



Basiswissen Statistik und Versuchsplanung

Moritz Körber

08.02.2016

Ablauf einer Studie



Ablauf einer Studie



1. Fragestellung

- Steht zu Beginn jeder Forschungsarbeit: **Was möchte ich herausfinden?**
- Meist sehr breit und vage
- Ausprägungen: Zusammenhänge, Unterschiede
- Beispiele:
 - Wie wirkt sich die Benutzung eines Mobiltelefons auf Autofahrer aus?
 - Wovon hängt die Wahrnehmung eines Gewichts ab?
 - Gibt es Unterschiede zwischen zwei Anzeige Konzepten?
 - ...

Ablauf einer Studie



2. Hypothesen

- Konkrete, **untersuchbare Aussagen**, abgeleitet aus der Fragestellung
- Empirische Vorhersage, welche mit der Wirklichkeit verglichen wird
- Überlegung: Wie kann ich die Fragestellung anhand konkreter, realer Beobachtungen (objektive und subjektive Daten) beantworten?
- Beispiele:
 - „*Personen, die während einer Fahrsimulatorenaufgabe ein Mobiltelefon nutzen, machen mehr Fehler in dieser Aufgabe als Personen, die kein Mobiltelefon nutzen.*“
 - „*Weibliche Versuchspersonen schätzen das Gewicht eines Wagens höher ein als männliche Versuchspersonen.*“
 - „*Wenn X, dann trifft Y ein.*“
 - „*Wenn nicht X, dann trifft auch nicht Y ein*“
 - ...

2. Hypothesen

- Aussagenform erlaubt, die Hypothesen in einem Versuch zu überprüfen, statistisch zu testen und zu widerlegen
- Ungerichtete vs. gerichtete Hypothesen:
 - **Ungerichtet:** ohne Vorwissen; keine Richtung oder Art des Effekts, nur dass ein Unterschied/Zusammenhang besteht
 - „Die neue und die alte Maus sind unterschiedlich gut.“
 - „Es besteht ein Zusammenhang zwischen Temperatur und Wohlbefinden.“
 - „Probanden mit Fahrerassistenzsystem verändern ihre Fahrleistung nach einer Eingewöhnungsfahrt.“
 - **Gerichtet:** auf Grund von Vorwissen; Richtung und Art des Effekts wird angegeben
 - „Die neue Maus ist *besser* als die alte Maus“
 - „Es besteht ein *positiver* Zusammenhang zwischen Temperatur und Wohlbefinden.“
 - „Die Bearbeitung einer Nebenaufgabe *erhöht* die Übernahmezeit.“

Ablauf einer Studie



3. Operationalisierung

- Die Hypothesen sind abstrakt gehalten
- Bei der Operationalisierung werden den Begriffen der Hypothese beobachtbare Phänomene und empirische Indikatoren zugeordnet
- Wie kann ich die Variablen der Hypothesen erheben/messen?
- Indikatoren können subjektiv (z. B. Fragebogen) oder objektiv (z. B. Reaktionszeit) sein
- Beispiel: *„Personen, die während einer Fahrsimulatoreaufgabe ein Mobiltelefon nutzen, machen mehr Fehler in dieser Aufgabe als Personen, die kein Mobiltelefon nutzen.“*
 - Wie kann ich *Fehler* konkret erheben?
 - Anzahl der Markierungsüberschreitungen
 - Anzahl der Kollisionen
 - ...

Von Fragestellung zu Operationalisierung

Vorgehen

Fragestellung

- Wie wirkt sich die Benutzung eines Mobiltelefons auf Autofahrer aus?

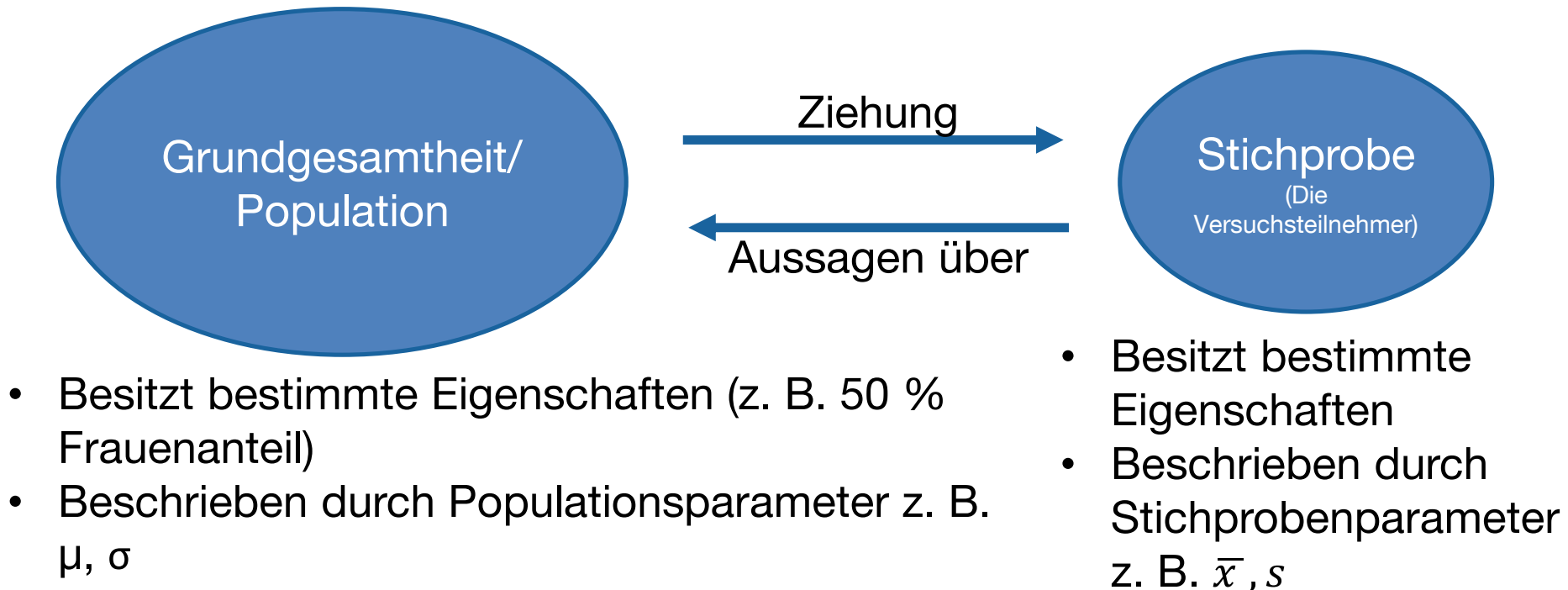
Hypothesen

- „*Personen, die während einer Fahrsimulatorenaufgabe ein Mobiltelefon nutzen, machen mehr Fehler in dieser Aufgabe als Personen, die kein Mobiltelefon nutzen.*“

Operationalisierung

- Anzahl der Überschreitungen der Fahrbahnmarkierung

Prinzip der Inferenzstatistik



Damit die Aussagen aus den Stichprobenergebnissen für die Population gelten, muss diese möglichst der Grundgesamtheit entsprechen (= repräsentative Stichprobe)

Ablauf einer Studie



Versuchsplanung

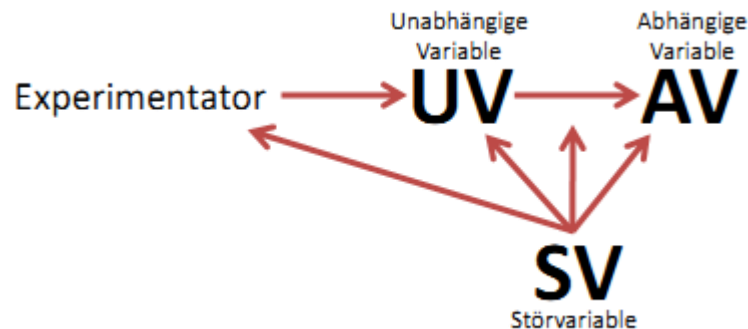
Zwei Arten von Versuchen:

- **Experiment:** Der Experimentierende greift in die Datenerhebung ein; Versuchspersonen werden verschiedenen Bedingungen zugeordnet, die miteinander verglichen werden; Bsp.: Vergleich zwischen zwei Anzeigekonzepten
- **Korrelative Studie:** Zusammenhänge zwischen zwei Variablen werden betrachtet; Bsp.: Zusammenhang zwischen Fahrerfahrung und Reaktionszeit bei einer kritischen Situation
- Nur bei Experimenten sind kausale Schlussfolgerungen möglich!

4. Versuchsplan

- **Kennzeichen eines Experiments:**

1. Mindestens eine Variable (UV; Bedingungen; Gruppen; Messzeitpunkte) wird systematisch variiert und der Einfluss auf eine weitere Variable (AV) registriert
2. Dabei werden die Störvariablen (SV) möglichst ausgeschaltet oder kontrolliert



- Kausalität: Durch das Design kann sichergestellt werden, dass die Veränderung in der AV nur durch die Unterschiede in der UV zustande kommen

4. Versuchsplan

- **Versuchsleiter/Experimentierender (VI)**
- **Versuchsperson (Vp)**
- **Unabhängige Variable (UV):** wird vom Experimentierenden/Versuchsleiter in einer empirischen Untersuchung gezielt durch Manipulation oder Selektion verändert/variiert; Bsp.: Mobiltelefonbenutzung während der Fahrt oder keine Mobiltelefonbenutzung während der Fahrt
- **Abhängige Variable (AV):** Die Variable, für welche die Auswirkungen der Variation der UV untersucht werden soll; wird in einer empirischen Untersuchung beobachtet und gemessen; Bsp.: Anzahl der Unfälle im Fahrversuch
- **Störvariable (SV):** unerwünschter Einfluss außerhalb der planmäßig variierten unabhängigen Variablen, der UV, AV und Versuchsleiter beeinflusst; liegen in den Vpn, im VI oder in der Untersuchungssituation; z. B. Lärm, Alter, Unterschiede im Fahrkönnen, Tageszeit der Testung, der Versuchsleiter selbst!

4. Versuchsplan – Möglichkeiten zur Kontrolle der Störvariablen (Auswahl)

- **Eliminieren:** z. B. bei äußeren Störeinflüssen; Lärm: ein schalldichtes Labor verwenden; führt im Extremfall zu kontrollierten Laborbedingungen.
- **Konstanthalten:** Führt zu standardisierten Bedingungen; evtl. schwer auf natürliche Umstände zu übertragen (zu generalisieren); Bsp.: Alle Teilnehmer sind demselben Lärmniveau ausgesetzt.
- **Aufnahme in den Versuchsplan:** Die Störvariable wird als UV aufgenommen und damit systematisch variiert; z. B. Alter, Geschlecht, Zugehörigkeit zu LMU oder TUM.
- **Ausbalancieren:** Bei wiederholten Messungen; Beispiel: Versuchsperson bewertet Bedienfreundlichkeit verschiedener Interaktionskonzepte nacheinander. Probleme: Transfereffekte (Bewertung des letzten Konzeptes „färbt“ Bewertung des nächsten Konzeptes, Trainingseffekt) und Reihenfolgeeffekte (bei letztem Konzept möglicherweise Motivationsverlust). Lösung: Ausbalancieren der Reihenfolge der Bedingungen. Die Reihenfolge kann für jede Versuchsperson zufällig ausgewählt werden oder systematisch variiert werden (z. B. lateinisches Quadrat).
- **Randomisieren:** zufällige Zuweisung von Probanden zu Untersuchungsbedingungen (UVs); Neutralisierung des Einflusses personenbezogener SV (z. B. Intelligenz). Damit gleichen sich Unterschiede aus und die UVs sind hinsichtlich der Versuchspersonen äquivalent. Empfehlung: mindestens 20 Versuchspersonen pro UV.

4. Versuchsplan

- Einteilungsmöglichkeiten für Studien:
 - Feldstudien:
 - natürlichen Bedingungen, außerhalb des Labors; z. B. Nutzung eines Smartphones im Alltag
 - Kontrolle der Störvariablen schwierig
 - Laborstudien:
 - kontrollierte Laborbedingungen, Fahrsimulator
 - Generalisierung der Ergebnisse auf natürliche Situationen schwierig
 - Ein- oder mehrfaktoriell (mehrere UVs)

4. Versuchsplan

Zwei Arten, die UV in den Versuchsplan einzubauen:

Within-subject design

Einer Gruppe VPs werden beide
Schilder nacheinander präsentiert



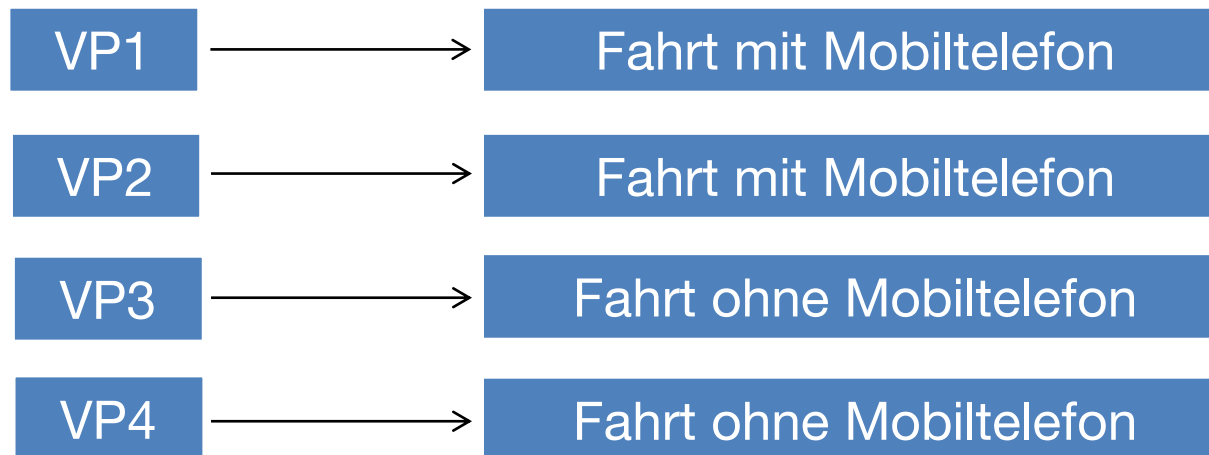
Between-subject design

Einer Gruppe VPs wird Schild 1 präsentiert und
einer anderen Gruppe wird Schild 2 präsentiert



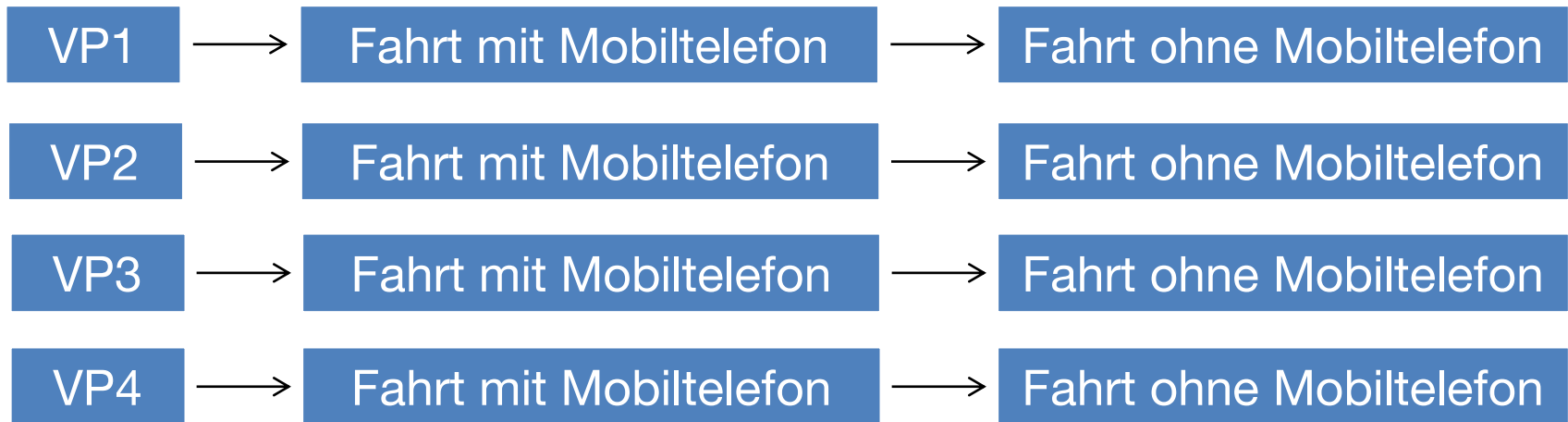
4. Versuchsplan

- Between-subject-Design: Jeder Proband fährt nur eine Fahrt.



4. Versuchsplan

- within-subject-design/Design mit Messwiederholung: Jeder Proband fährt jede Fahrt:



- **Vorteile:** Weniger Versuchspersonen benötigt, höhere Teststärke durch weniger Fehlervarianz
- **Nachteile:** Reihenfolgeeffekte (z. B. Ermüdung im letzten Durchgang, Training) und Transfereffekte (Bedingung wirkt sich auf die nächste(n) Bedingung(en) aus; z. B. negatives Bild -> positives Bild $\neq$ positives Bild -> negatives Bild)
- **Lösung:** Reihenfolge der Bedingungen randomisieren oder systematisch variieren (z. B. Lat. Quadrat)

Mehrfaktorielle Versuchspläne

Beispiel: 2 UVs (UV a = Fahrerassistenzsystem; UV b = Güte der Instruktion)

		UV a		
		a1	a2	a3
UV b	b1	a1b1	a2b1	a3b1
	b2	a1b2	a2b2	a3b3

Jede Kombination der beiden UVs stellt eine Gruppe/Versuchsbedingung dar

Korrelation und Kausalität

- Bei einer korrelativen Studie kann nur festgestellt werden, dass zwei Variablen zusammenhängen, jedoch können im Gegensatz zum Experiment keine kausalen Schlüsse gezogen werden. Durch die systematische Variation der UV im Experiment können Veränderungen in der AV kausal auf die Variation zurückgeführt werden. Bei der Korrelation geht dies nicht.
- Ein beliebtes Beispiel zum Thema „Korrelation ist nicht Kausalität“ ist der positive Zusammenhang der Anzahl der Störche und der Geburtenrate in einem Gebiet: je mehr Störche, desto mehr Kinder. Der Storch bringt also Kinder, logisch, oder? Der Zusammenhang kommt jedoch nur durch die Drittvariable *Urbanisierung* zustande: In ländlichen Gebieten leben mehr Störche und dort werden auch mehr Kinder geboren.

4. Versuchsplan – Qualitätskriterien

1. **Interne Validität:** inwieweit wurden die Störvariablen kontrolliert
 - Auch: Inwieweit können Alternativerklärungen für die Effekte ausgeschlossen werden
 - Wie eindeutig ist die Änderung der AV auf die Variation der UV zurückführbar?
 - Dazu müssen alle Probanden möglichst dieselben standardisierten Bedingungen haben
 - Strecke
 - Handy
 - Fahrsituation
 - Tageszeit
 - Instruktion
 - ...
 - Gering bei Feldstudien, hoch bei Laborstudien

4. Versuchsplan – Qualitätskriterien

2. Externe Validität: inwieweit sind die Ergebnisse verallgemeinerbar oder auf andere Situationen übertragbar

- Anforderungen:
 - Realistisches Fahrszenario
 - Strecke möglichst wirklichkeitsnah
 - Stichprobe möglichst repräsentativ
 - ...
- Hoch bei Feldstudien, gering bei Laborstudien

4. Versuchsplan – Qualitätskriterien

3. Konstruktvalidität: Güte der Operationalisierung

- Inwieweit entspricht die Operationalisierung dem zu messenden Konstrukt
- Bsp. hohe Konstruktvalidität: Fahrprüfung entspricht sehr gut dem Konstrukt „Fahrfähigkeit“
- Bsp. niedrige Konstruktvalidität: Intelligenz mit Thermometer messen

4. Versuchsplan

- Vorversuch:
 - Versuch mit 1–5 Versuchspersonen
 - Daten fließen nicht in die spätere Gesamtauswertung des Versuchs ein
 - Dient zum Test des entwickelten Versuchs
- Vorteile:
 - Hardware kann getestet werden
 - Probleme im Ablauf können aufgedeckt werden
 - Verständnisprobleme der Versuchsperson können aufgedeckt werden
 - Die Ausprägung der AV kann grob eingeschätzt werden
 - Versuchsdauer kann abgeschätzt werden
- Die Durchführung eines Vorversuchs ist unbedingt zu empfehlen!

Ablauf einer Studie



5. Erhebung

Stichprobe (= teilnehmende Probanden)

- Mit/ohne Zufallsauswahl (jeder hat die gleiche Chance in die Stichprobe zu kommen)
- Mit/ohne Schichtung (Einteilung der Grundgesamtheit in Subpopulationen/Schichten)
- Je größer die Stichprobe, desto genauer können Parameter geschätzt werden. Die geplante Größe ist also abhängig von den erwarteten Effekten (z. B. Mittelwertsunterschiede und Streuung -> Analyse der Teststärke/Power), aber auch von praktischen Überlegungen (wie viele Versuchspersonen kann ich wahrscheinlich überhaupt anwerben?)

5. Erhebung

Stichprobe (= teilnehmende Probanden)

- Durch die Wahl der Stichprobe wird bestimmt, welche Schlussfolgerungen durch die Ergebnisse getroffen werden können
- Bsp.: Stichprobe besteht nur aus weibliche Personen zwischen 60 und 85 Jahren => Ergebnisse sind über die Gruppe der weibliche Senioren hinaus schwer generalisierbar
- Es ist aber auch wichtig zu definieren, wer aus der Stichprobe ausgeschlossen werden soll (Bsp.: Nicht-Muttersprachler)

Deswegen: Die Auswahl der Stichprobe muss zur Fragestellung passen!

5. Erhebung

Stichprobe (= teilnehmende Probanden)

- Wenn Teilnehmer von der Datenauswertung ausgeschlossen wurden, so muss dies aufgeführt und begründet werden.
- Mögliche Gründe:
 - Fehler während der Datenaufzeichnung
 - Dem Proband wurde übel
 - Der Proband hatte die Aufgabe nicht verstanden
- Bsp. Stichprobenbeschreibung: Insgesamt nahmen $n = 44$ Versuchspersonen am Versuch teil. Auf Grund von Abstürzen des Fahrsimulators konnten drei Versuchspersonen nicht ausgewertet werden. Die Stichprobe zur Auswertung umfasst damit $n = 41$ Personen.

5. Erhebung

- Bezüglich der Stichprobe sollten Sie folgende Angaben erheben und in Ihrem Text aufführen:
 - Anzahl der Teilnehmer
 - Anzahl und Anteil weiblicher & männlicher Teilnehmer
 - Mittelwert, Standardabweichung und Spannweite des Alters
 - Wie sind die Versuchspersonen geworben worden
 - Wie wurden sie für die Teilnahme belohnt
 - Gegebenenfalls gesundheitliche/körperliche Voraussetzungen (bspw. Farbfehlsichtigkeiten, Sehhilfen, weitere körperliche Einschränkungen, u.U. anthropometrische Gegebenheiten)
 - Gegebenenfalls Erklärung für die Homogenität der Stichprobe (z. B. alles männliche Maschinenbaustudenten zwischen 20 und 25, weil nur in Garching geworben)
 - Alle weiteren Angaben sind studienabhängig. Grundsätzlich sollte man alle Charakteristiken berichten, welche für die Ergebnisse oder die Studie relevant sind; bei einer Fahrsimulatorstudie bspw. die Fahrerfahrung.

5. Erhebung

Ethik

- Der Versuchsleiter muss die Probanden auf Folgendes hinweisen:
 - Die Daten werden anonymisiert
 - Der Versuch kann jederzeit abgebrochen werden
 - Die Instruktion muss dem entsprechen, was gemessen wird (Transparenz)
 - Der Versuch sollte nicht mehr als 1 Stunde Fahrzeit betragen (ansonsten Ethikkommission beurteilen lassen)
 - Die VP sollte nicht mit einem Übermaß an Messtechnik und Fragebögen “belastet” werden (Verhältnismäßigkeit)
- Die Versuchsperson muss per Unterschrift bestätigen, dass sie über die Teilnahmebedingungen informiert wurde und mit diesen einverstanden ist.

Ablauf einer Studie



Statistiksoftware



- am Lehrstuhl verfügbar, sonst teuer
- Einfach, schnell
- Sehr kleiner Funktionsumfang



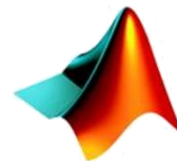
- am Lehrstuhl verfügbar, sonst teuer
- Graphische Benutzeroberfläche



- kostenlos
- Sehr flexibel
- Schwierigerer Einstieg



- kostenlos
- Graphische Benutzeroberfläche
- Bayes-Statistik inkludiert



MATLAB

- Für Studenten kostenlos
- Automatische Datenaufbereitung möglich
- Sehr flexibel

6. Auswertung

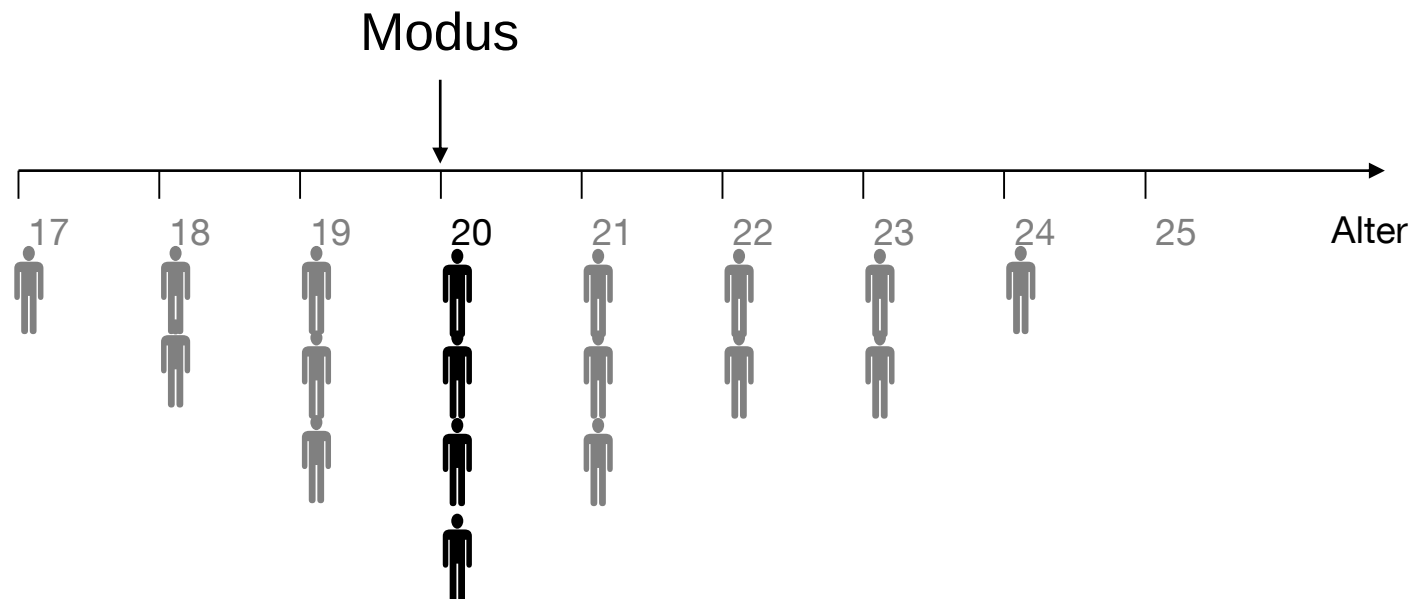
- Wie wird aus einer Beobachtung eines Merkmals eine Zahl? Messtheorie!
- Empirischen Relationen (z. B. A wiegt mehr als B) werden Zahlen zugeordnet (z. B. $A > B$)
- Je nach Erhebung haben die Daten unterschiedlichen Informationsgehalt
- Das Skalenniveau gibt an, welche Aussagen bezüglich der Zahlen getroffen werden darf

	Skala	Mögliche Aussagen	Beispiele
Metrisch	Nominalskala	A und B sind gleich oder A und B sind verschieden	Geschlecht, Religion, Rückennummern im Sport
	Ordinalskala	Rangfolge; A ist größer/kleiner B	Plätze beim Wettrennen, militärische Ränge
	Intervallskala	Differenzen sind sinnvoll interpretierbar: Der Unterschied zwischen 10° und 20° ist so groß wie zwischen 20° und 30° ; willkürlicher Nullpunkt	Temperatur
	Verhältnisskala	Sinnvolle Verhältnisse: A ist doppelt so groß wie B; Natürlicher Nullpunkt	Größe, Zeit

6. Auswertung

Deskriptive Statistik: Maße der zentralen Tendenz

- **Modalwert/Modus:** Der häufigste Wert oder die häufigste Kategorie/Gruppe/Klasse

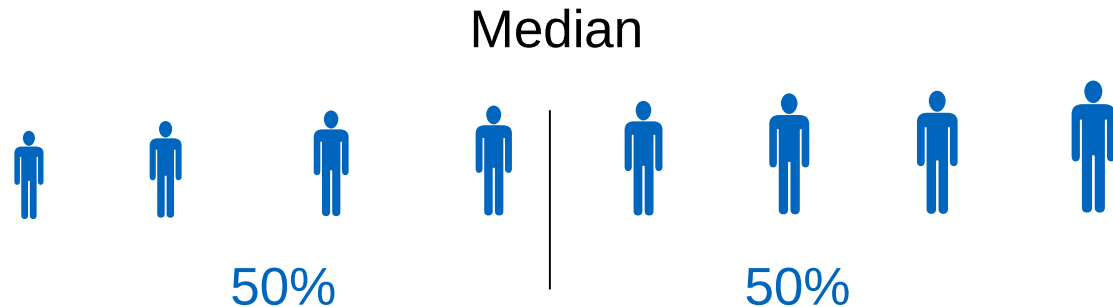


6. Auswertung

Deskriptive Statistik: Maße der zentralen Tendenz

- **Median:** Die Mitte der Verteilung; der Wert über dem 50% der Werte liegen und unter dem 50% der Werte liegen; ab ordinalskalierten Daten

Bsp. Körpergröße:
50% der
Teilnehmer sind
größer als der
Median



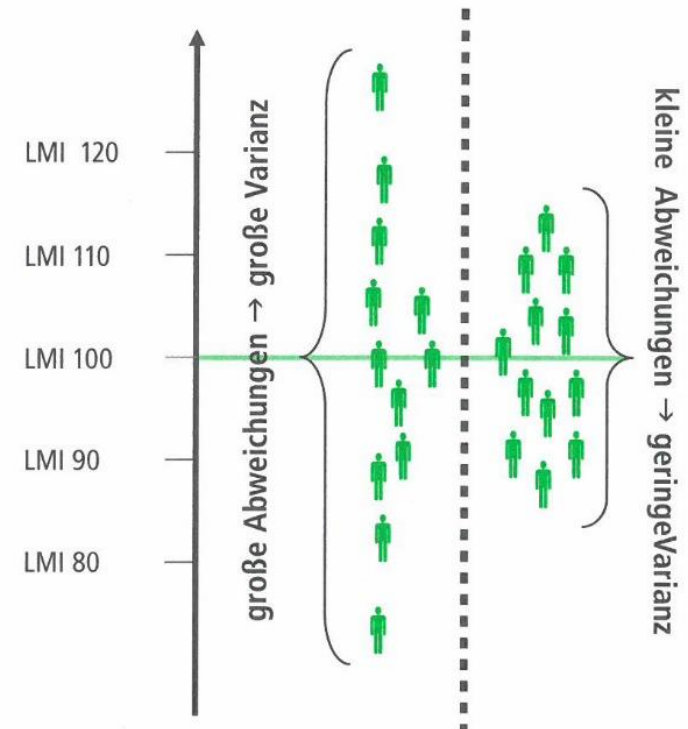
- **Arithmetisches Mittel:** Mittelwert der beobachteten Daten (nur bei metrischen Daten!)

$$AM = \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

6. Auswertung

Deskriptive Statistik: Dispersionsmaße

- **Varianz:** Mittlere quadrierte Streuung der Messwerte um den Mittelwert
 - Stichprobenvarianz: $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$
 - Schätzer für die Varianz der Grundgesamtheit: $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$
- **Standardabweichung:** Wurzel aus der Varianz; = Varianz der Werte in der ursprünglichen Messeinheit
- **Spannweite:** Differenz zwischen dem größten und kleinsten beobachteten Wert

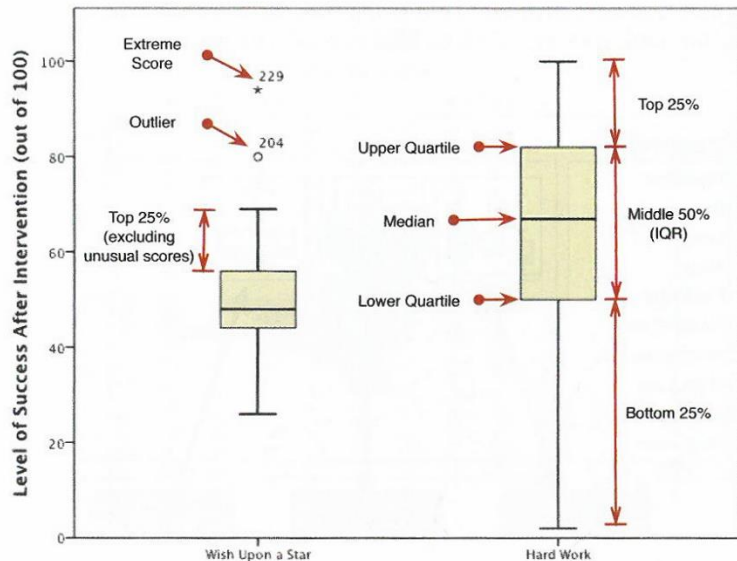


6. Auswertung

Deskriptive Statistik: Ergebnisse aus einem Fahrversuch

	Anzahl Unfälle ohne Handy	Anzahl Unfälle mit Handy
Valid	30	30
Missing	0	0
Mean	1.233	3.200
Median	1.000	3.000
Mode	0.000	3.000
Std. Deviation	1.165	1.919
Minimum	0.000	0.000
Maximum	3.000	6.000

Diagramme helfen, die Daten zu visualisieren



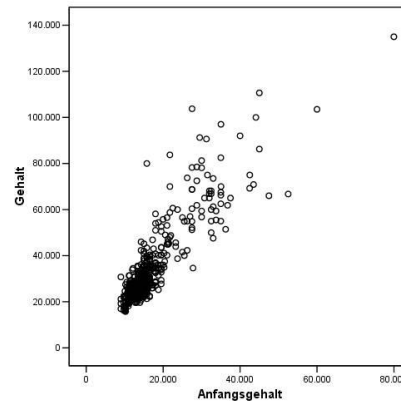
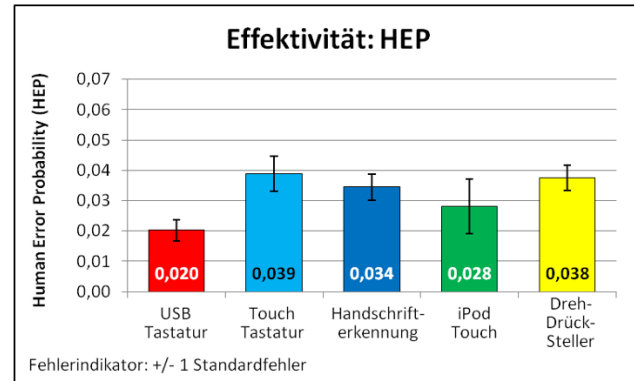
Strategy for Success
Boxplot



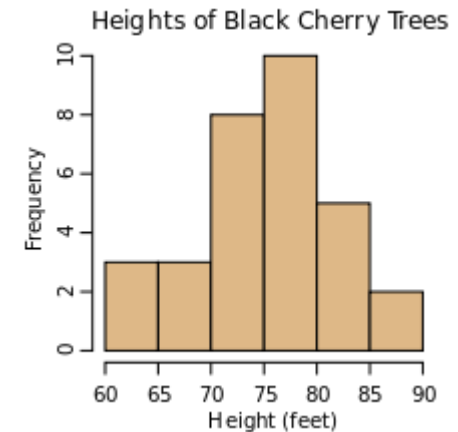
Kreisdiagramm

■ 1. Quartal ■ 2. Quartal ■ 3. Quartal ■ 4. Quartal

Balken-
diagramm



Streudiagramm; für
Zusammenhänge



Histogramm

Diagramme

- Achten Sie darauf, dass alle Achsen beschriftet sind und alle verwendeten Einheiten angegeben sind. Symbole müssen in einer Legende erklärt werden.
- Für jeden berichteten Mittelwert muss einer der folgenden Fehlerbalken angegeben werden: Standardabweichung, Standardfehler, 95%-Konfidenzintervall oder Spannweite. (Angabe des Typs in der Diagrammbeschriftung oder -legende!)
- Mehr dazu im Styleguide des LfE

Konfidenzintervalle

- Form der Intervallschätzung
- Gibt die Genauigkeit der Schätzung an
- Ein $(1-\alpha)100$ % Konfidenzintervall (KI) ist ein Intervall für einen unbekannten Parameter Θ (z. B. Mittelwert), das anhand einer Stichprobe konstruiert wird. Es wird so ausgewählt, dass bei einer Konstruktion einer großen Anzahl dieser Intervalle anhand von unabhängigen Stichproben $(1-\alpha)100$ Prozent der Intervalle den Parameter Θ enthalten. Die Schätzung des Parameters und die Grenzen eines einzigen KIs unterliegen dem Stichprobenfehler (sampling error) und ändern sich deshalb von Stichprobe zu Stichprobe

Konfidenzintervalle

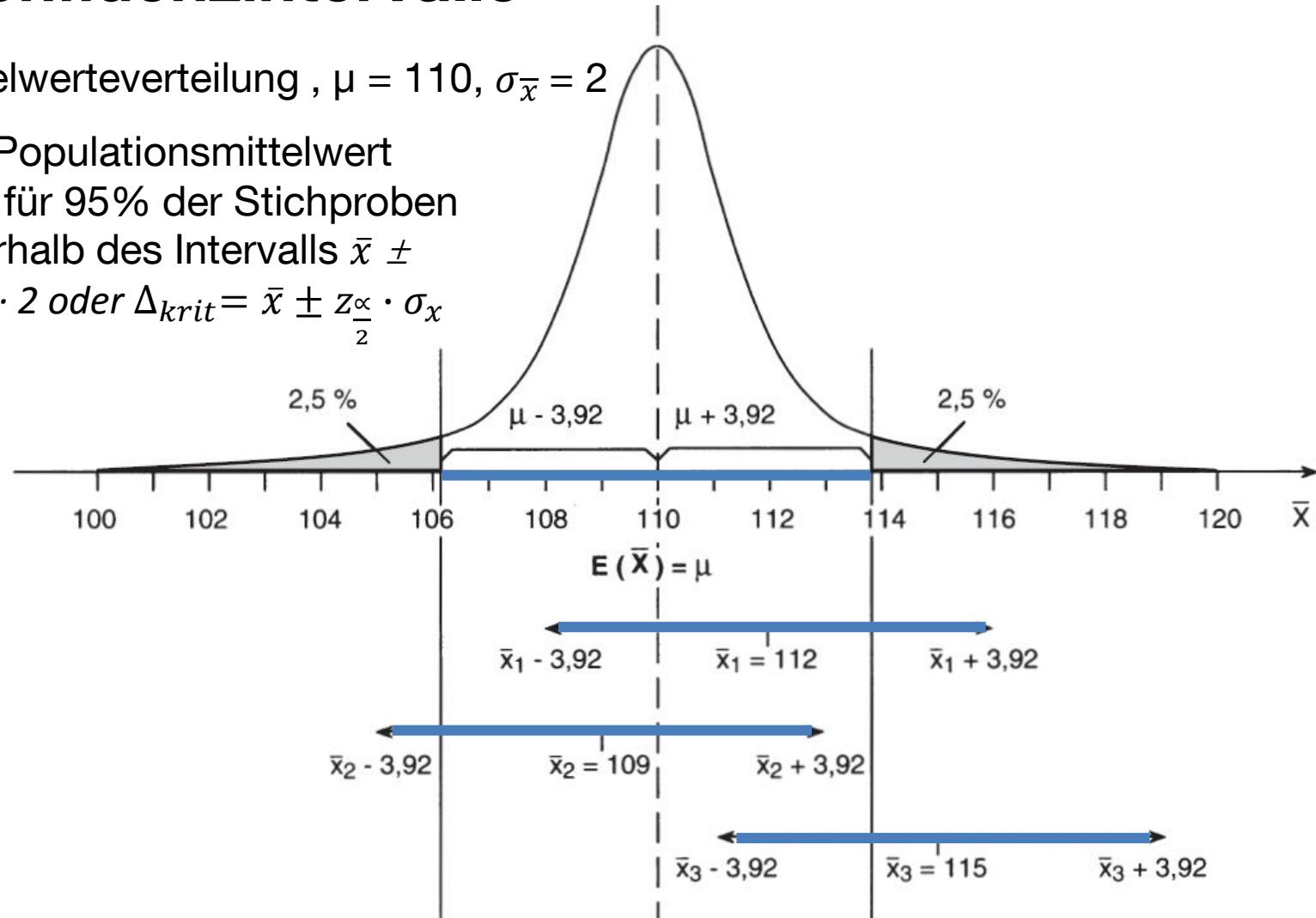
- Prinzip: Verteilung des Parameters und deren Streuung muss bekannt sein
- 2,5 % größte/kleinste Werte der Verteilung bestimmen, per Standardfehler in die ursprüngliche Einheit umrechnen
- Intervallgrenzen um Parameter legen
- Beispiel: Stichprobenmittelwert, Populationsvarianz unbekannt:

$$\left[\bar{x} - t_{\left(1-\frac{\alpha}{2}\right)}(df) \frac{\hat{\sigma}}{\sqrt{n}}; \bar{x} + t_{\left(1-\frac{\alpha}{2}\right)}(df) \frac{\hat{\sigma}}{\sqrt{n}} \right]$$

Konfidenzintervalle

Mittelwerteverteilung , $\mu = 110$, $\sigma_{\bar{x}} = 2$

Der Populationsmittelwert
liegt für 95% der Stichproben
innerhalb des Intervalls $\bar{x} \pm$
 $1.96 \cdot 2$ oder $\Delta_{krit} = \bar{x} \pm z_{\frac{\alpha}{2}} \cdot \sigma_x$



Hypothesentests

- Hypothesen können nur widerlegt (falsifiziert), jedoch nie bewiesen (verifiziert) werden
- Analogie: Hypothese, dass alle Schwäne weiß sind, kann nie verifiziert werden, egal wieviel weiße Schwäne man findet. Findet man jedoch nur einen einzigen schwarzen Schwan, so ist die Hypothese widerlegt.
- Prinzip des Hypothesentests:
 - Es werden zwei Hypothesen formuliert: Null- und Alternativhypothese
 - Diese bilden ein Paar von Aussagen, die sich gegenseitig ausschließen.
 - Das erlaubt es, von der Widerlegung der einen auf die (wahrscheinliche) Gültigkeit der anderen schließen zu können.



Trauerschwan (*Cygnus atratus*)

Hypothesen – Beispiele (ungerichtet)

Die Nullhypothese besagt normalerweise, dass es keinen Effekt gibt, d. h. kein Unterschied oder Zusammenhang

Nullhypothese H_0

„Personen, die während einer Fahrsimulatoraufgabe ein Mobiltelefon nutzen, machen genauso viele Fehler in dieser Aufgabe wie Personen, die kein Mobiltelefon nutzen.“

Die Alternativhypothese stellt die Forschungshypothese dar, welche man annehmen möchte

Alternativhypothese H_1

„Personen, die während einer Fahrsimulatoraufgabe ein Mobiltelefon nutzen, mehr oder weniger Fehler in dieser Aufgabe als Personen, die kein Mobiltelefon nutzen.“

6. Auswertung

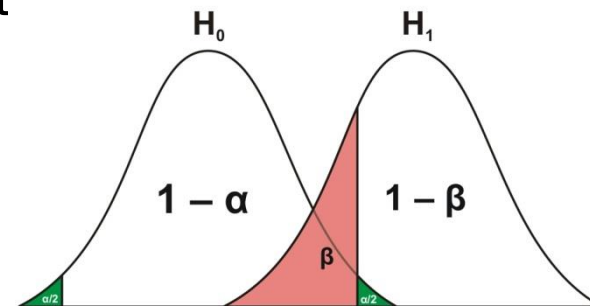
- Um zu einer Entscheidung zu kommen, ob man die H_0 verwirft oder als wahr erachtet, legt man eine Wahrscheinlichkeit fest, ab der man das Auftreten der Daten unter der H_0 als zu unwahrscheinlich sieht
- Konvention: Irrtumswahrscheinlichkeit = Signifikanzniveau = $\alpha = 5\%$
- In diesem Fall: Wenn die Wahrscheinlichkeit, dass die Daten unter Gültigkeit der H_0 (es gibt keinen Unterschied zwischen dem Fahren mit und ohne Mobiltelefon) entstanden sind, unter 5% liegt, dann nimmt man an, dass die H_0 wohl nicht wahr ist

=> H_0 verwerfen, H_1 annehmen

- Die Wahrscheinlichkeit, dass die Daten (oder extremere) unter Gültigkeit der H_0 auftreten nennt man p-Value

6. Auswertung

- Ein Blick auf die Daten zeigt, dass die Probanden mit Mobiltelefon mehr Unfälle gebaut haben. Statistische Tests geben an, wie wahrscheinlich es ist, dass dieses Ergebnis entsteht, wenn es keinen Unterschied gibt, dass heißt, wenn die H_0 wahr ist
- Hier konkret: Wie wahrscheinlich ist es, dass ich in meiner Untersuchung einen Mittelwertsunterschied von 1,96 Unfällen als Ergebnis erhalte, wenn das Mobiltelefon keinen Einfluss hat?
- Wenn die berechnete Wahrscheinlichkeit unter der festgelegten Irrtumswahrscheinlichkeit α , dann wird die H_0 abgelehnt. Das Ergebnis ist statistisch ist signifikant



6. Auswertung

Vorgehen zusammengefasst

1. Statistischer Test ermittelt Wahrscheinlichkeit, mit der die erhaltenen Daten (oder extremere) bei Gültigkeit der H_0 auftreten.
2. Ist diese kleiner als 5% bezeichnen wir das Ergebnis als signifikant, die Daten lassen sich kaum mit der H_0 vereinbaren
3. Verwerfen der H_0 und Annahme der H_1
4. Ist die Wahrscheinlichkeit über 5%, dann spricht nicht genug gegen die H_0 , um diese abzulehnen

6. Auswertung

- Je nach Versuchsdesign, Skalenniveau und Eigenschaften der Daten ergeben sich unterschiedliche Auswertungsmöglichkeiten
- Die statistischen Tests liefern den gesuchten p-Value, den ich zur Entscheidung benötige
- Statistische Tests haben bestimmte Voraussetzungen, welche von den Daten erfüllt werden müssen
- Gängige Tests: t-Test, Varianzanalyse, lineare Regression

6. Auswertung

Die Wahl des statistischen Tests hängt von der Nullhypothese und von den Daten (Skalenniveau, Versuchsplan etc.) ab

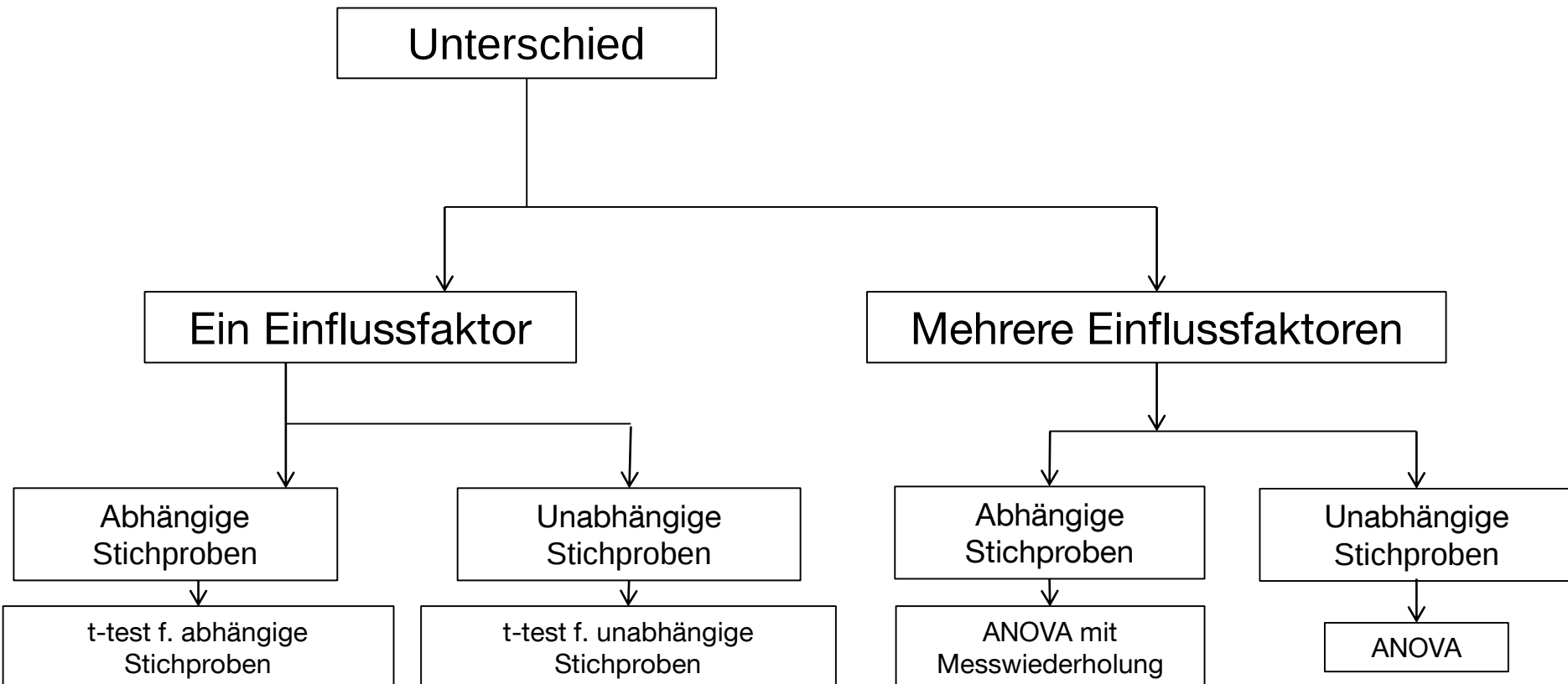
Zusammenhang

Unterschied

6. Auswertung

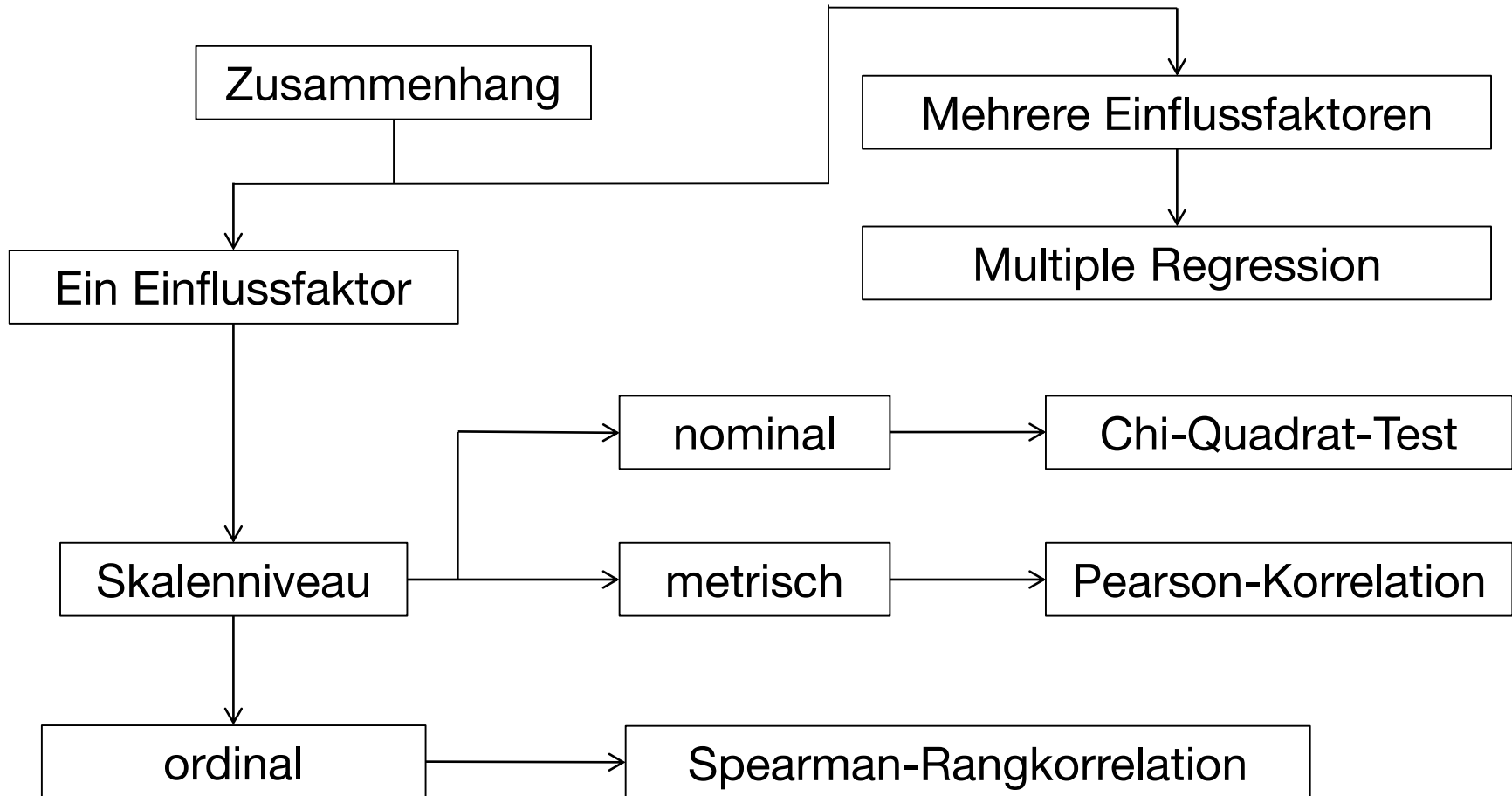
- Wann welcher Test?

Between-Design: unabhängige Stichproben
Within-design: abhängige Stichproben
ANOVA = Varianzanalyse



6. Auswertung

- Wann welcher Test?



6. Auswertung

Hypothese, die in dieser Studie getestet werden soll: „*Personen, die während einer Fahrsimulatorenaufgabe ein Mobiltelefon nutzen, machen genauso viele Fehler in dieser Aufgabe wie Personen, die kein Mobiltelefon nutzen.*“

- Unterschied zwischen zwei Gruppen (mit/ohne Mobiltelefon)
- Within-subject Design
- Metrische abhängige Variable (Fehleranzahl)

Unterschied

Ein Einflussfaktor:
Mobiltelefon

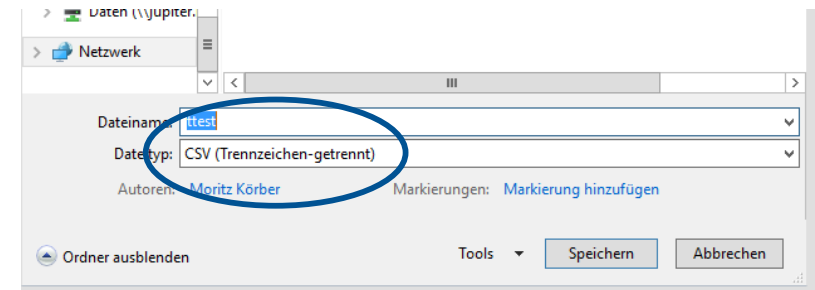
Abhängige Messung
(within-subject Design)

t-test f. abhängige
Stichproben

Datenanalyse mit JASP

Daten in Excel eintragen und als .csv speichern

	A	B	C	D	E	F
1	Anzahl Unfälle ohne Handy	Anzahl Unfälle mit Handy				
2	1	6				
3	1	6				
4	0	3				
5	2	2				
6	2	0				
7	3	2				
8	0	3				
9	2	2				
10	2	2				
11	3	2				
12	2	3				
13	2	4				
14	3	5				
15	2	3				
16	0	3				
17	0	2				
18	2	6				
19	0	5				
20	2	3				
21	1	1				
22	0	1				
23	0	1				
24	0	5				
25	0	6				
26	0	3				
27	3	6				
28	1	0				
29	3	1				
30	0	6				
31	0	4				



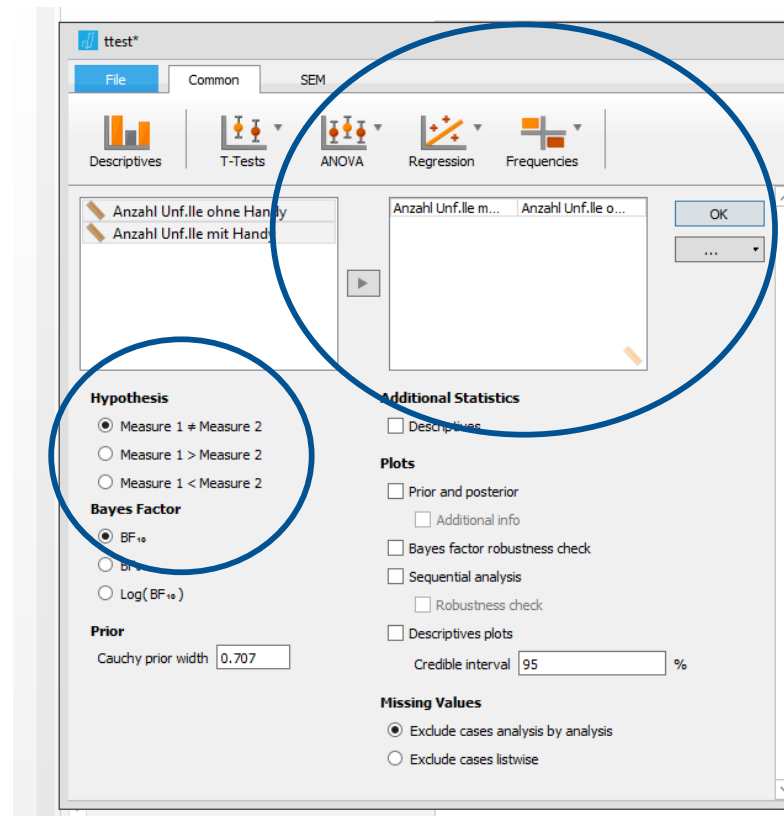
Datenanalyse mit JASP

Unter *File* die Datei öffnen und t-Tests *paired samples t-test* auswählen

	Anzahl Unfälle ohne Handy	Anzahl Unfälle mit Handy
1	1	6
2	1	6
3	0	3
4	2	2
5	2	0
6	3	2
7	0	3
8	2	2
9	2	2
10	3	2
11	2	3
12	2	4
13	3	5
14	2	3
15	0	3
16	0	2

Datenanalyse mit JASP

Die zu untersuchenden Variablen in das rechte Feld kopieren und die passende Alternativhypothese auswählen



Versuchsplanung



7. Interpretation

- Konvention: Irrtumswahrscheinlichkeit = Signifikanzniveau = 5 %
- Testentscheidung nach p-Value (Wahrscheinlichkeit, dass die Daten unter der H_0 entstanden sind)
- Gewähltes Verfahren t-Test mit JASP ausführen
- Ergebnis: p-Value $p < .001$ (kleiner 0,1 %)
 $\Rightarrow H_0$ verwerfen

T-Test ▼

Paired Samples T-Test

			t	df	p	Mean Difference	SE Difference
Anzahl Unf.ile ohne Handy	-	Anzahl Unf.ile mit Handy	-4.572	29	< .001	-1.967	0.430

7. Interpretation

Wir lehnen also die Nullhypothese ab und nehmen die Alternativhypothese an:

Nullhypothese H_0

~~„Personen, die während einer Fahrsimulatoraufgabe ein Mobiltelefon nutzen, machen genauso viele Fehler in dieser Aufgabe wie Personen, die kein Mobiltelefon nutzen.“~~

Alternativhypothese H_1

„Personen, die während einer Fahrsimulatoraufgabe ein Mobiltelefon nutzen, machen mehr oder weniger Fehler in dieser Aufgabe als Personen, die kein Mobiltelefon nutzen.“



7. Interpretation

- In der Diskussion/Interpretation werden die berichteten Ergebnisse kritisch diskutiert und für den Leser „übersetzt“:
 - Was kam raus?
 - Wie sicher kann ich mir mit den Ergebnissen sein?
 - Welche Schlussfolgerungen können aus den Ergebnissen gezogen werden?
- In der Diskussion können Sie zeigen, dass Sie nicht nur auswerten können, sondern auch Ihre Ergebnisse verstehen
- Die Auswertung erfolgt immer im Sinne der Fragestellung; diese soll mit der Auswertung beantwortet werden
- Habe ich mit meiner Auswertung das herausgefunden, was ich mich zu Beginn der Studie gefragt habe?
- Mit Hilfe der statistischen Methoden kann durch die Daten beantwortet werden, ob die Hypothesen durch den Versuch bestätigt oder widerlegt wurden

7. Interpretation

- Signifikanz vs. Bedeutsamkeit
 - Ein Unterschied kann statistisch signifikant, aber praktisch belanglos und somit unbedeutend sein; z. B. ein Unterschied von 15 ms bei einer Gesamtbedienzeit von 10 min
 - Bei genügend großer Stichprobe können selbst minimal kleine Effekt signifikant werden
 - Wichtig: inhaltliche Gewichtung und praktische Bedeutsamkeit!
 - Eine Möglichkeit, Unterschiede (oder Zusammenhänge) auf eine standardisierte Art auszudrücken, bieten Effektgrößen



Effektgrößen

- Ist abhängig von der Studie und der Forschungsfrage:
 - Kann sich um die zentrale Tendenz, Variabilität, Zusammenhänge, Differenzen, Chancen, Verhältnisse, Raten, Dauer, Abweichung, Anteile, ... drehen
 - Wie groß ist der untersuchte Effekt/Unterschied/Zusammenhang?
 - Bsp.: Unterschied in der Bearbeitungsdauer zwischen zwei Bediensystemen -> Effektgröße ist der Unterschied in der Zeit in ms
 - Bsp.: Abnahme von Unfällen nach Übungsfahrt -> Effektgröße ist die absolute oder prozentuale Abnahme der Unfallhäufigkeit
- Ist nicht das Signifikanzniveau/p-Value!
- Unabhängig von der Stichprobengröße

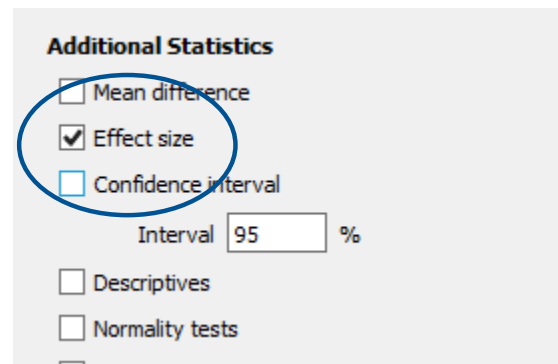
Effektgrößen

- Effektgrößen können standardisiert werden, um sie über Studien hinweg zu vergleichen
- Bekanntestes Maß: Cohens d (bei t-Test)

$$\delta = \frac{\mu_1 - \mu_2}{\sigma}$$

= Unterschied der Mittelwerte im Verhältnis zur Standardabweichung

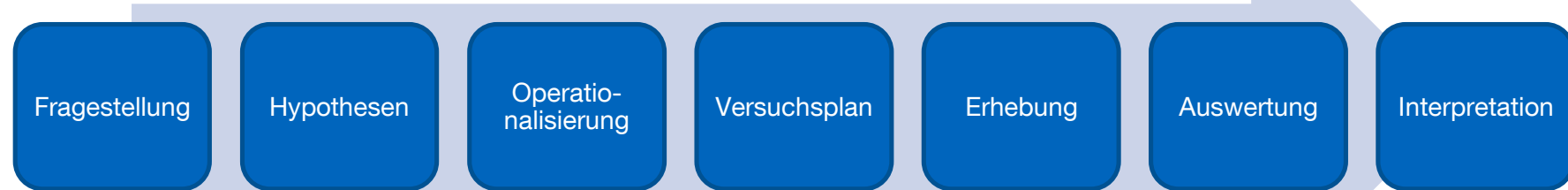
- Cohens d in JASP:



Interpretation und Diskussion

- Fokussieren Sie sich bei der Interpretation nicht auf den p-Value eines Signifikanztests, sondern sehen Sie ihn als eines von mehreren verfügbaren Auswertungsmethoden
- Deskriptive Werte, Konfidenzintervalle, Effektgrößen und eine inhaltliche Gewichtung geben Ihnen und dem Leser wesentlich mehr Informationen und ein detailreicheres Bild

Planungsschritte vor der Untersuchung



- Fragestellung entwickeln: Was will ich herausfinden?
- Wie kann ich Hinweise für eine Antwort auf die Fragestellung finden (Hypothesen aufstellen)?
 - Welche Variablen muss ich dazu erheben und wie kann ich das tun (Operationalisierung)?
 - Between- oder within-Design?
 - Welche Personen muss ich dafür untersuchen? Die Stichprobe muss möglichst repräsentativ sein

Nach dem Versuch:

- Mit welchen statistischen Methoden kann ich die Hypothesen prüfen?
- Wurden die Hypothesen durch die Daten bestätigt oder widerlegt?
⇒ Fragestellung beantworten

Weitere Quellen

- LfE Styleguide für Studienarbeiten

<http://www.lfe.mw.tum.de/lehre/studienarbeiten/styleguide-fuer-studienarbeiten/>

- LfE Richtlinien für Studienarbeiten

<http://www.lfe.mw.tum.de/lehre/studienarbeiten/>

Zur weiteren Vertiefung...

- Statistik

- Eid, M., Gollwitzer, M., & Schmitt, M. (2015). *Statistik und Forschungsmethoden: Mit Online-Materialien* (4., überarb. u. erw. Aufl.). Weinheim: Beltz.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*. 6. vollst. überarb. und akt. Auflage. Heidelberg: Springer Medizin Verlag.
- Bühner, M. & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson Studium.
- Field, A. (2005). *Discovering statistics using SPSS*. (2nd ed.). London: Sage.

- Versuchsplanung

- Huber, O. (2000). *Das psychologische Experiment: Eine Einführung* (3. Aufl., unverändert nach der 2. ergänzten und korr. Aufl.). *Aus dem Programm Huber. Psychologie-Lehrbuch*. Bern, Göttingen, Toronto [etc.]: Huber.

- Wissenschaftliches Schreiben

- *APA Manual (Publication manual of the American Psychological Association)* (6th ed). (2010). Washington, DC: American Psychological Association.
- Kornmeier, M. (2013). *Wissenschaftlich schreiben leicht gemacht: Für Bachelor, Master und Dissertation* (6, aktualisierte und erw. Aufl). *UTB: 3154 : Arbeitshilfen*. Bern [u.a.]: Haupt.