot \text{Alter} + 69,18$$

Das erwartete Gewicht einer 20-jährigen Person kann so „vorhergesagt“
werden: $\text{Gewicht (20 Jahre)} = 0,15 \cdot 20 + 69,18 = 72,18$



Regression ■ Mit CODAP

Die Regressionsgleichung lautet

$$\text{Gewicht} = 0,15 \cdot \text{Alter} + 69,18$$

Das erwartete Gewicht einer 20-jährigen Person kann so „vorhergesagt“ werden: $\text{Gewicht (20 Jahre)} = 0,15 \cdot 20 + 69,18 = 72,18$

2024i testimate

Datensatz: **Placebo-Studie**, 1363 Fälle (ist nicht zufällig) ?

Gewicht

Zielvariable/erstes Merkmal 123 🗑️

Alter

Prädiktor/zweites Merkmal 123 🗑️

Lineare Regression von (Gewicht) als eine Funktion von (Alter) ▾

Wie hängt (Gewicht) von (Alter) ab?
 LSRL: $\text{Gewicht} = 0,1513 (\text{Alter}) - \text{NaN}$
 $N = 1363, \rho = 0,1492, r^2 = 0,02227$

▼ Regression Zusammenfassung

Steigung	0,1513	95% KI = [0,09801, 0,2047]
y-Achsenabschnitt	69,18	95% KI = [67,08, 71,27]

Test auf Steigung $\neq 0$
 $t = 5,57, P < 0,0001$
 $df = 1361, \alpha = 0,05, t^* = 1,96,$

Allgemeinere Interpretation der Steigung:

Pro Lebensjahr steigt das Gewicht im Schnitt um 0,15 kg.

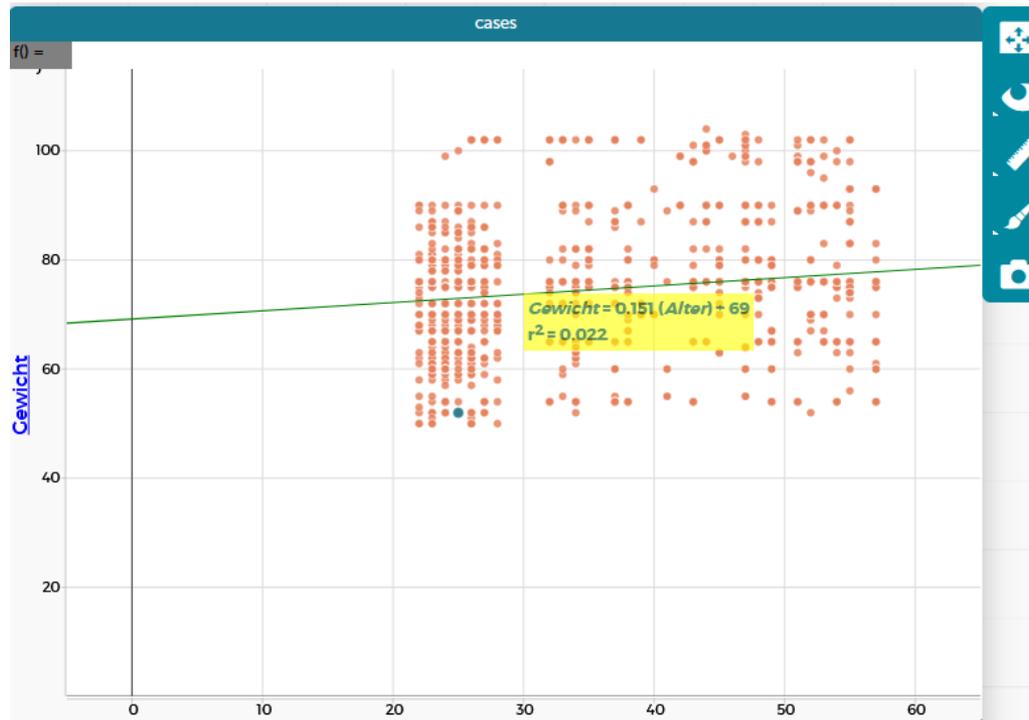
Tieferegehende Überlegungen:

- Wäre hier eine Trennung nach Geschlecht zielführend?
- Zusätzlich hat auch die Körpergröße einen entscheidenden Einfluss auf das Gewicht. Dafür kontrollieren?

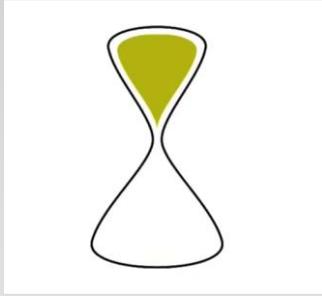


Regression ■ Mit CODAP

Bevor die Regressionsanalyse durchgeführt wird, lohnt sich allerdings ein Blick auf das zugehörige **Streudiagramm** (mithilfe von „Graph“). In das Streudiagramm kann man sich ebenfalls die Regressionsgerade einzeichnen und deren Funktion ausgeben lassen:



Im vorliegenden Beispiel erklärt das Modell nur sehr wenig Varianz!



Zeit: 7 Minuten

Workshop-Phase

Bearbeiten Sie
bitte Übung 6





Mögliche Lösung (Korrelation):

2024i testimate

Datensatz: **Mikrozensus_2010**, 500 Fälle (ist nicht zufällig)

Wohnfläche

Zielvariable/erstes Merkmal 123

Alter

Prädiktor/zweites Merkmal 123

Korrelation von (Wohnfläche) mit (Alter) ▼

Ist die Korrelation ρ zwischen (Wohnfläche) und (Alter) $\neq 0$?

$\rho = -0,1208, r^2 = 0,0146, N = 493$
 $t = -2,7, P = 0,007225$
 $KI = [-0,2072, -0,03267]$
 $df = 491, \alpha = 0,05, t^* = 1,96,$

Teste Korrelation $\neq 0$, conf = %

▼ Testergebnisse an CODAP übermitteln

vorliegender Test

Es gibt einen **kleinen (!) negativen Zusammenhang** zwischen den Merkmalen Alter und Wohnfläche. Ältere Menschen wohnen also tendenziell in kleineren Wohnungen.

Dieser sehr kleine Zusammenhang ist **signifikant**, was aber vielleicht auch an unserer großen Stichprobe liegt.



Mögliche Lösung (Korrelation):

2024i testimate

Datensatz: **Mikrozensus_2011**, 100 Fälle (ist nicht zufällig)

Alter
Zielvariable/erstes Merkmal 123

Wohnfläche
Prädiktor/zweites Merkmal 123

Korrelation von (Alter) mit (Wohnfläche)

Ist die Korrelation ρ zwischen (Alter) und (Wohnfläche) $\neq 0$?

$\rho = 0,1142$, $r^2 = 0,01304$, $N = 97$
 $t = 1,12$, $P = 0,2655$
 $KI = [-0,08984, 0,309]$
 $df = 95$, $\alpha = 0,05$, $t^* = 1,99$,

Teste Korrelation $\neq 0$, conf = 95 %

▼ Testergebnisse an CODAP übermitteln

vorliegender Test

Hier ein Beispiel mit nur 100 gezogenen Fällen aus dem Mikrozensus-Datensatz. Im vorliegenden Beispiel ist der kleine Effekt nun nicht mehr signifikant.

Signifikanz sagt also nichts über die Größe eines Effekts, sondern hängt mit der Größe der Stichprobe zusammen.



Mögliche Lösung (Korrelation):

2024i testimate

Datensatz: **Mikrozensus_2010**, 500 Fälle (ist nicht zufällig)

Geburtsjahr

Zielvariable/erstes Merkmal

Alter

Prädiktor/zweites Merkmal

Korrelation von (Geburtsjahr) mit (Alter) v

Ist die Korrelation ρ zwischen (Geburtsjahr) und (Alter) $\neq 0$?

$\rho = -0,9998, r^2 = 0,9995, N = 500$
 $t = -1040, P < 0,0001$
 $KI = [-0,9998, -0,9997]$
 $df = 498, \alpha = 0,05, t^* = 1,96,$

Teste Korrelation , conf = %

Amüsant: Alter und Geburtsjahr hängen auch sehr stark (natürlich negativ) zusammen.

Warum wird die Korrelation nicht exakt (sondern nur fast) -1? Haben Sie Ideen? 😊



Mögliche Lösung (Regression):

Es gibt einen **kleinen negativen Zusammenhang** zwischen den Merkmalen Alter und Gesamtmiete. Ältere Menschen wohnen also tendenziell in günstigeren Wohnungen. (Passt ja auch dazu, dass diese häufiger in kleineren Wohnungen wohnen.)

2024i | testimate

Datensatz: **Mikrozensus_2010**, 215 Fälle (ist nicht zufällig)

Gesamtmiete
Zielvariable/erstes Merkmal 123

Alter
Prädiktor/zweites Merkmal 123

Lineare Regression von (Gesamtmiete) als eine Funktion von (Alter) ▾

Wie hängt (Gesamtmiete) von (Alter) ab?
LSRL: Gesamtmiete = -1,767 (Alter) - NaN
N = 215, $\rho = -0,14$, $r^2 = 0,01959$

▼ Regression Zusammenfassung

Steigung	-1,767	95% KI = [-3,455, -0,07872]
y-Achsenabschnitt	625,4	95% KI = [544,7, 706]

Test auf Steigung $\neq 0$
t = -2,06, P = 0,04032
df = 213, $\alpha = 0,05$, $t^* = 1,97$,

Die zugehörige Regressionsgleichung ist links zu sehen:
 $y = 625,4m^2 - 1,767 \cdot \text{Alter}$

Das Modell erklärt allerdings nur wenig Varianz (nämlich nur ca. 2%).

Das Simpson-Paradoxon



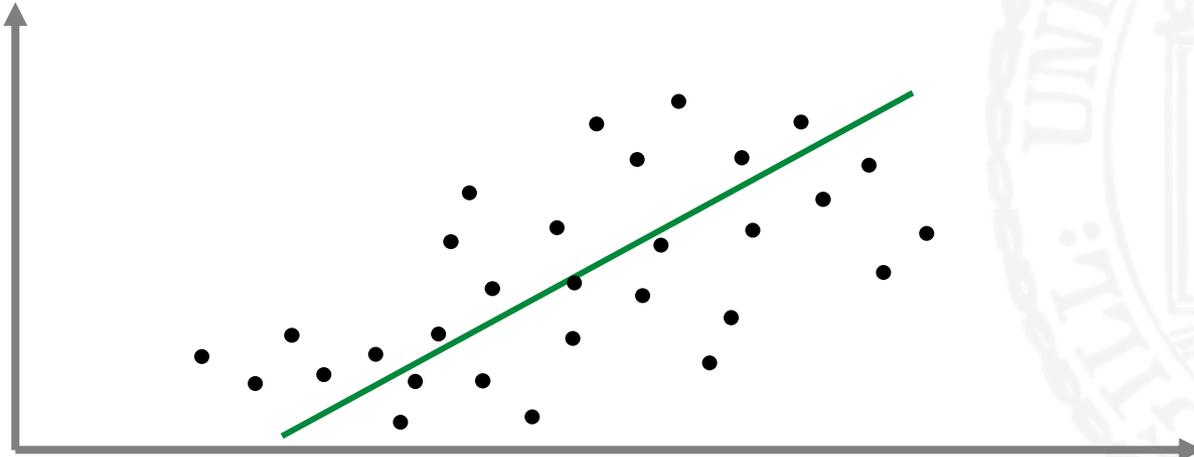
Das Simpson-Paradoxon



vgl. statistiker-blog.de (Weigel, 2021)

	Geimpft			Ungeimpft		
	Gesamt	Verstorben	Tote je 100.000	Gesamt	Verstorben	Tote je 100.000
Gesamt	2.000.000	42	2,10	2.000.000	35	1,75

Sterbewahrscheinlichkeit



Anteil geimpfter
Personen

Das Simpson-Paradoxon



vgl. statistiker-blog.de (Weigel, 2021)

	Geimpft			Ungeimpft		
Alters- klasse	Gesamt	Verstorben	Tote je 100.000	Gesamt	Verstorben	Tote je 100.000
10-40	500.000	2	0,40	1.500.000	15	1,00
40-60	1.500.000	40	2,67	500.000	20	4,00
Gesamt	2.000.000	42	2,10	2.000.000	35	1,75

Sowohl in der Altersgruppe der jüngeren Personen, als auch in der Altersgruppe der älteren Personen ist die Sterblichkeitsrate unter den Geimpften niedriger als unter den Ungeimpften.

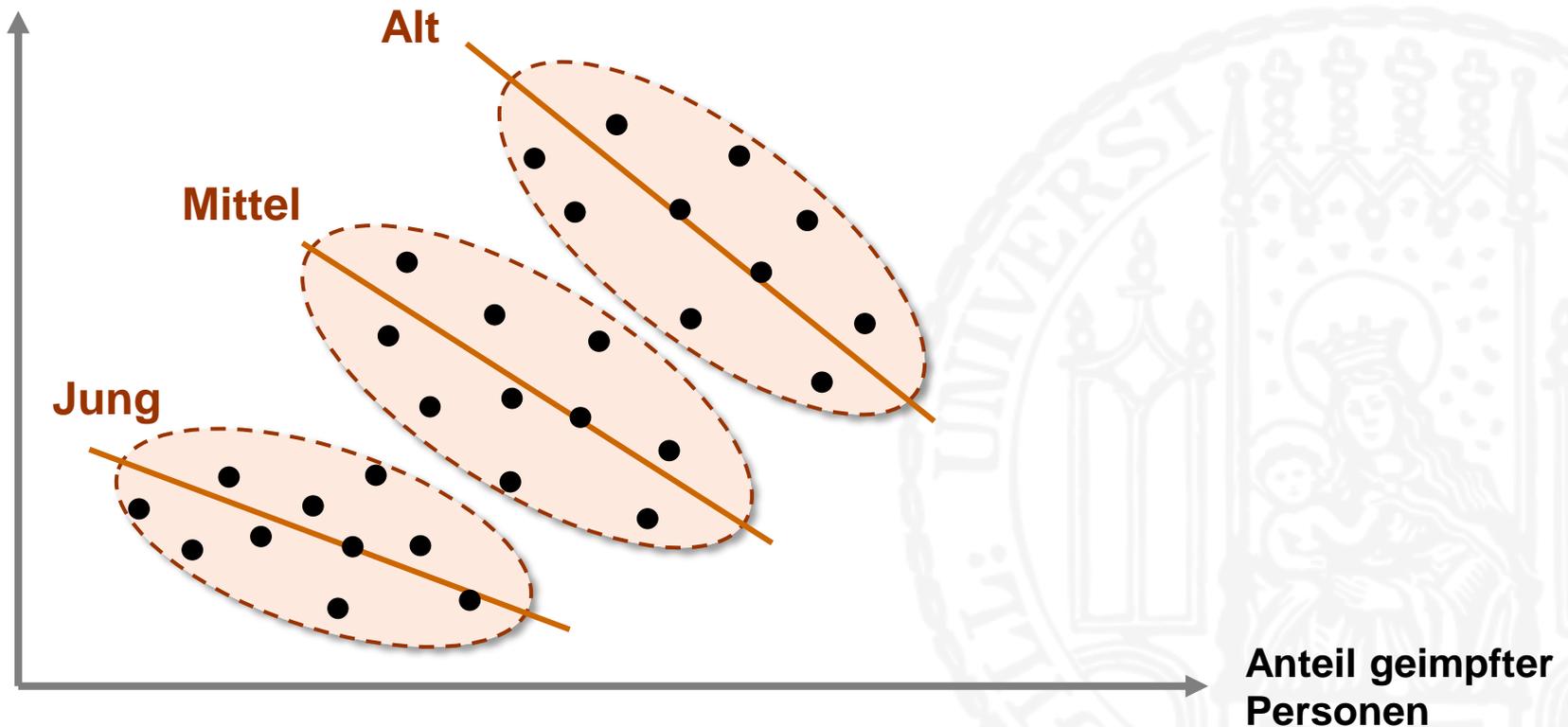
Trennt man allerdings nicht nach Altersgruppen, so ist die Sterblichkeitsrate unter den Ungeimpften niedriger als unter den Geimpften.

Grund für die Paradoxie: Diese Zahlen kamen zu einem Zeitpunkt zustande, an dem die meisten älteren Personen bereits geimpft waren und die jüngeren Personen erst mit den Impfungen begonnen haben.



Simpson-Paradoxon erklärt über Regression

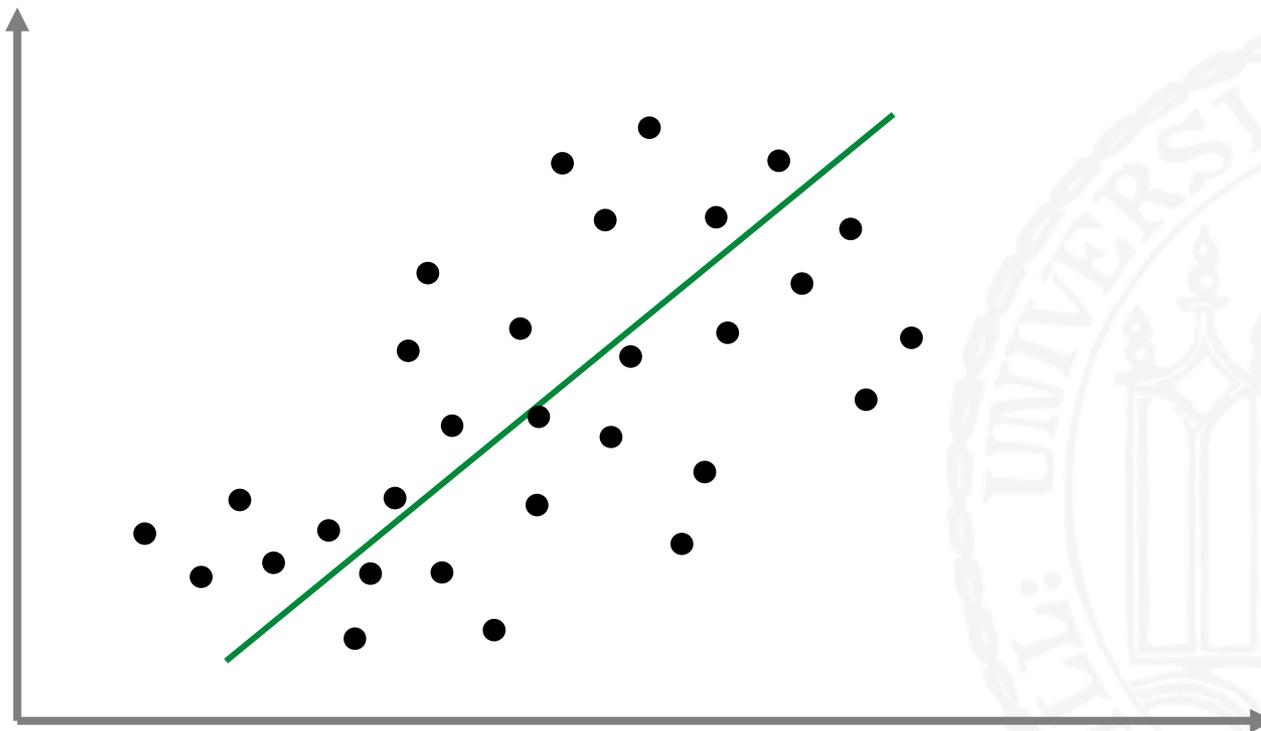
Sterbewahrscheinlichkeit





Simpson-Paradoxon erklärt über Regression

Sterbewahrscheinlichkeit



Anteil geimpfter
Personen

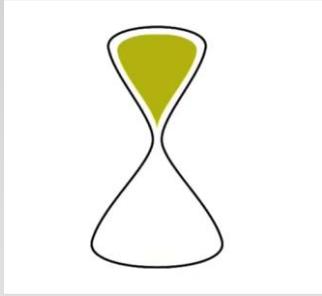
Simpson-Paradoxon

Das Simpson-Paradoxon beschreibt eine **Scheinkorrelation**, also eine Korrelation, die in Wirklichkeit nicht existiert oder sogar in umgekehrter Richtung gilt.

Beim Simpson-Paradoxon fällt die Bewertung verschiedener Gruppen unterschiedlich aus, je nachdem ob man die **Gruppen einzeln** betrachtet oder **zusammengefasst** betrachtet.

Wichtig ist also immer nach möglichen **Hintergrundvariablen** zu suchen, die für den gefundenen Effekt verantwortlich sein könnten.





Workshop-Phase

Bearbeiten Sie
bitte Übung 7

Zeit: 7 Minuten





Mögliche Lösung:

	Frauen			Männer			Gesamt		
	beworben	ange- nommen	Quote	beworben	ange- nommen	Quote	beworben	ange- nommen	Quote
Studien- gang 1	100	80	80%	400	280	70%	500	360	72%
Studien- gang 2	400	160	40%	100	20	20%	500	180	36%
Gesamt	500	240	48%	500	300	60%	1.000	540	54%

a) Argumentieren Sie mit obenstehender Tabelle, warum der Universität vorgeworfen wurde, dass Frauen bei der Studienzulassung benachteiligt wurden.

Fasst man beide Studiengänge zusammen, so werden nur 48% der Frauen, aber 60% der Männer zu einem Studium an der Universität zugelassen.



Mögliche Lösung:

	Frauen			Männer			Gesamt		
	beworben	ange- nommen	Quote	beworben	ange- nommen	Quote	beworben	ange- nommen	Quote
Studien- gang 1	100	80	80%	400	280	70%	500	360	72%
Studien- gang 2	400	160	40%	100	20	20%	500	180	36%
Gesamt	500	240	48%	500	300	60%	1.000	540	54%

b) Argumentieren Sie nun, warum man bei den obigen Annahmehquoten umgekehrt auch von einer Diskriminierung der Männer sprechen könnte.

Betrachtet man jeden der beiden Studiengänge einzeln, so ist in beiden Studiengängen die Annahmehquote bei den Frauen höher als bei den Männern.



Mögliche Lösung:

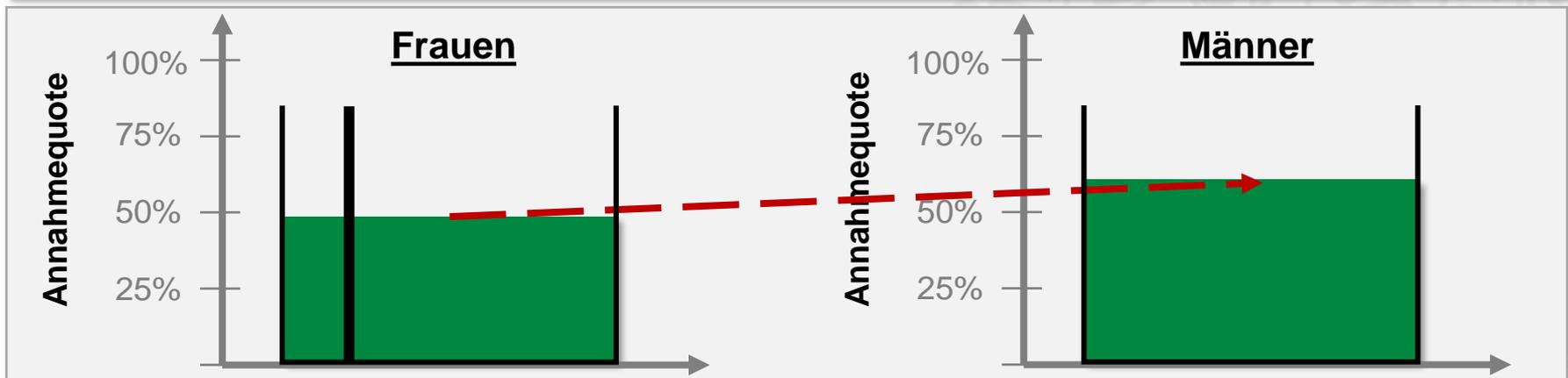
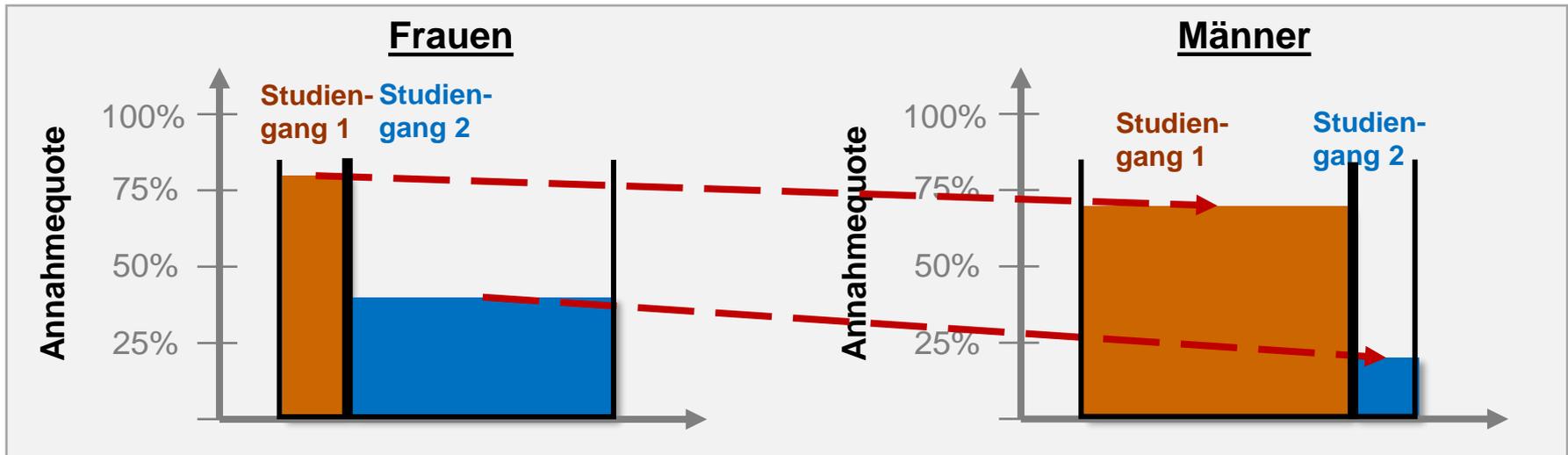
	Frauen			Männer			Gesamt		
	beworben	ange- nommen	Quote	beworben	ange- nommen	Quote	beworben	ange- nommen	Quote
Studien- gang 1	100	80	80%	400	280	70%	500	360	72%
Studien- gang 2	400	160	40%	100	20	20%	500	180	36%
Gesamt	500	240	48%	500	300	60%	1.000	540	54%

c) Erklären Sie nun Ihrem Sitznachbarn, wodurch in obigem Beispiel die Paradoxie zustande kommt.

Die Studiengänge unterscheiden sich stark bezüglich der Annahmehquoten (in Studiengang 1 höher als in Studiengang 2). Frauen bewerben sich häufiger auf den Studiengang mit den niedrigen Annahmehquoten. Die Universität sollte also zwei Diskussionen führen: 1. Warum werden Männer pro Studiengang benachteiligt? 2. Warum werden in dem Studiengang, auf den sich überwiegend Frauen bewerben nicht mehr Plätze geschaffen; bzw. warum bewerben sich hier mehr Frauen?



Visualisierung des Simpson-Paradoxon





Abschließendes Fazit ■ Inhaltlich

- Merkmale können **nominal, ordinal oder metrisch** skaliert sein.
- In der quantitativen empirischen Forschung stellt man in den meisten Fällen **Abhängigkeitshypothesen, Zusammenhangshypothesen oder Unterschiedshypothesen** (Zusammenhangshypothesen sind hierbei das übergeordnete Konzept).
- **Chi-Quadrat-Tests** werden bei **Abhängigkeitshypothesen** durchgeführt.
- **t-Tests** eignen sich bei **Unterschiedshypothesen** (z.B. auch mit Excel möglich).
- Zusammenhangshypothesen können mithilfe von **Korrelation** (von Kausalität abgrenzen!) und **Regression** untersucht werden (z.B. mithilfe von Fathom).
- Das **Simpson-Paradoxon** tritt regelmäßig in der Realität auf und kann gut mit dem Thema „Regression“ verknüpft werden.
- **p-Werte** sind mit **vielen Fehlvorstellungen** verbunden.



Abschließendes Fazit ▪ Didaktisch

- **Verständnisorientierung vor Kalkülorientierung:** Nehmen Sie sich viel Zeit für die Erarbeitung der grundlegenden Ideen und Konzepte! Diese sollten nicht durch konkrete Berechnungen und Prozeduren (Kochrezepte) überlagert werden!
- **Aufbau negativen Wissens (Oser et al. 1999):** p-Werte sind mit vielen Fehlvorstellungen verbunden. Unterrichten Sie daher unbedingt auch typische Verwechslungskandidaten und auch die Kontroversen um Signifikanztests (und ggfs. Alternativen – wie z.B. Effektstärken)! Trotzdem: Der p-Wert ist eine didaktische Chance!
- **Simulationen (Biehler & Prömmel):** Simulationen eignen sich bei Hypothesentests, um ein Verständnis für den p-Wert zu erzeugen.
- **Wissenschaftspropädeutisches Potential des Vertiefungsmoduls Statistik:** Insgesamt eignet sich das Statistik-Modul sehr gut für eine **Vorbereitung für viele Studiengänge.**



Zum Abschluss

Wie hat es Ihnen heute gefallen?



Vielen Dank für die Aufmerksamkeit!

(Fragen, Anregungen, Kritik auch
gerne an Karin.Binder@lmu.de)

