Quantitative Forschungsmethoden

19.10.2022

Tarek Carls

Wer bin ich?

- Tarek Carls
- Studium Psychologie MSc in Regensburg
- Arbeitsaufenthalte in Regensburg (Uni Regensburg) und Los Angeles (UCLA)
- Projektleiter & Wissenschaftlicher Mitarbeiter an der Universität St.Gallen (HSG)

Woran forsche ich?

- PhD General Management (BWL)
- Themenschwerpunkt New Work
- Wie kommen ArbeitnehmerInnen mit dem digitalen Wandel zurecht?
- Welche förderlichen/hinderlichen Faktoren gibt es für gesundes Arbeiten am Arbeitsplatz?

Wie forsche ich?

- Nahezu ausschliesslich quantitativ
- Wir verwenden grösstenteils Längsschnittsdaten
- Statistisches Modellieren
  - SEM (Structural Equation Modeling)
  - Cross Lagged Panel Models
  - GLM
  - Hierarchical Linear Modeling
  - Psychometrie
  - …..

Wie erreichen Sie mich?

- Bei Fragen: Schreiben Sie mir sehr gern!
- LinkedIn: Tarek Carls
- ResearchGate: Tarek Carls
- Mail to: tarek.carls@unisg.ch
- Ich bemühe mich, so schnell wie möglich zu antworten ☺

Lernziele

- Warum tun wir all das überhaupt?
- Zufallsvariablen und Verteilungen
- Stichprobe vs. Population
- T-Tests und ANOVA
- Lineare Regression
- Logistische Regression
- … und was Sie noch interessiert?

Agenda

02.11.21 – Wiederholung / Population & Stichprobe

08.11.21 - t-Tests und ANOVA (Analysis of Variance)

15.11.21 – Lineare Regression & Logistische Regression

18.11.21 – Anwendung in SPSS

22.11.21 – Anwendung in SPSS

27.11.21 - Wiederholung

Warum ist all das wichtig?

Statistische Methodik als wichtigstes Handwerkzeug für Forschende und Praktizierende

Fundierte Methodenkenntnisse am Arbeitsmarkt sehr geschätzt

Besseres Verständnis von Wissenschaft im Allgemeinen

Wissenschaftliche Standards (Replication crisis)

Kapitel 1 - Wiederholung

- Diskrete/stetige Zufallsvariablen
- Population und Stichprobe

Zufallsexperiment

= Vorgänge, die sich beliebig oft wiederholen lassen, deren Ausgang aber *nicht vorhersagbar* ist (Werfen einer Münze, eines Würfels etc)

Ergebnismenge = potentielle Ausgänge eines Zufallsexperiments (Ω, sprich: *Omega*)

$$\Omega_{Würfel} = \{1, 2, 3, 4, 5, 6\}$$
$$\Omega_{Münze} = \{Kopf, Zahl\}$$

Zufallsvariable

Eine Zufallsvariable ist eine Abbildung dieser Ergebnismenge Ω. Diese Zufallsvariable kann **abzählbar viele** Ausprägungen haben (diskrete ZV) oder **unendlich viele** Ausprägungen haben (stetige ZV).

$$X\colon\ \Omega \to \Omega'$$

Was sind gute Beispiele für diskrete und stetige Zufallsvariablen?

Wahrscheinlichkeits-funktion

… für **diskrete** Zufallsvariablen:

Betrachten wir das Werfen zweier Münzen

Zufallsvariable			Wahrscheinlichkeitsfunktion
Ergebnis ω aus Ω		zugeordnete Zahl in Ω'	Wahrscheinlichkeit des Auftretens des Elementes in Ω'
(W,W)	$\rightarrow$	0	0.25
(W,Z)	$\rightarrow$	1	0.5
(Z,W)	$\rightarrow$	1	
(Z,Z)	$\rightarrow$	2	0.25

Dichtefunktion

Wie sieht die Wahrscheinlichkeitsfunktion für stetige Zufallsvariablen aus?

Beispiel: Messen der Körpergrösse eines zufällig gezogenen Menschen

Wahrscheinlichkeitsfunktion einer stetigen Zufallsvariable ist die Dichtefunktion

→ Die konkrete W'keit eines einzelnen Ereignisses einer stetigen Zufallsvariable ist immer 0. Wahrscheinlichkeiten beziehen sich daher auf einen bestimmten Bereich, nicht auf einen bestimmten Wert!

Dichtefunktion

Die Wahrscheinlichkeit, dass ein Wert aus einem Wertebereich stammt, lässt sich durch das Integral der Dichtefunktion in diesem Bereich beschreiben.

$$P(i \leq x \leq j) = \int_{i}^{j} f(x)dx$$

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

Verteilungsfunktion

Ein Wert der Verteilungsfunktion F(x) gibt die aufsummierte Wahrscheinlichkeit aller Werte, die **kleiner oder gleich x** sind.

$$\lim_{\to +\infty} F(x) = 1$$
$$\lim_{\to -\infty} F(x) = 0$$

Für diskrete Zufallsvariablen:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

Für stetige Zufallsvariablen:

$$G(x) = P(X \leq x) = \int_{x_i \leq x} g(x_i)$$

Verteilungsfunktion

Was wäre hier der Wert der Verteilungsfunktion?

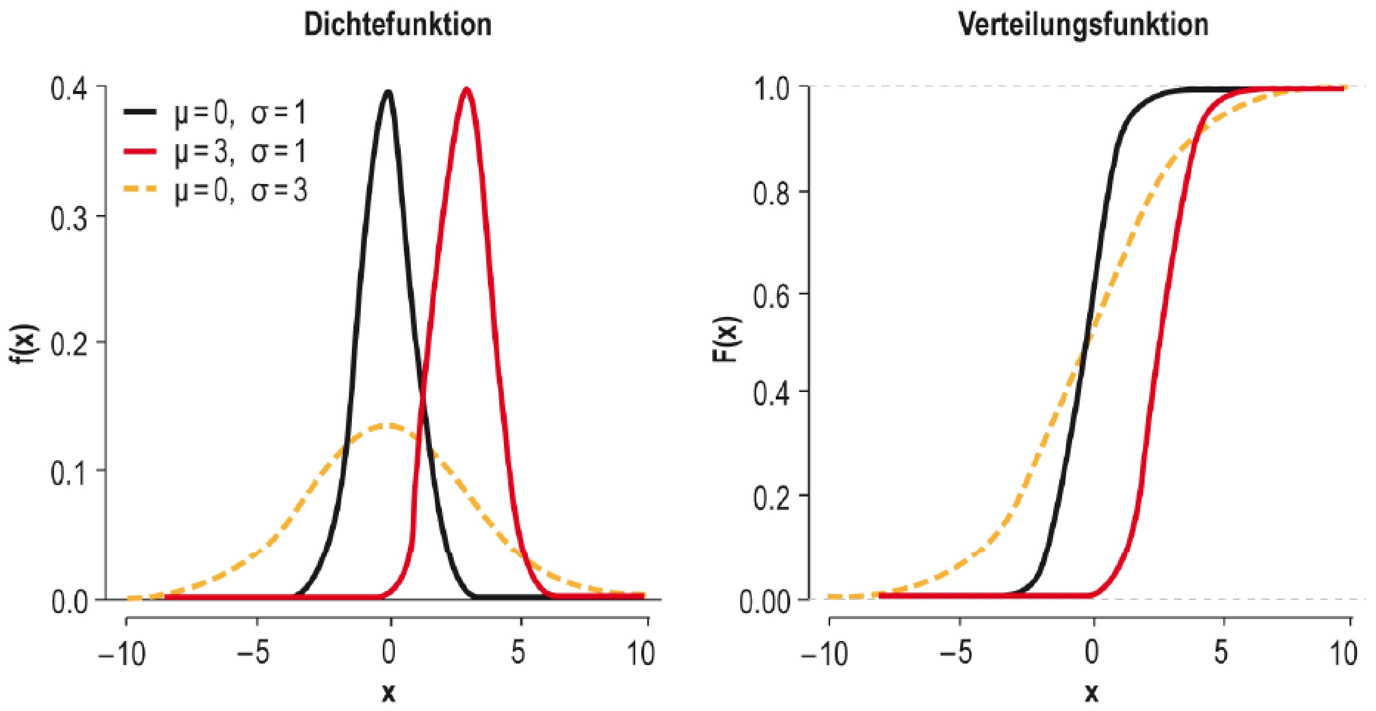
Für

$$F(0) =$$
$$F(1) =$$
$$F(2) =$$

Zufallsvariable			Wahrscheinlichkeitsfunktion
Ergebnis ω aus Ω		zugeordnete Zahl in Ω'	Wahrscheinlichkeit des Auftretens des Elementes in Ω'
(W,W)	$\rightarrow$	0	0.25
(W,Z)	$\rightarrow$	1	0.5
(Z,W)	$\rightarrow$	1	
(Z,Z)	$\rightarrow$	2	0.25

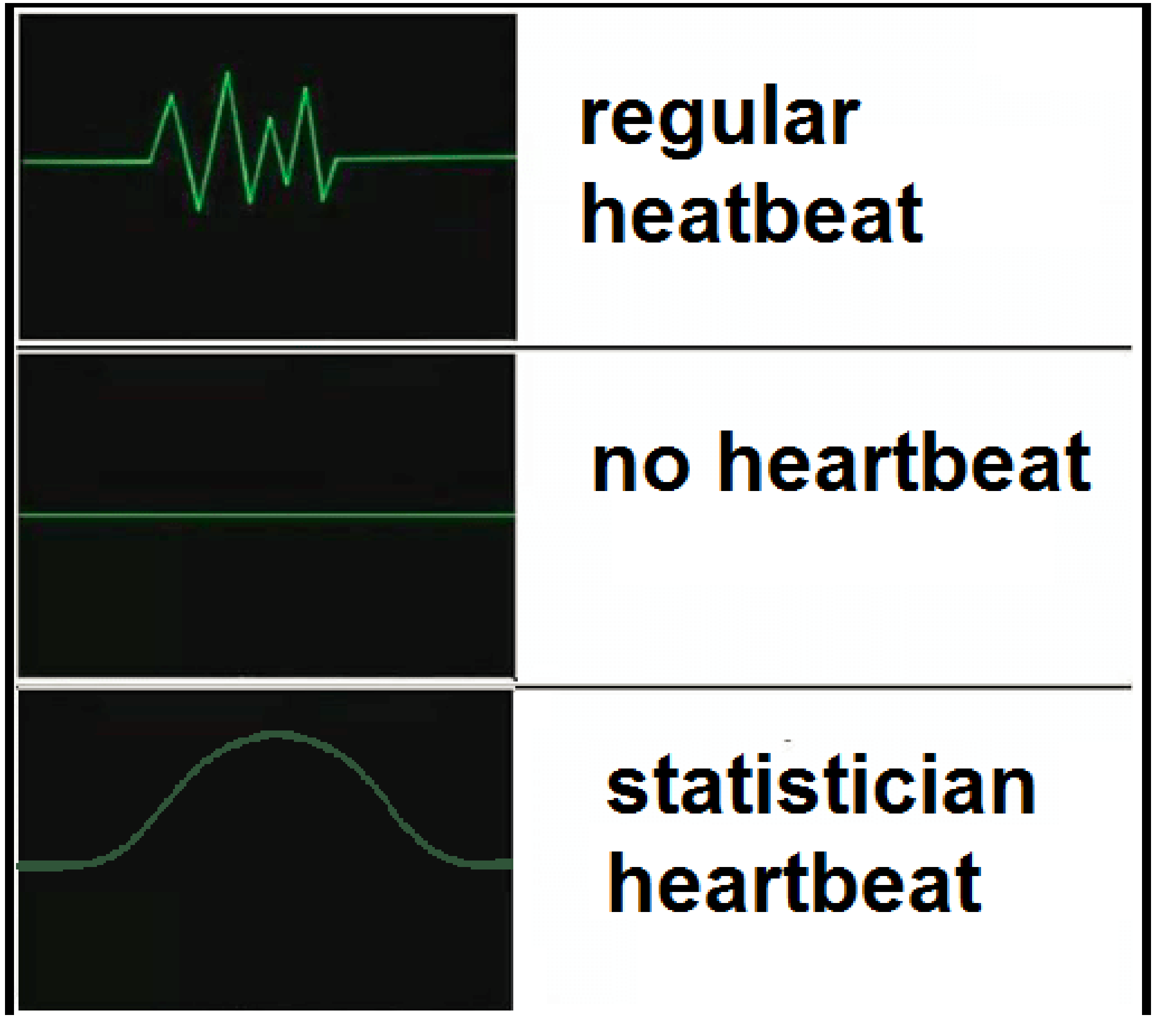
Verteilungsfunktion vs. Dichtefunktion

- Verteilungsfunktion: $\int_{-\infty}^{\infty} f(x)dx = \infty$
- Dichtefunktion: $\int_{-\infty}^{\infty} f(x)dx = 1$



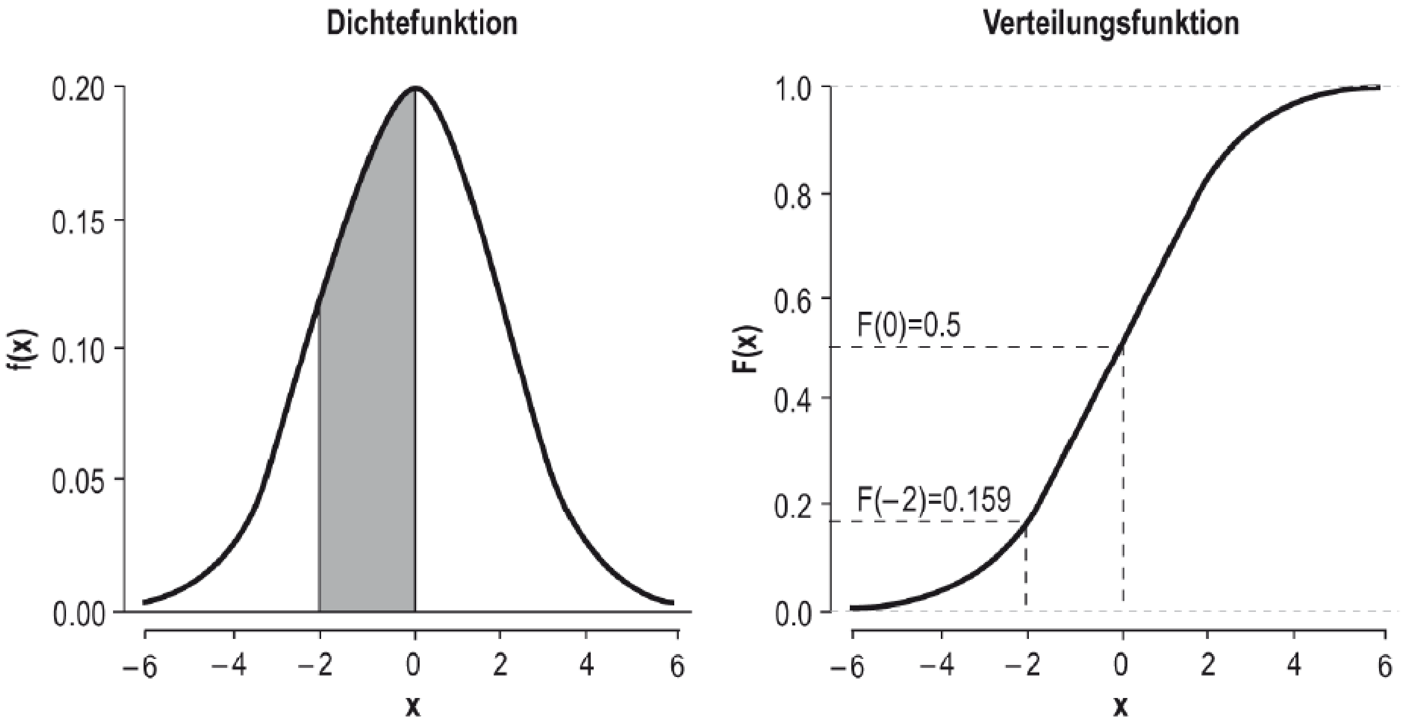
Die Normalverteilung

- $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$
- Definiert durch zwei Parameter:
  - Erwartungswert μ *($\mu \in \mathbb{R}$), sprich: my*
  - Standardabweichung σ $(\sigma > 0)$, sprich: sigma
- Die Wahrscheinlichkeit, dass ein Wert aus einem bestimmten Wertebereich stammt, ist definiert durch die Fläche unter der Funktion in dem entsprechenden Wertebereich



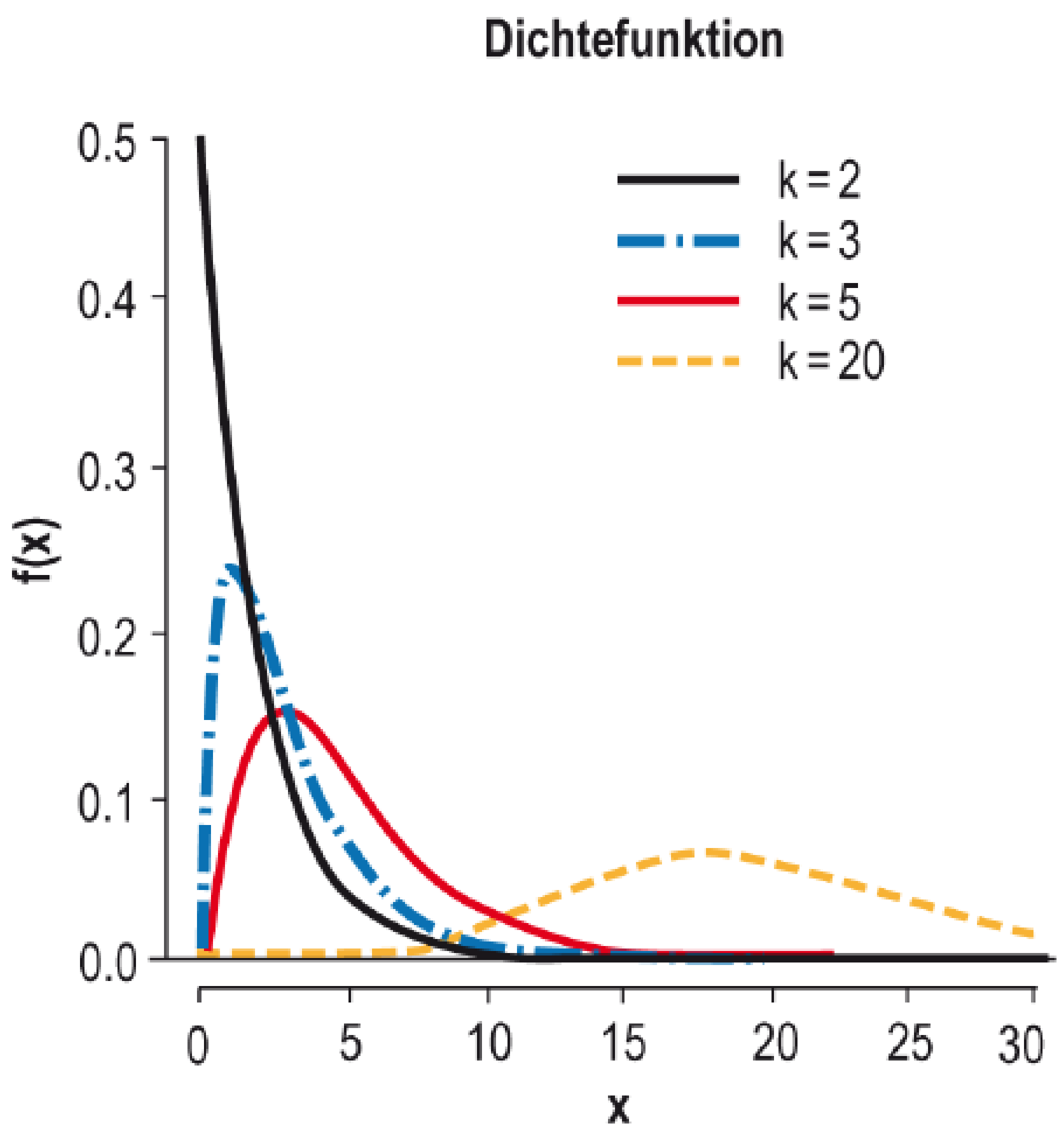
Die Normalverteilung

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



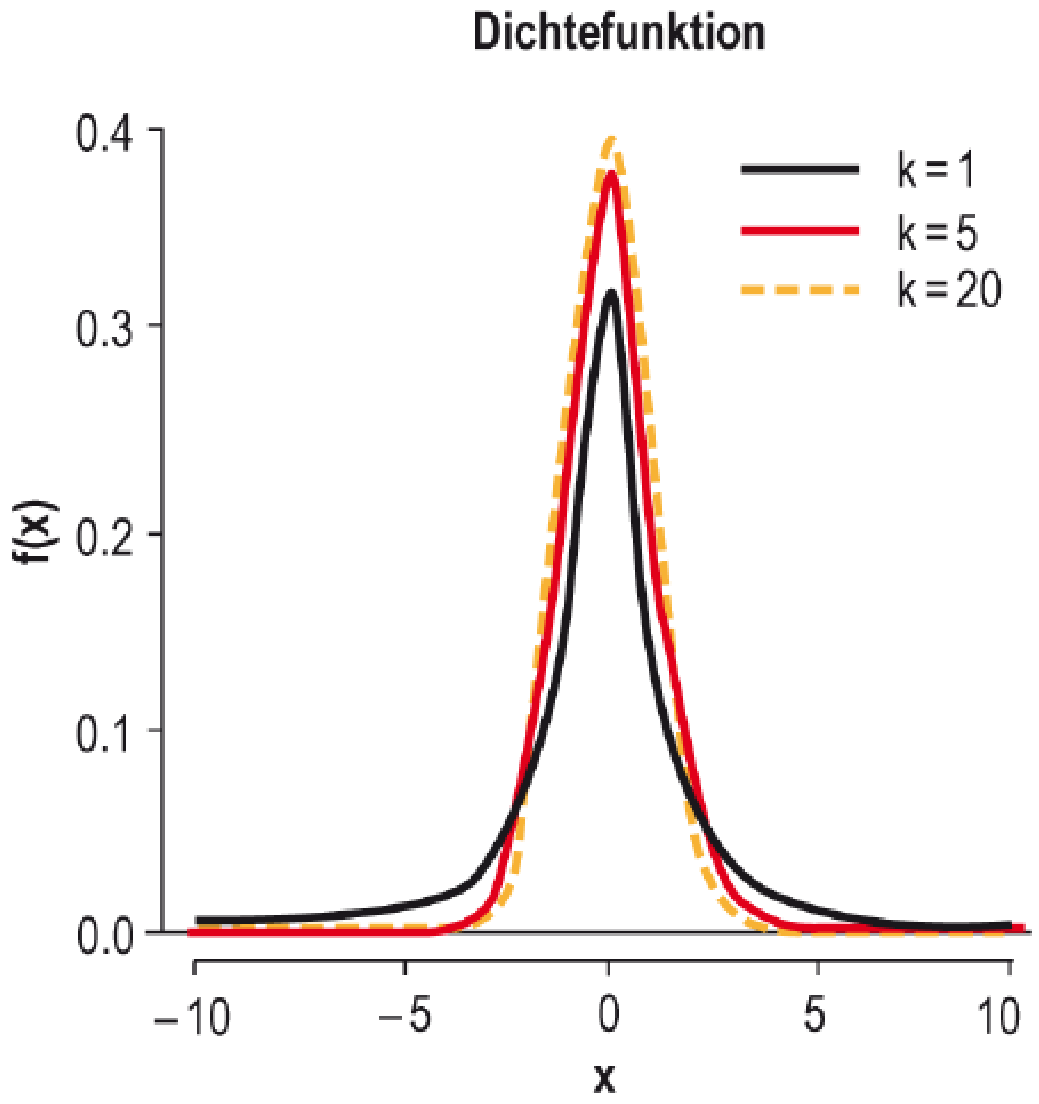
Die χ^2-Verteilung

- Eine χ^2-verteilte („Chi-Quadrat“) Zufallsvariable mit k Freiheitsgraden ergibt sich durch Addieren von k quadrierten standard-normalverteilten Zufallsvariablen.
- $\chi^2 = X_1^2 + X_2^2 + \cdots + X_k^2$
- Prüfstatistik beim χ^2-Test



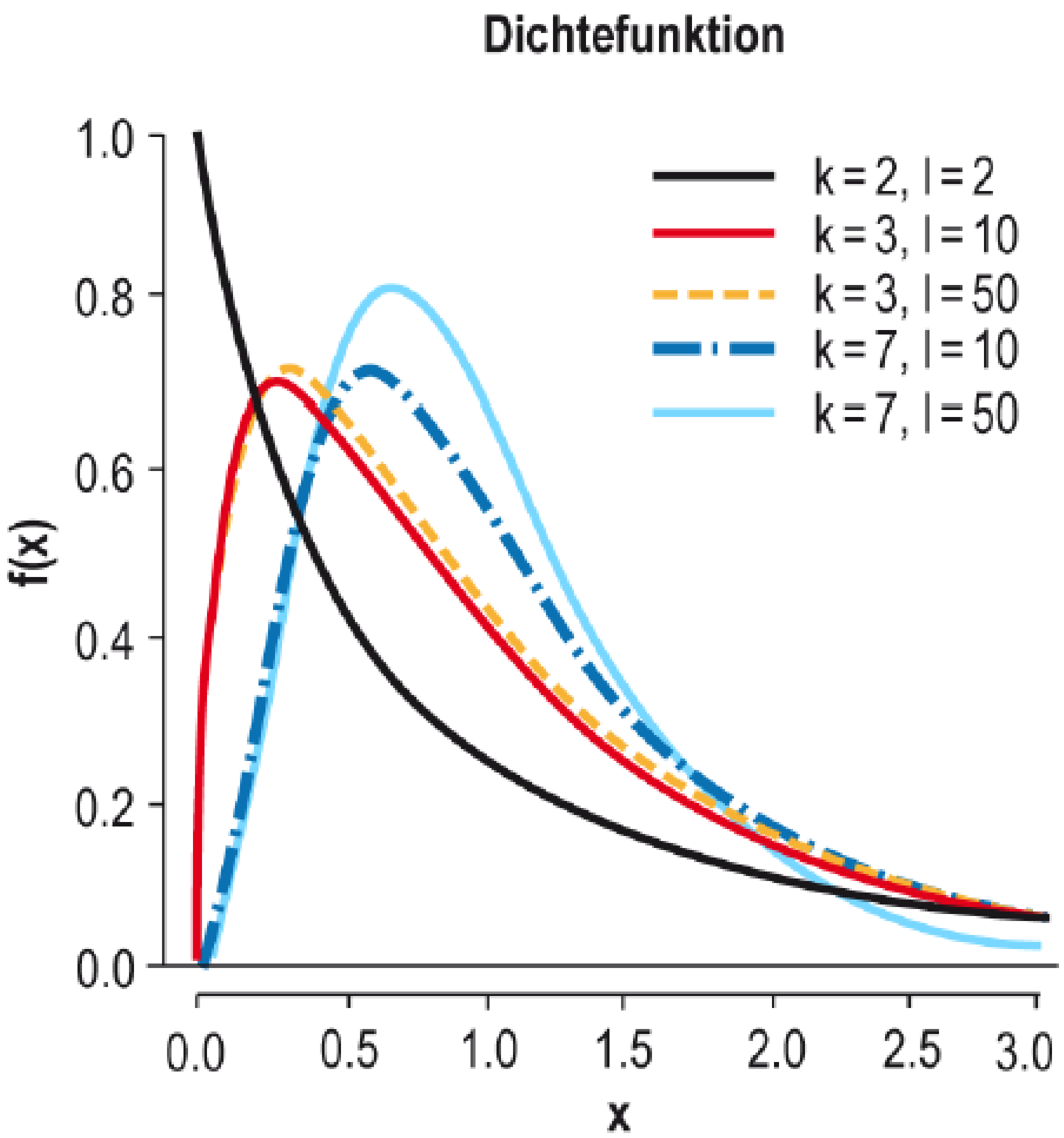
Die t-Verteilung

- Die t-Verteilung ergibt sich aus dem Zusammenspiel einer standardnormalverteilten Zufallsvariablen X und einer mit k Freiheitsgraden χ^2-verteilten Zufallsvariablen $Y \sim \chi_k^2$.
- $T = \frac{X}{\sqrt{\frac{1}{k}Y}}$
- Prüfstatistik beim t-Test



Die F-Verteilung

- $F = \frac{\frac{X}{k}}{\frac{Y}{l}}$
- Die F-Verteilung ergibt sich aus dem Zusammenspiel zweier voneinander unabhängiger und χ^2-verteilter Zufallsvariablen X und Y mit k bzw. l Freiheitsgraden
- Prüfstatistik für die ANOVA



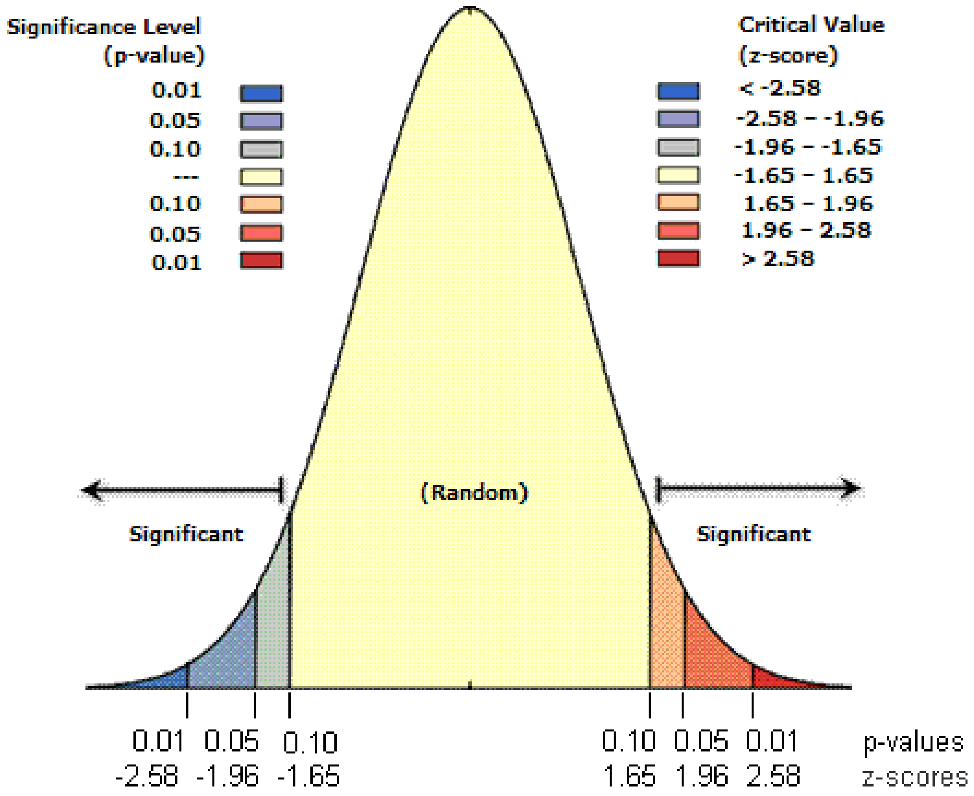
Konfidenzintervalle

- Bisher haben wir viel über Punktschätzer geredet
- Aber: Ein Punktschätzer allein bietet wenig Auskunft über die Präzision der Schätzung
- Sinnvolle Ergänzung zu einem Punktschätzer: das Konfidenzintervall
- Übliche Schreibweise: $[T \pm c \cdot SE_T]$
- T ist unser Schätzwert
- c ist der «Sicherheitsparameter»
- SE ist die Standardabweichung

Konfidenzintervalle

$$\left[M_X \pm t_{n-1;\frac{\alpha}{2}} \cdot \frac{\hat{S}_X}{\sqrt{n}}\right]$$

- Mittelwert berechnen
- Standardabweichung berechnen
- T-Wert ablesen (siehe Anhang)
- Formel anwenden



Konfidenzintervalle: Übungsaufgabe

Aufgabe 2: Statistisches Testen

Es ist ein weit verbreitetes Klischee, dass in Bayern mehr Alkohol konsumiert wird als in anderen Bundesländern. Im Rahmen eines Forschungsprojekts für Suchtprävention soll diese These wissenschaftlich untersucht werden. Sie veranstalten deshalb zwei identische Partys in München und Hamburg und messen am Ende des Abends den Atemalkoholgehalt (in mg/l) bei von je 10 Party-Besucher*innen.

Nehmen Sie im Weiteren an, dass das Merkmal Atemalkoholgehalt normalverteilt ist.

Atemalkoholgehalt in München (M)	0.5	0.7	1.2	0.0	1.5	0.6	0.3	0.2	0.1	0.9
Atemalkoholgehalt in Hamburg (H)	0.4	0.8	1.1	1.2	0.0	0.3	0.4	0.2	0.1	0.1

a) Schätzen Sie ein 95%-Konfidenzintervall für den Erwartungswert des Atemalkoholgehalts bei den Münchner Party-Besucher:innen (t-Verteilung im Anhang).

a) Schätzen Sie ein 95%-Konfidenzintervall für den Erwartungswert des Atemalkoholgehalts bei den Münchner Party-Besucher:innen. => SB 1; Statistisches Testen, S.46

$$\left[M_X \pm t_{n-1;\frac{\alpha}{2}} \cdot \frac{\hat{S}_X}{\sqrt{n}}\right]$$

- Mittelwert berechnen
- Standardabweichung berechnen
- T-Wert ablesen (siehe Anhang)
- Formel anwenden

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{10}(0,5+0,7+\ldots+0,9) = 0,6$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{1}{10-1}((0,5-0,6)^2+\ldots+(0,9-0,6)^2) = 0,24$$

$$t(0,975;9) = 2,262$$

$$KI = 0,6 \pm 2,262\frac{\sqrt{0,24}}{\sqrt{10}} \Rightarrow [0,6-0,35;\ 0,6+0,35] = \mathbf{[0,25;\ 0,95]}$$

n \ α	0,10 0,20	0,05 0,10	0,025 0,05	0,01 0,02	0,005 0,01	0,001 0,002	0,0005 0,001	einseitig zweiseitig
1	3,078	6,314	12,71	31,82	63,66	318,3	636,6	
2	1,886	2,920	4,303	6,965	9,925	22,33	31,56	
3	1,638	2,353	3,182	4,541	5,841	10,22	12,92	
4	1,533	2,132	2,776	3,747	4,604	7,173	8,610	
5	1,476	2,015	2,571	3,365	4,032	5,893	6,869	
6	1,440	1,943	2,447	3,143	3,707	5,208	5,959	
7	1,415	1,895	2,365	2,998	3,499	4,785	5,408	
8	1,397	1,860	2,306	2,896	3,355	4,501	5,041	
9	1,383	1,833	2,262	2,821	3,250	4,297	4,781	
10	1,372	1,812	2,228	2,764	3,169	4,144	4,587	
11	1,363	1,796	2,201	2,718	3,106	4,025	4,437	
12	1,356	1,782	2,179	2,681	3,055	3,930	4,318	
13	1,350	1,771	2,160	2,650	3,012	3,852	4,221	
14	1,345	1,761	2,145	2,624	2,977	3,787	4,140	
15	1,341	1,753	2,131	2,602	2,947	3,733	4,073	
16	1,337	1,746	2,120	2,583	2,921	3,686	4,015	
17	1,333	1,740	2,110	2,567	2,898	3,646	3,965	
18	1,330	1,734	2,101	2,552	2,878	3,610	3,922	
19	1,328	1,729	2,093	2,539	2,861	3,579	3,883	
20	1,325	1,725	2,086	2,528	2,845	3,552	3,850	
21	1,323	1,721	2,080	2,518	2,831	3,527	3,819	
22	1,321	1,717	2,074	2,508	2,819	3,505	3,792	
23	1,319	1,714	2,069	2,500	2,807	3,485	3,768	

Stichprobe und Population

Inferenzstatistik

→ Was ist das?

2.1 Das Verhältnis von Stichprobe zu Population

- Sie vergleichen die Ergonomie zweier Produkte miteinander.
- 10 VPn bewerten die Ergonomie jeweils eines Produktes auf einer Skala von 1 – 10.
- $M_A = 5.6,\ M_B = 6.3$
- Welches Produkt ist ergonomischer?

Produkt A	Produkt B
6	7
5	6
7	5
1	6
6	4
7	6
5	10
7	6
6	6
6	7

2.1 Das Verhältnis von Stichprobe zu Population

- Welches Produkt ist ergonomischer?
- Jeweils zwei starke Ausreisser in den Gruppen
- Wenn wir diese entfernen, dreht sich das Bild um
- $M_A = 6.1,\ M_B = 5.89$

Produkt A	Produkt B
6	7
5	6
7	5
1	6
6	4
7	6
5	10
7	6
6	6
6	7

2.1 Das Verhältnis von Stichprobe zu Population

- Das tatsächliche Ergebnis einer Studie ist immer von der zufälligen Ziehung der Stichprobe abhängig
- Hätten wir nur minimal andere Menschen befragt, wären wir zu einem anderen Ergebnis gekommen
- Wir sind als Forschende NICHT an Aussagen über die Stichprobe interessiert, sondern an Aussagen über die Population
  - «Bewerten **diese 10 Personen** die Ergonomie von Produkt A höher als von Produkt B?»
  - «Bewerten **alle Personen** die Ergonomie von Produkt A im Mittel höher als von Produkt B?»
- Mittelwert der Stichprobe dient als **Schätzwert** für den Populationsmittelwert

2.2 Parameter vs. Statistiken: Punktschätzung

- Stichprobenstatistik → kennen wir
- Populationsparameter → kennen wir nicht
- Stichprobe:
  - Stichprobenmittelwert M
  - Stichprobenvarianz S^2
- Population:
  - Populationsmittelwert μ
  - Populationsvarianz σ^2

2.2.1 Der Populationsmittelwert μ

- Population von fünf Elementen mit den Werten 2, 4, 6, 6, 7
- $\mu = \frac{2+4+6+6+7}{5} = 5$
- $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{(2-5)^2+(4-5)^2+(6-5)^2+(6-5)^2+(7-5)^2}{5} = 3.2$
- (Frage: Wie unterscheidet sich die Berechnung der Varianz der Stichprobe von der Berechnung der Varianz der Population?)
- Aus dieser Population ziehen wir nun eine Stichprobe mit $n = 2$

2.2.1 Der Populationsmittelwert μ

- Frage 1: Ist diese Zufallsvariable stetig oder diskret?
- Frage 2: Welche Wahrscheinlichkeit hat ein einzelnes Ereignis (eine Zelle)?

		Werte der Populationselemente				
		2	**4**	**6**	**6**	**7**
Werte der Populationselemente	**2**	2	3	4	4	4.5
	4	3	4	5	5	5.5
	6	4	5	6	6	6.5
	6	4	5	6	6	6.5
	7	4.5	5.5	6.5	6.5	7

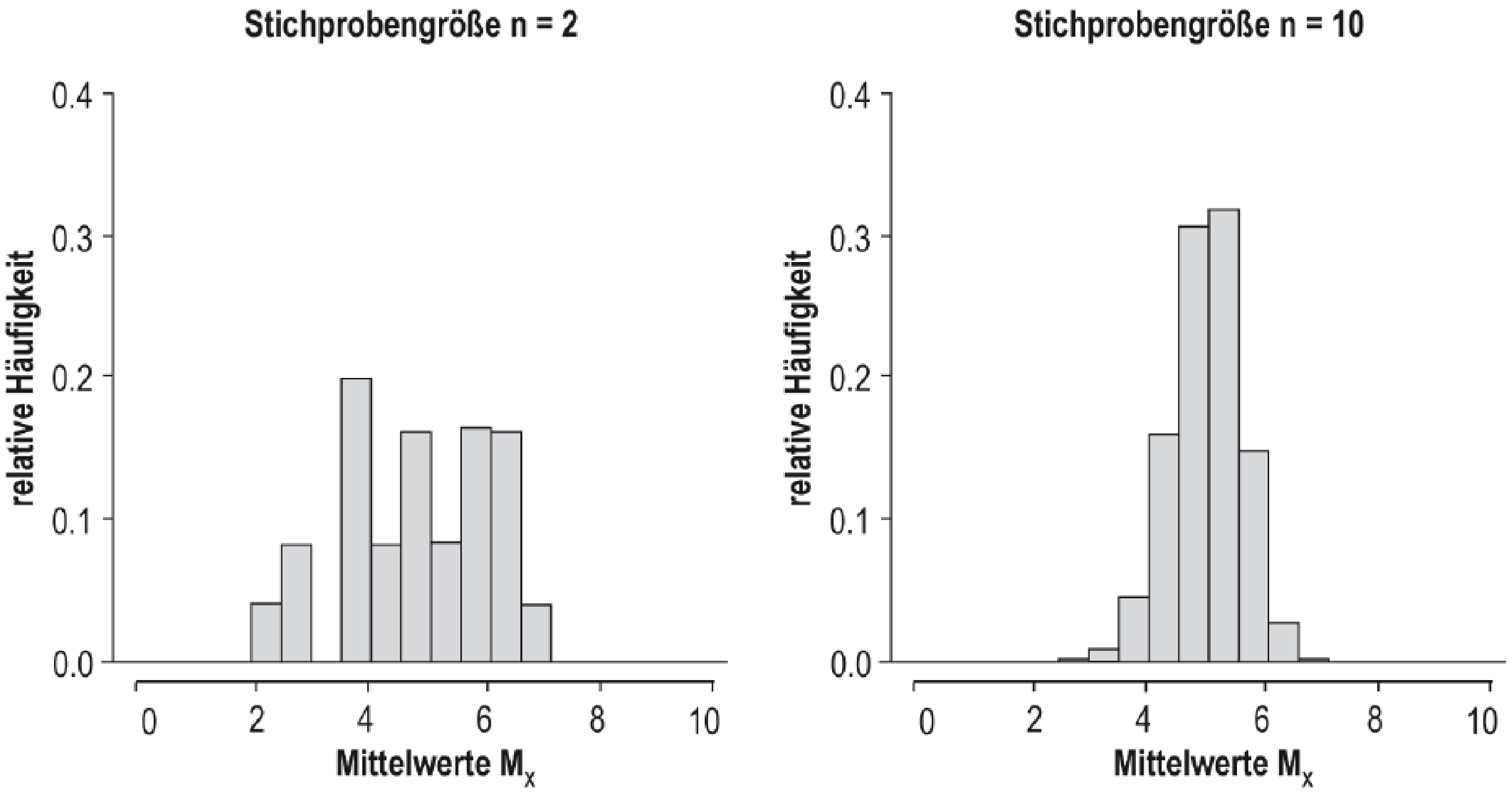
2.2.1 Der Populationsmittelwert μ

- Wahrscheinlichkeitsfunktion für $\bar{X}$
- Erwartungswert von $\bar{X}$: $E(\bar{X}) = \mu$
- Varianz von $\bar{X}$: $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$

M_X	2	3	4	4.5	5	5.5	6	6.5	7
$P(\bar{X} = M_X)$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{5}{25}$	$\frac{2}{25}$	$\frac{4}{25}$	$\frac{2}{25}$	$\frac{4}{25}$	$\frac{4}{25}$	$\frac{1}{25}$

2.2.1 Der Populationsmittelwert μ

Was sehen wir hier? Woran nähert sich das an?



2.2.1 Der Populationsmittelwert μ

- Standardfehler des Mittelwerts ergibt sich aus der Wurzel der Varianz der Zufallsvariablen
- $SE = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$
- Die Mittelwerte der Stichprobe liegen um den «wahren» Populationsmittelwert μ herum, wobei Werte nahe μ häufiger sind als Ausreisser

2.2.2 Gütekriterien für Schätzer

- Wir gehen davon aus, dass der Erwartungswert $E(\bar{X})$ ein guter Schätzer für den Populationsmittelwert μ ist. Aber warum?
- Gute Schätzer definieren sich durch zwei Eigenschaften: Erwartungstreue und Konsistenz
- **Erwartungstreue**: $E(T) = \tau$
- **Konsistenz**: Die W'keit, dass der Schätzer nah am zu schätzenden Parameter liegt, steigt mit der Stichprobengrösse

2.2.3 Die Populationsvarianz

- Ist die Stichprobenvarianz ein erwartungstreuer Schätzer für die Populationsvarianz?
- $E(S^2) \neq \sigma^2$
- $E(S^2) = \frac{n-1}{n}\sigma^2$
- $\frac{\sum_{i=1}^{n}(x_i - M)^2}{n-1}$
- In der Anwendung wird nahezu immer der erwartungstreue Schätzer verwendet (Division durch n-1)

2.3 Arten statistischer Hypothesen

- Uns interessieren nicht die Stichproben, sondern die Population. Deswegen brauchen wir Hypothesen
  - Unterschiedshypothesen
  - Zusammenhangshypothesen
  - Gerichtete Hypothesen
  - Ungerichtete Hypothesen
  - Nullyhpothese
  - Alternativhypothese

2.3.1 Unterschiedshypothesen

- Beispiel: «Produkt A unterscheidet sich hinsichtlich seiner wahrgenommenen Ergonomie von Produkt B»
  - Was ist hier die Nullhypothese, was ist die Alternativhypothese?
- Grundsätzlich: Nullhypothese meist «das Gegenteil von dem, was wir zeigen wollen»
- Ungerichtete Hypothese:
  - $Nullhypothese\ H_0{:}\ \mu_A = \mu_B$ (Die Ergonomie beider Produkte unterscheidet sich nicht)
  - $Ungerichtete\ Alternativhypothese\ H_1{:}\ \mu_A \neq \mu_B$ (Die Ergonomie beider Produkte unterscheidet sich)

2.3.1 Unterschiedshypothesen

- Sie sind Produktdesigner. Sie haben das ursprüngliche Produkt (Produkt A) hinsichtlich seiner Ergonomie verbessert und einen Nachfolger entwickelt (Produkt B)
- Gerichtete Hypothese:
  - $Nullhypothese\ H_0\colon\ \mu_A \geq \mu_B$ (Die Ergonomie von Produkt B ist kleiner als die von Produkt A oder beide Produkte unterscheiden sich nicht hinsichtlich ihrer Ergonomie)
  - $Gerichtete\ Alternativhypothese\ H_1\colon\ \mu_A < \mu_B$ (Die Ergonomie von Produkt B ist grösser als die von Produkt A)

2.3.2 Zusammenhangshypothesen

- ….formulieren einen Zusammenhang von zwei oder mehreren Variablen in einer Population
- Korrelation
  - …auf Populationsebene: ρ (sprich: *rho*)
  - …auf Stichprobenebene: r
- Beispiel: Wir untersuchen den Zusammenhang von Tenure (Dauer Organisationszugehörigkeit) und Jahreseinkommen.
  - Gerichtete oder ungerichtete Hypothese?
  - Gerichtete Hypothese:
    - $H_1: \rho > 0$
    - $H_0: \rho \leq 0$
  - Fallen Ihnen Beispiele für eine ungerichtete Zusammenhangshypothese ein?

2.3.2 Zusammenhangshypothesen

- **Beispiel**: Wir untersuchen den Zusammenhang von Körpergrösse und Extraversion.
  - Ungerichtete Hypothese:
    - $H_1: \rho \neq 0$
    - $H_0: \rho = 0$

2.4. Logik des Nullhypothesen-Signifikanztests

Wie funktioniert ein Signifikanztest?

2.4. Logik des Nullhypothesen-Signifikanztests

Wie funktioniert ein Signifikanztest?

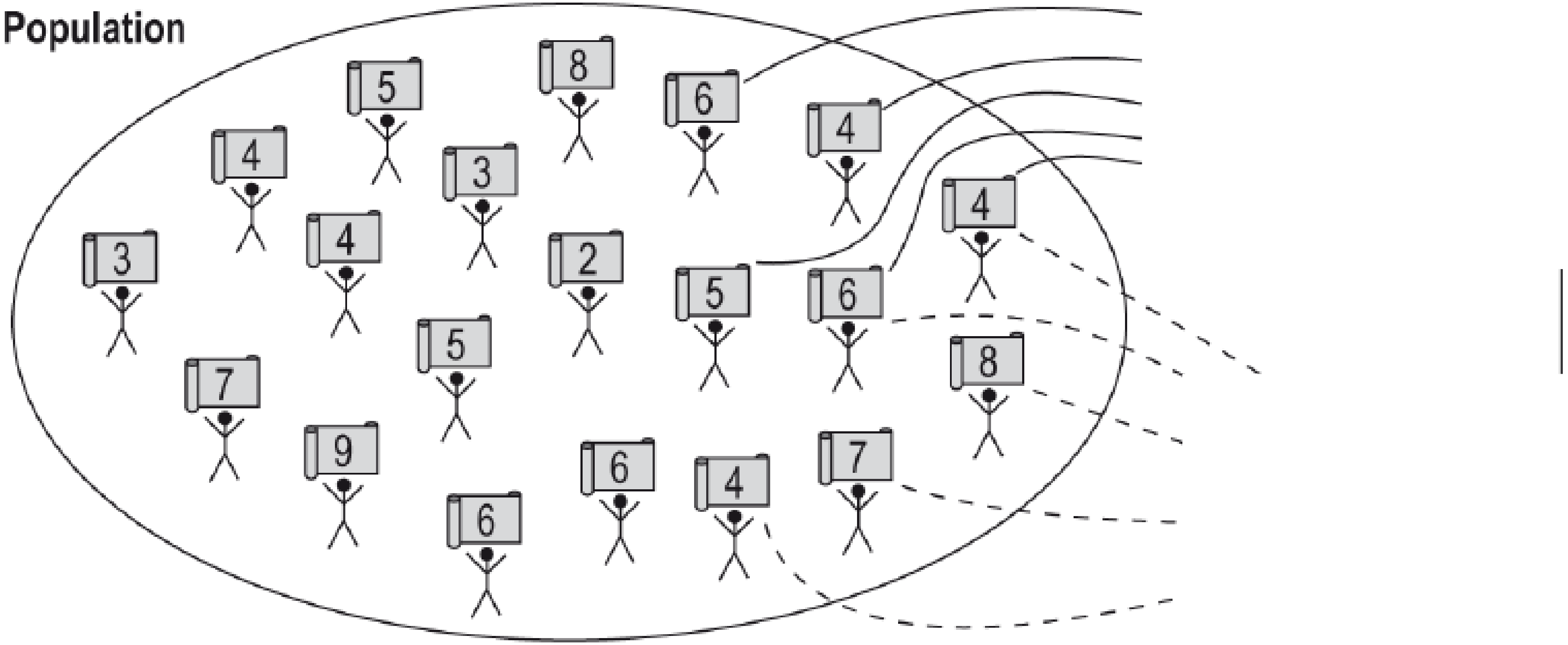
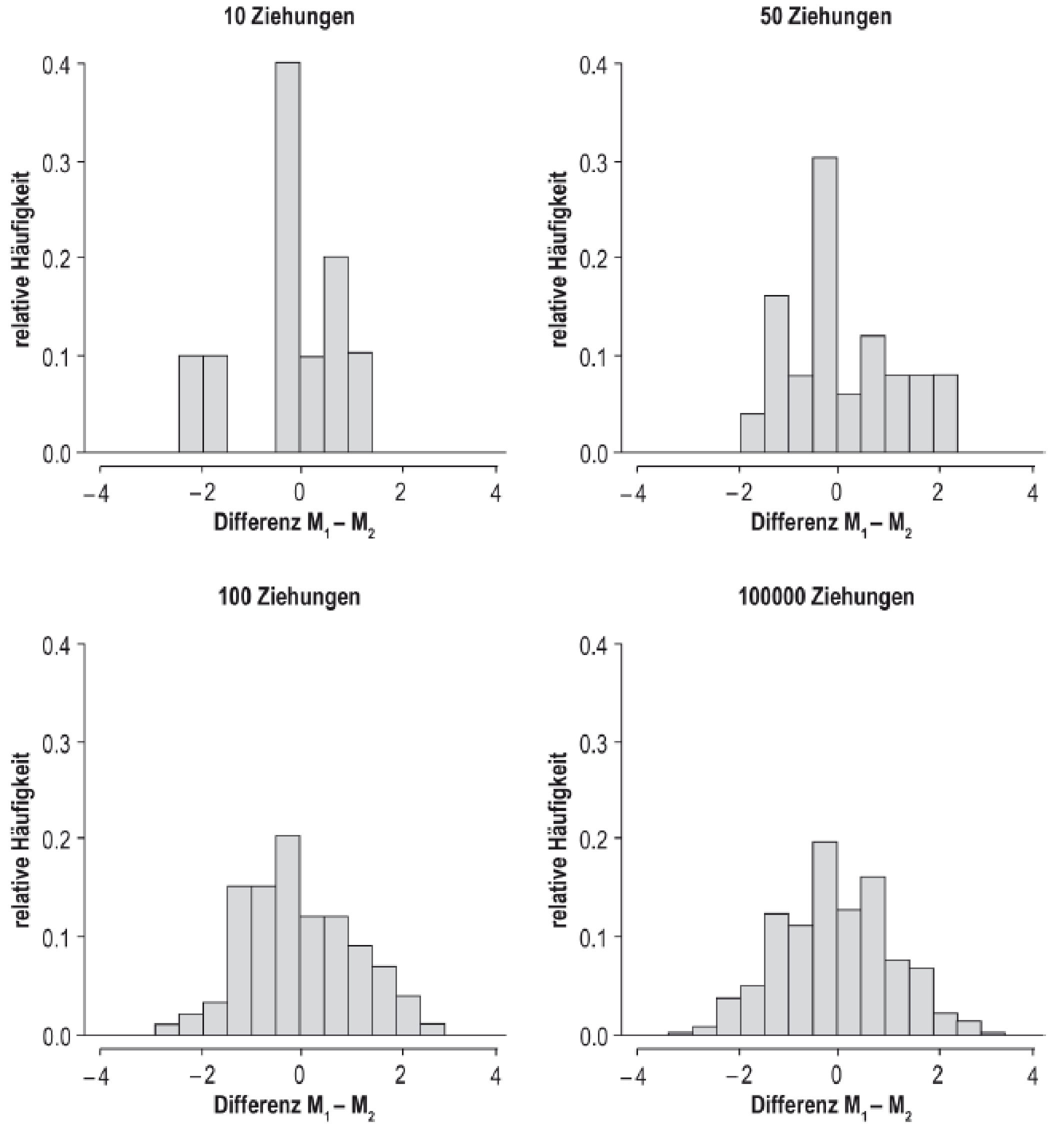


Abb. 2.2: *Zwei Stichproben vom Umfang n = 5 werden aus der gleichen Population gezogen*

2.4. Logik des Nullhypothesen-Signifikanztests

- Wir wechseln mal in R und spielen das mal durch…
- Beobachtungen um 0 herum sind offensichtlich häufiger
- Verteilung nähert sich einer Normalverteilung an



2.4. Prüfgrösse

- Was ist die Prüfgrösse?
- In unserem Beispiel: Die Differenz der zwei Mittelwerte
- Die Prüfgrösse nimmt grössere Werte an, je eher die Daten gegen die Nullhypothese sprechen
- Nullhypothese: Es gibt keinen Unterschied zwischen den Mittelwerten beider Stichproben
- In unserem Beispiel: Gilt die Nullhypothese oder die Alternativhypothese?

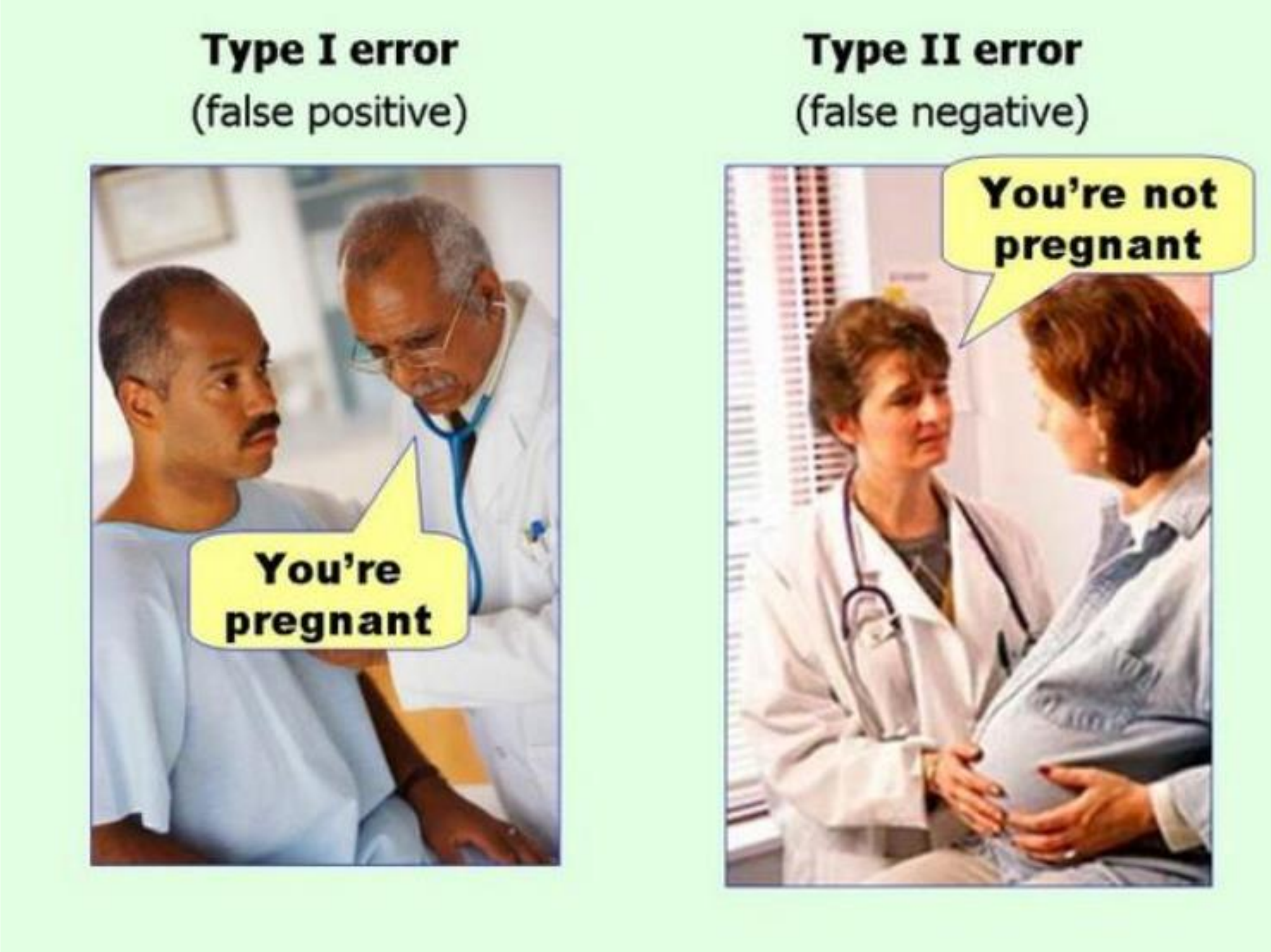
2.4. Logik des Nullhypothesen-Signifikanztests

- Grundsätzlich: Wir sprechen von einer **positiven Entscheidung, wenn die Nullhypothese abgelehnt wird**. Wir sprechen von einer negativen Entscheidung, wenn die Nullhypothese angenommen wird

	Null Hypothesis is TRUE	Null Hypothesis is FALSE
Reject null hypothesis	Type I Error (False positive)	Correct Outcome! (True positive)
Fail to reject null hypothesis	Correct Outcome! (True negative)	Type II Error (False negative)

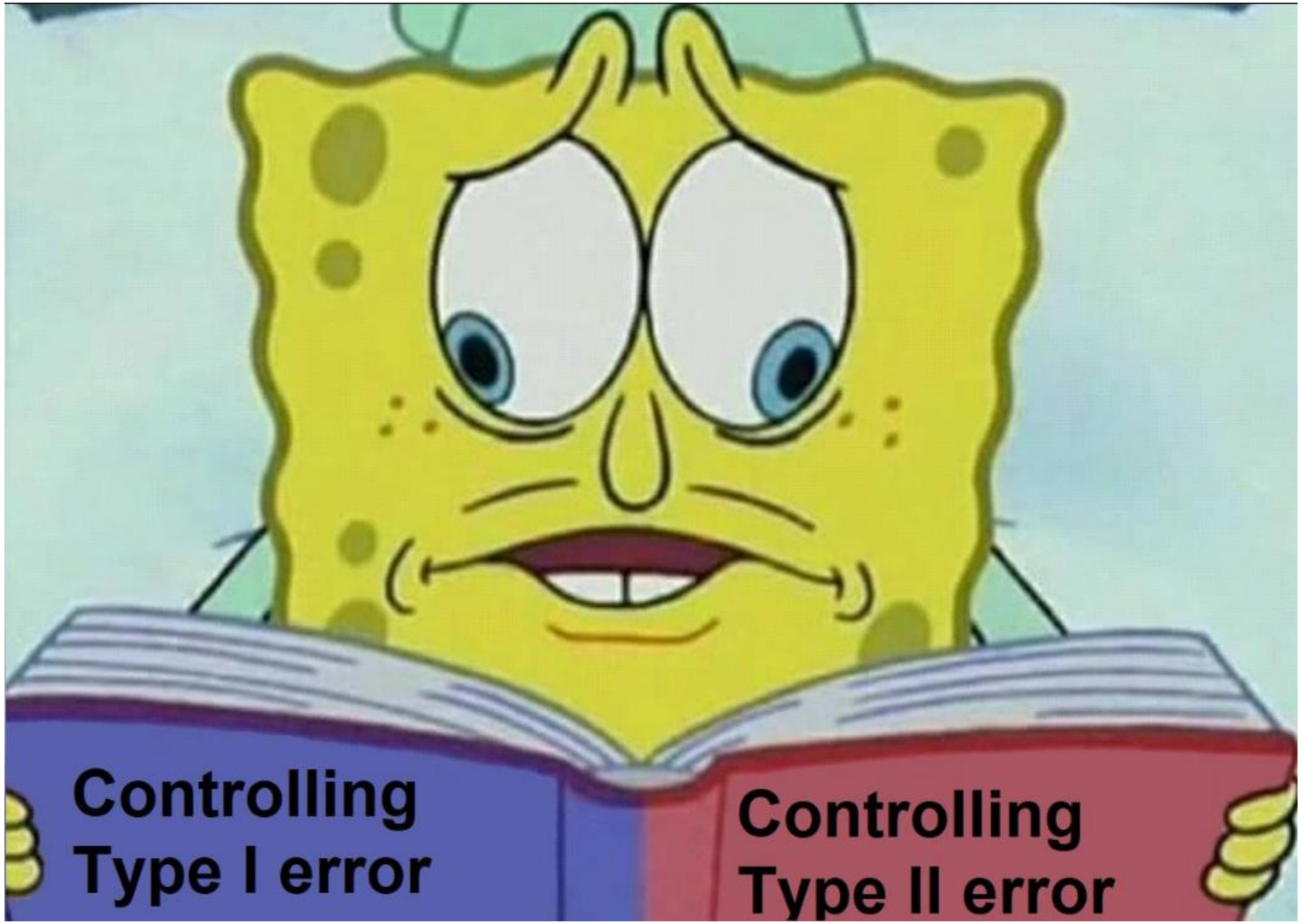
2.4. Logik des Nullhypothesen-Signifikanztests

Welcher Fehler wiegt schwerer? Typ I oder Typ II?



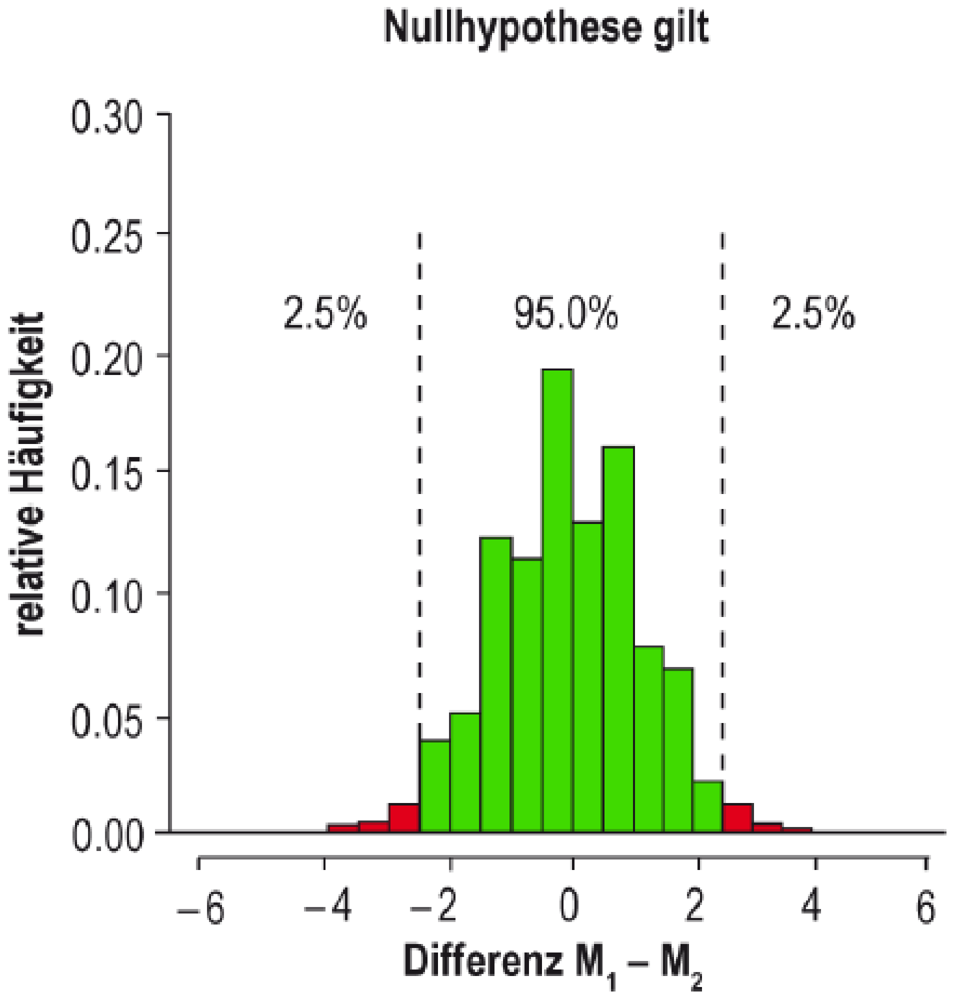
2.4. Logik des Nullhypothesen-Signifikanztests

- Type – I Error (alpha-Fehler): Wahrscheinlichkeit, dass die Nullhypothese fälschlicherweise abgelehnt wird
- Type – II Error (beta-Fehler): Wahrscheinlichkeit, dass die Nullhypothese fälschlicherweise angenommen wird



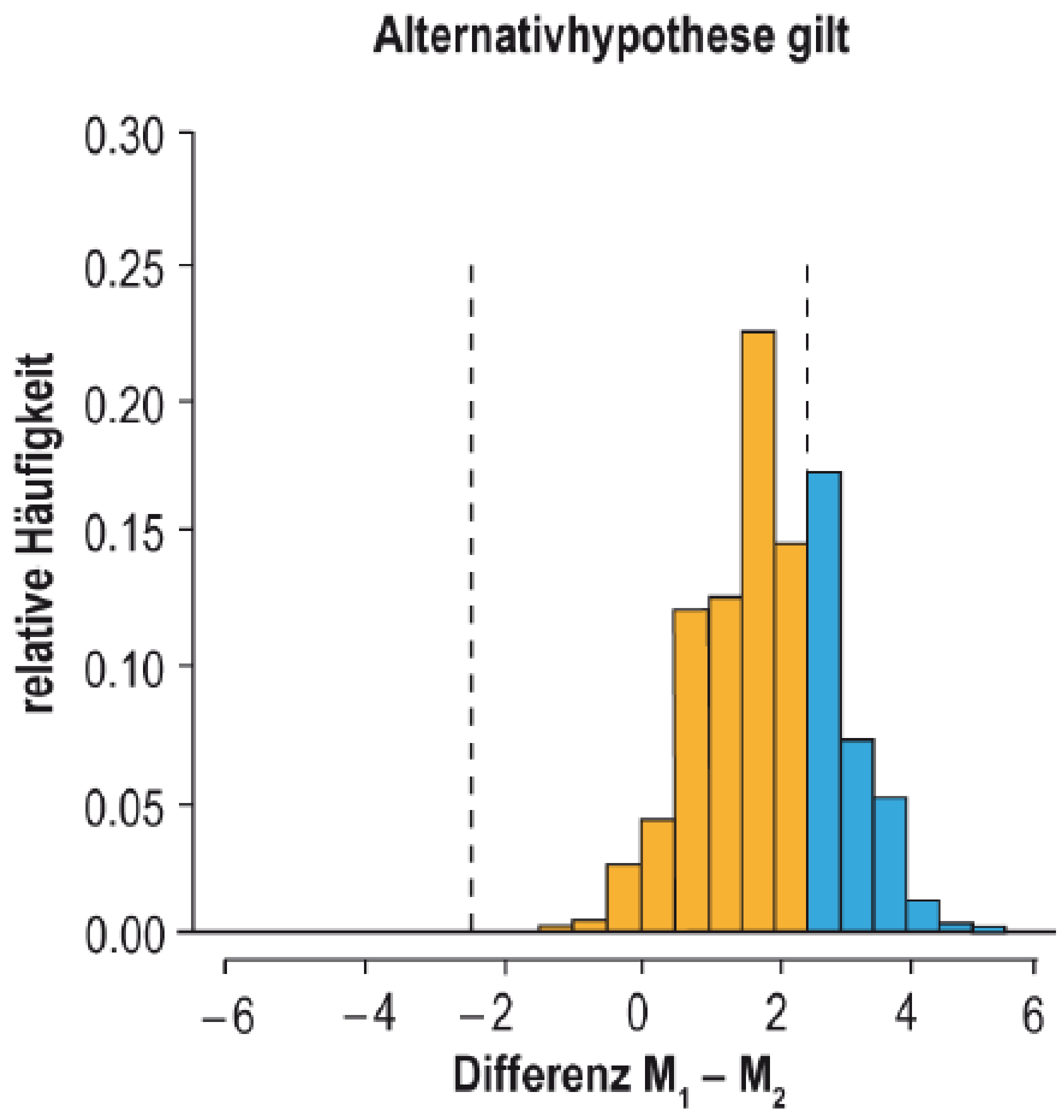
2.4. Logik des Nullhypothesen-Signifikanztests

- Fall 1: $M_1 - M_2 = 1.5$
    - Nullhypothese angenommen
    - Alternativhypothese abgelehnt
    - True negative
- Fall 2: $M_1 - M_2 = -3.5$
    - Nullhypothese abgelehnt
    - Alternativhypothese angenommen
    - Type I – Fehler (false positive)
- Fall 3: $M_1 - M_2 = 4$
    - Nullhypothese abgelehnt
    - Alternativhypothese angenommen
    - Type I – Fehler (false positive)



2.4. Logik des Nullhypothesen-Signifikanztests

- Fall 1: $M_1 - M_2 = 1.5$
  - Nullhypothese angenommen
  - Alternativhypothese abgelehnt
  - Type II – Fehler (false negative)
- Fall 2: $M_1 - M_2 = -3.5$
  - Nullhypothese abgelehnt
  - Alternativhypothese angenommen
  - True positive
- Fall 3: $M_1 - M_2 = 4$
  - Nullhypothese abgelehnt
  - Alternativhypothese angenommen
  - Type I – Fehler (false positive)



Testen von Mittelwertsunterschieden

- t-Test
  - Unabhänige Gruppen
  - Abhängige Gruppen
- ANOVA
  - Einfaktoriell
  - Mehrfaktoriell

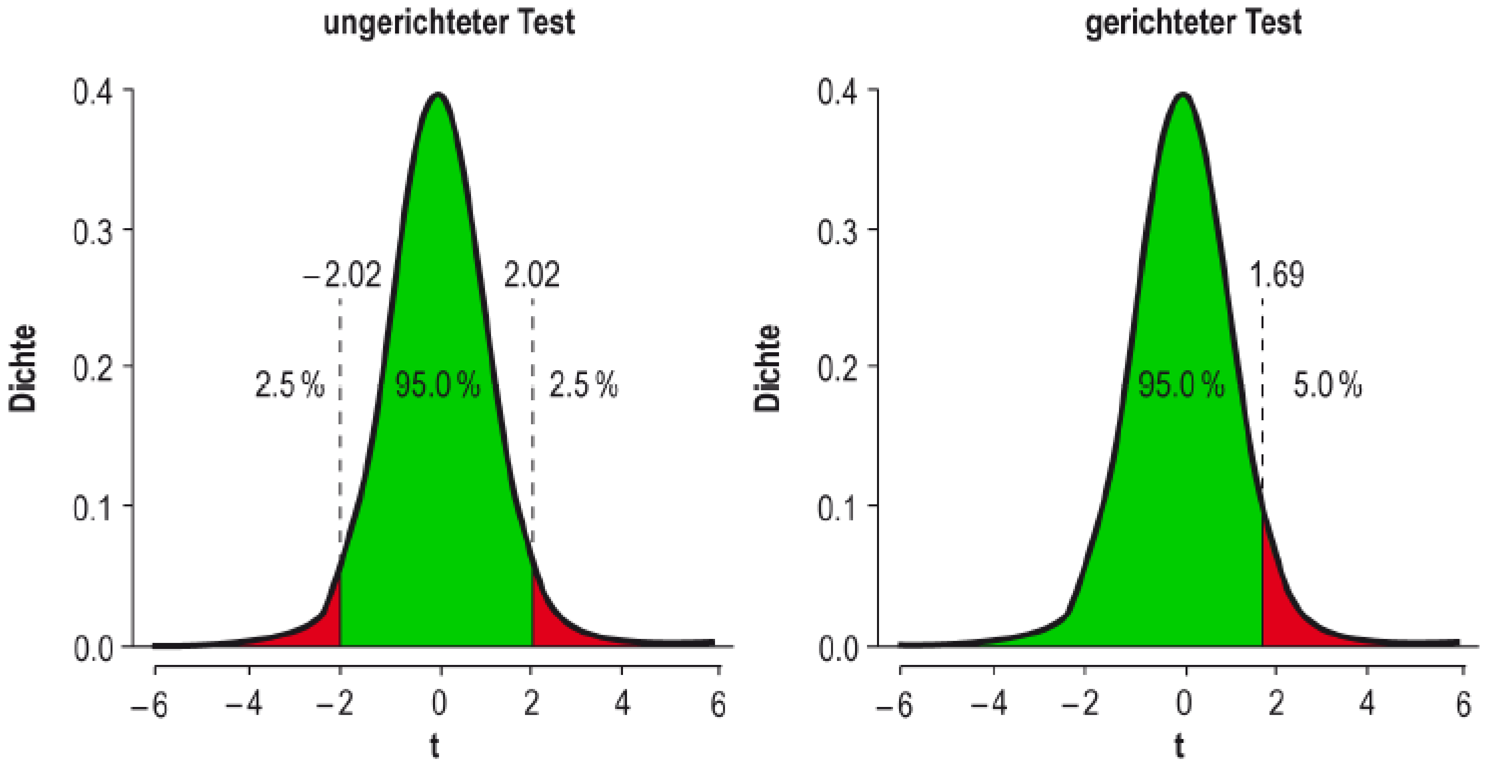
T-Test für zwei unabhängige Stichproben (ungerichteter Test)

- $H_1\colon\ \mu_A \neq\ \mu_B, H_0\colon\ \mu_A = \mu_B$
- Erster Blick:
  - $M_A = 6.1, S_A^2 = 0.89$
  - $M_B = 5.25, S_B^2 = 1.39$
- Wie machen wir nun weiter? Wir definieren eine Zufallsvariable $M_A - M_B$
- $t = \frac{M_A - M_B}{\sqrt{\frac{\hat{S}_A^2 + \hat{S}_B^2}{n}}}$
- Prüfgrösse
  - Bruch nimmt extremere Werte an, je eher die Daten gegen die Nullhypothese sprechen
  - t-verteilt mit $n_A + n_B - 2$ Freiheitsgraden, bei $n_A \approx n_B$

Berechnung des t-Wertes

$$t = \frac{M_A - M_B}{\sqrt{\frac{\hat{S}_A^2 + \hat{S}_B^2}{n}}}$$

Kritische T-Werte



- Dies sind diejenigen Werte links von dem und rechts von dem 2.5 % der Gesamtfläche unter der Kurve liegen, insgesamt also 5 % der Gesamtfläche.

Woher kriegen wir die kritischen t-Werte?

- Die „klassische“ Methode bieten Tabellen verschiedener Verteilungen, wie sie in vielen Statistiklehrbüchern zu finden sind (z. B. Bortz & Schuster, 2010; siehe auch Tabelle 6 im Anhang von Studienbrief 5 des Moduls Statistik 1). Hier müssen wir i. d. R. zunächst die entsprechende Zeile für die adäquaten Freiheitsgrade wählen.
- Alternativ: mittels Software

Entscheidungsregeln

- Wenn der empirische t-Wert grösser oder gleich dem kritischen t-Wert ist, dann betrachten wir das Ergebnis unserer Studie (oder ein noch extremeres) als hinreichend unwahrscheinlich, wenn die Nullhypothese tatsächlich gilt. Dies gibt uns ausreichend Zweifel an genau dieser Annahme und wir entscheiden uns für die Alternativhypothese.
- Wenn aber der empirische t-Wert kleiner als der kritische t-Wert ist, dann liegt ein Ergebnis vor, welches bei Gültigkeit der Nullhypothese so wahrscheinlich ist, dass wir wenig Anlass zum Zweifel an der Gültigkeit der Nullhypothese haben und wir uns folglich für ihre Beibehaltung entscheiden.

$$t = \frac{M_A - M_B}{\sqrt{\frac{\hat{S}_A^2 + \hat{S}_B^2}{n}}} = \frac{6.1 - 5.25}{\sqrt{\frac{\frac{20}{20-1} \cdot 0.89 + \frac{20}{20-1} \cdot 1.3875}{20}}} = \frac{0.85}{\sqrt{\frac{2.397368}{20}}} = 2.455085$$

Bestimmung des empirischen t-Werts

- Stichprobe 1: $M_A = 6.1, S_A^2 = 0.89$
- Stichprobe 2: $M_B = 5.25, S_b^2 = 1.3875$

Bestimmung des empirischen t-Werts

- Dieser empirische t-Wert ist also im Betrag größer als der kritische t-Wert von 2.02. Gemäß der Entscheidungsregel halten wir das Ergebnis der Studie also für so unwahrscheinlich, würde die Nullhypothese gelten, dass wir an genau dieser Annahme zweifeln und uns stattdessen für die Alternativhypothese entscheiden. In diesem Fall redet man auch von einem „signifikanten Ergebnis“, wobei „signifikant“ hier nicht bedeutet, dass es sich um ein inhaltlich bedeutendes Ergebnis handelt

P-value

- Prüfgrösse allein ist nicht entscheidend
- Zur Prüfgrösse wird ein p-Wert zugeordnet. Dieser p-Wert gibt die Wahrscheinlichkeit des Ergebnisses (oder eines extremeren Ergebnisses) unter Annahme der Nullhypothese an.
- Er sagt NICHTS (!!) über die Wahrscheinlichkeit der Nullhypothese resp. der Alternativhypothese aus
- P-Value: Fläche unter der t-Verteilung von $-\infty$ bis zum jeweiligen t-Wert (und bei ungerichteten Hypothesen auch die jeweilige Gegenseite.

Voraussetzungen für den T-Test und den sog. Welch-Test

- Drei wichtige Voraussetzungen
  - Die beiden Stichproben müssen zufällig und unabhängig voneinander gezogen worden sein.
  - Das betrachtete Merkmal, also im Beispiel der Fähigkeitstest, den die Kinder durchgeführt haben, muss als in der Population normalverteilt angenommen werden. Während sich die Populationsmittelwerte unterscheiden können, wird darüber hinaus vorausgesetzt, dass die Varianz beider Populationen identisch ist, d. h. beide korrigierten Stichprobenvarianzen auch Schätzer für die gleiche Populationsvarianz sind. Diese Voraussetzung wird Varianzhomogenität genannt und kann bspw. mit dem Levene-Test (Levene, 1960) überprüft werden
  - Da in die Berechnung von t Mittelwerte und Varianzen eingehen, muss für das betrachtete Merkmal (mindestens) Intervallskalenniveau unterstellt werden.

Was passiert, wenn Annahmen verletzt werden?

- Verletzungen solcher Voraussetzungen führen i. d. R. zu einem „liberalen“ Verhalten, da die Verteilung nicht mehr genau derjenigen entspricht, die sie sein sollte. In der Folge steigt die Wahrscheinlichkeit einer Entscheidung für die Alternativhypothese, obwohl die Nullhypothese gilt
- Robust gegen Verletzungen der Normalverteilung (in Grenzen) und moderaten Verletzungen der Varianzhomogenitätsannahme
- Aber wenn die Varianzhomogenität definitiv und stark verletzt wird, empfiehlt sich der Welch-Test

T-Test: Effektstärke

- Bisher haben wir nur die Prüfgrösse und die zugeordnete Wahrscheinlichkeit (p-Value) kennengelernt
- Diese Werte sagen uns, ob ein Test signifikant wird oder nicht, aber sie sagen uns nichts darüber aus, wie stark ein Effekt ist
- Effektstärke: Differenz der Mittelwerte, normiert an der Standardabweichung

$$d = \frac{M_A - M_B}{\hat{\sigma}}.$$

< 0,5 – kleiner Effekt

0,5 – 0,8 – mittelgradiger Effekt

> 0,8 – großer Effekt

T-Test: Power

- Power= Wahrscheinlichkeit, dass ein Test einen existierenden Effekt auch tatsächlich findet.
- Angenommen, ein statistischer Test hat eine Power von $p = 0.80$. Wenn wir nun annehmen, dass der zu untersuchende Effekt tatsächlich existiert, liefert uns der Test in 80% der Fälle ein positives Ergebnis.
- Grösste Einflussnahme auf die Power ist über die Stichprobengrösse möglich

Der T-Test: Schritt für Schritt

1. Annahmen prüfen
2. Signifikanzniveau festlegen
3. Hypothesen aufstellen
4. Testwert berechnen
5. Ablehnungsbereich definieren
6. Testentscheidung fällen

Probieren wir das mal gemeinsam an der Übungsaufgabe aus!

Durchführung des t-Tests in SPSS

Wir probieren das jetzt mal selbst aus ☺ Dazu laden wir uns das berühmte cars.sav - Dataset herunter. Darin sind verschiedene Kennzahlen zu Fahrzeugen enthalten. Wir interessieren uns dafür, ob der Kraftstoffverbrauch über die Zeit zurückgegangen ist. Ganz konkret möchten wir wissen, ob der Kraftstoffverbrauch der Fahrzeuge aus dem Jahr 1975 geringer ist als der Kraftstoffverbrauch der Fahrzeuge aus dem Jahr 1970.

Durchführung des t-Tests in SPSS

	mpg	year
1	18	70
2	15	70
3	18	70
4	16	70
5	17	70
6	15	70
7	14	70
8	14	70
9	14	70
10	15	70
11	.	70
12	.	70
13	.	70
14	.	70
15	.	70
16	15	70
17	14	70
18	.	70
19	15	70
20	14	70
21	24	70

	mpg	year
166	16	75
167	14	75
168	17	75
169	16	75
170	15	75
171	18	75
172	21	75
173	20	75
174	13	75
175	29	75
176	23	75
177	20	75
178	23	75
179	24	75
180	25	75
181	24	75
182	18	75
183	29	75
184	19	75
185	23	75
186	23	75

Data View Variable View

Unterscheiden sich die Jahrgänge voneinander? Was meint ihr?

Durchführung des t-Tests in SPSS

Was wir hier sehen, ist die Syntax von SPSS. Das ganze kann man natürlich auch über die grafische Oberfläche lösen, aber so ist es einfach cooler ☺

```
DATASET ACTIVATE DataSet1.
T-TEST GROUPS=year(70 75) /* Hier wählen wir sowohl die Testart als auch die Gruppierung aus
  /VARIABLES=mpg /* Hier wählen wir unsere Zielvariable aus
  /CRITERIA=CI(.95). /* Hier wählen wir das Konfidenzintervall unserer Schätzung aus (aka der Gegenwert zum p-value)
```

Durchführung des t-Tests in SPSS

Erster Eindruck: Gruppen sind hinsichtlich ihrer N und SE vergleichbar

Group Statistics

	Model Year (modulo 100)	N	Mean	Std. Deviation	Std. Error Mean
Miles per Gallon	70	28	18.00	5.164	.976
	75	30	20.27	4.941	.902

Durchführung des t-Tests in SPSS

Ergebnisse:

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means							
						Significance				95% Confidence Interval of the Difference	
		F	Sig.	t	df	One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference	Lower	Upper
Miles per Gallon	Equal variances assumed	.206	.652	-1.708	56	.047	.093	-2.267	1.327	-4.925	.391
	Equal variances not assumed			-1.706	55.277	.047	.094	-2.267	1.329	-4.930	.396

Einseitig oder zweiseitig?

Levene-Test. Sind die Varianzen identisch?

Signifikant oder nicht signifikant?

Durchführung des t-Tests in SPSS

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Miles per Gallon	Cohen's d	5.050	-.449	-.969	.075
	Hedges' correction	5.118	-.443	-.956	.074
	Glass's delta	4.941	-.459	-.983	.073

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control group.

- Effektstärke
- Siginifikanz an sich sagt nicht über die Grösse des Effekts aus. Daher IMMER die Effektstärke berichten

$$d = \frac{M_A - M_B}{\hat{\sigma}}.$$

$$\delta = \frac{\mu_A - \mu_B}{\sigma}.$$

Durchführung des t-Tests in SPSS

- Jetzt machen wir das mal selbst..

Durchführung des t-Tests mit verbundenen Stichproben

Was heisst überhaupt «verbundene» Stichproben?

Was wären Beispiele für verbundene Stichproben?

Praxisbeispiel: t-Test mit abhängigen Stichproben

- Sie entwickeln ein Medikament. Dazu haben Sie eine Untersuchung mit einer Stichprobe gemacht. Diese Stichprobe hat eine Messung vor der Gabe des Medikaments und nach der Gabe des Medikaments gemacht. Es handelt sich also um eine Messwiederholung. Nun möchten Sie wissen, ob das Medikament wirkt (also die Testwerte nach Gabe des Medikaments signifikant reduziert sind)
- Was ist die Null-, was ist die Alternativhypothese?
- Öffnet den Datensatz newdrug.sav

Durchführung des t-Tests mit abhängigen Stichproben

Skript für die Durchführung des t-Tests mit verbundenen Stichproben (Messwiederholung)

```
DATASET ACTIVATE DataSet1.
T-TEST PAIRS=Before_exp_BP WITH After_exp_BP (PAIRED)
  /CRITERIA=CI(.95)
  /MISSING ANALYSIS
```

Durchführung des t-Tests mit abhängigen Stichproben

Paired Samples Statistics

Means und Standardabweichung

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Before_exp_BP	98.3020	50	5.16770	.73082
	After_exp_BP	88.5980	50	4.56135	.64507

Paired Samples Correlations

Korrelation… was heisst das?

		N	Correlation	Significance	
				One-Sided p	Two-Sided p
Pair 1	Before_exp_BP & After_exp_BP	50	.193	.090	.180

Durchführung des t-Tests mit abhängigen Stichproben

Ergebnisse

Paired Samples Test

		Paired Differences					t	df	Significance	
					95% Confidence Interval of the Difference					
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			One-Sided p	Two-Sided p
Pair 1	Before_exp_BP - After_exp_BP	9.70400	6.19858	.87661	7.94238	11.46562	11.070	49	<.001	<.001

Durchführung des t-Tests mit abhängigen Stichproben

Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
Pair 1	Before_exp_BP - After_exp_BP	Cohen's d	6.19858	1.566	1.147	1.977
		Hedges' correction	6.24653	1.554	1.138	1.962

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Zusammenfassung

Wir haben gelernt…

- Was ein t-Test ist
- Worin sich ein t-Test für unabhängige Stichproben von einem t-Test für abhängige Stichproben unterscheidet
- Wie man einen t-Test für unabhängige Stichproben durchführt
- Wie man einen t-Test für verbundene Stichproben durchführt

ANOVA

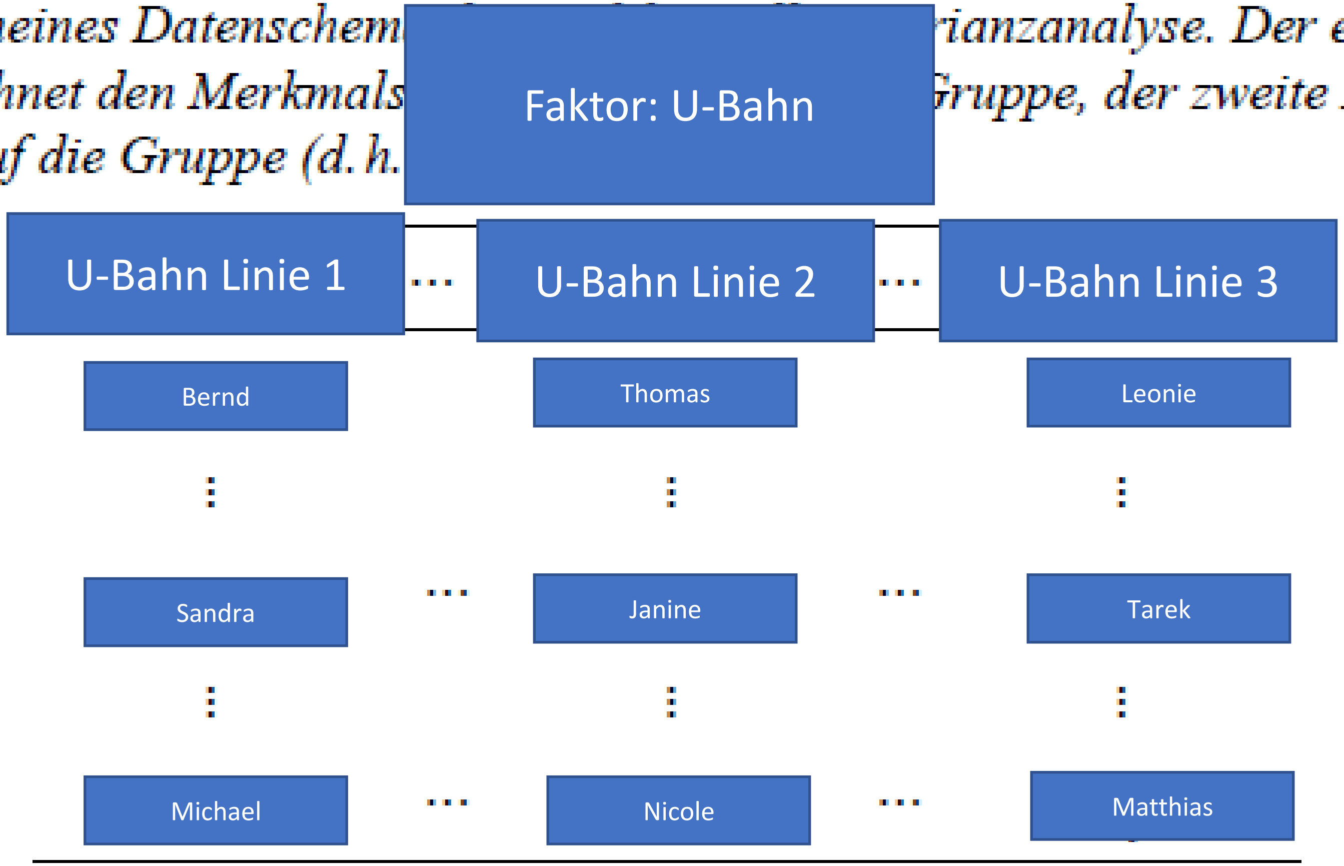
- Einfaktorielle ANOVA
- Einfaktorielle ANOVA mit Messwiederholung
- Mehrfaktorielle ANOVA
- Mehrfaktorielle ANOVA mit Messwiederholung

ANOVA - Varianzanalyse

- ANOVA = **An**alysis **o**f **va**riance
- Prüfung des Effekts einer unabhängigen Variable (oder mehrerer…) auf eine abhängige Variable
- Unabhängige Variable: Kategoriale Variable
- Unabhängige Variable wird als Faktor bezeichnet.
- Die einzelnen Kategorien des Faktors werden als Faktorstufen bezeichnet

ANOVA - Datenschema

Allgemeines Datenschem… rianzanalyse. Der erste Index (i) bezeichnet den Merkmals… Gruppe, der zweite Index (j) bezieht sich auf die Gruppe (d. h.…



ANOVA - Annahmen

- Varianzhomogenität $y_{ij} \sim N(\mu_j, \sigma^2)$
- Fehlerterm normalverteilt mit Erwartungswert 0 $e_{ij} \sim N(0, \sigma^2)$.
- Modellgleichung:

$$y_{ij} = \mu + \alpha_j + e_{ij}$$

ANOVA - Hypothesen

- Wir sind wieder in der Inferenzstatistik, daher gibt es auch hier eine Nullhypothese und eine Alternativhypothese
- Nullhypothese: $\alpha_j = 0$ $H_0\colon \mu_1 = \mu_2 = \ldots = \mu_J$,
- Alternativhypothese: $H_1\colon \mu_k \neq \mu_m$ für mindestens ein Paar k, m $(k, m \in \{1, \ldots, J\})$.

ANOVA – Prüfgrösse

- Prüfgrösse ist der F-Bruch
- F-Bruch = eine F-verteilte Zufallsvariable
- Ganz allgemein gesagt, bezeichnet der F-Bruch das Verhältnis der Varianz zwischen den Gruppen zur Varianz innerhalb der Gruppen
- Einfach ausgedrückt:

$$F = \frac{\text{Variabilität zwischen den Gruppen}}{\text{Variabilität innerhalb der Gruppen}} .$$

ANOVA – Prüfgrösse

$$F = \frac{\text{Variabilität zwischen den Gruppen}}{\text{Variabilität innerhalb der Gruppen}} = \frac{\text{Effekte+Messfehler}}{\text{Messfehler}} \; .$$

- Welchen Wert nimmt der F-Bruch für die Nullhypothese an?
- Wie verhält sich der F-Bruch, wenn die Daten zunehmend gegen die Nullhypothese sprechen?

ANOVA - Quadratsummenzerlegung

- Quadratsumme = Varianz der Daten
- Drei «Arten» der Quadratsumme: Quadratsumme innerhalb der Gruppen, Quadratsumme zwischen den Gruppen, Gesamtquadratsumme $SS_{\text{tot}} = SS_A + SS_w$.
- Grundprinzip der Varianzanalyse!! Die Varianz der Daten kann aufgeteilt werden in einen Teil, der auf den Effekt zurückgeht und einen Teil, der auf den Messfehler zurückgeht

ANOVA – Quadratsumme zwischen den Gruppen

- Summe der Abweichungen der Gruppenmittelwerte vom Gesamtmittelwert, multipliziert mit der Gruppengrösse

$$SS_A = \sum_{j=1}^{J} n_j (M_j - M)^2 \,.$$

ANOVA – Quadratsumme innerhalb der Gruppen

- Quadrierte Summe der Abweichungen der einzelnen Beobachtungen vom jeweiligen Gruppenmittelwert für alle Beobachtungen

$$SS_w = \sum_{j=1}^{J} \sum_{i=1}^{n_j} (y_{ij} - M_j)^2 = \sum_{j=1}^{J} n_j S_j^2 \,.$$

ANOVA – mittlere Quadratsumme

- Damit die Quadratsummen vergleichbar werden, müssen sie durch ihre Freiheitsgrade dividiert werden.
- Freiheitsgrade innerhalb der Gruppen: Alle Datenpunkte – Anzahl Gruppen
- Freiheitsgrade zwischen den Gruppen: Anzahl Gruppen – 1

- Die SS_A hat $J-1$ Freiheitsgrade, daher berechnet sich die Mittlere Quadratsumme zwischen den Gruppen (engl.: *Mean Square between; MS_A*) als:

$$MS_A = \frac{SS_A}{J-1} \quad .$$

ANOVA – mittlere Quadratsumme

- Die SS_A hat $J - 1$ Freiheitsgrade, daher berechnet sich die Mittlere Quadratsumme zwischen den Gruppen (engl.: *Mean Square between; MS_A*) als:

$$MS_A = \frac{SS_A}{J-1} \quad .$$

- SS_w hat $N - J$ Freiheitsgrade; N bezeichnet dabei wieder die Gesamtzahl der Beobachtungen. Daher berechnet sich die Mittlere Quadratsumme innerhalb der Gruppen (engl.: *Mean Square within; MS_w*) als:

$$MS_w = \frac{SS_\omega}{N-J} \quad .$$

ANOVA – der F-Bruch

$$F = \frac{\text{Variabilität zwischen den Gruppen}}{\text{Variabilität innerhalb der Gruppen}}.$$

$$F = \frac{MS_A}{MS_w}.$$

Das Beispiel aus dem Studienbrief

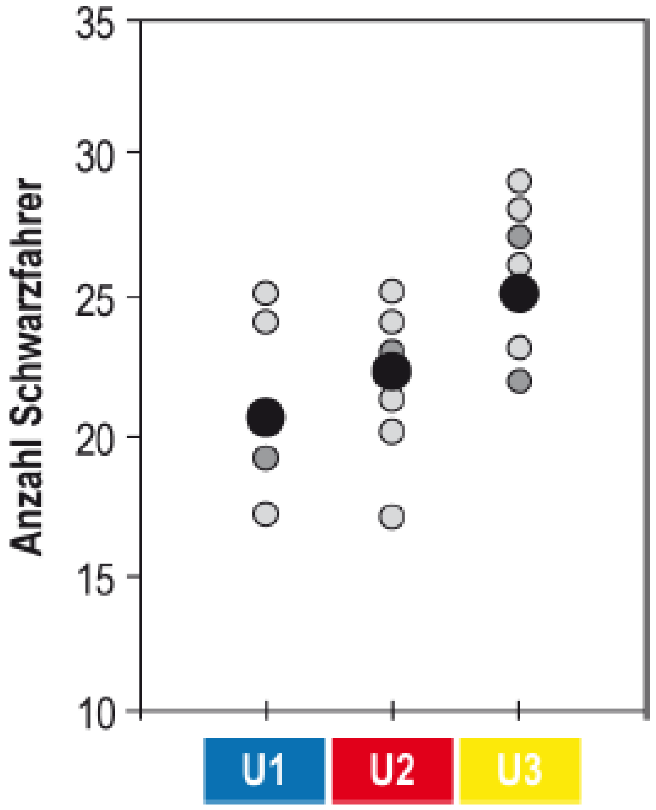


Abb. 3.1: Fiktive Datensituation zur Berechnung einer einfaktoriellen Varianzanalyse. Dargestellt ist die Anzahl an Personen, die bei jeweils 10 Kontrollen in den drei Hamburger U-Bahn-Linien U1, U2 und U3 als Schwarzfahrer erfasst wurden. Große schwarze Punkte entsprechen den Mittelwerten der jeweiligen Gruppen, während kleine graue Punkte die Ergebnisse einzelner Kontrollgänge symbolisieren; eine dunklere Schattierung zeigt hierbei das mehrfache Vorliegen desselben Wertes an

Das Beispiel aus dem Studienbrief

- Die Daten:

	U-Bahn-Linie		
	U1	U2	U3
	19	24	23
	20	23	22
	24	23	23
	20	22	22
	21	20	22
	25	25	26
	21	23	27
	17	17	27
	19	23	28
	19	21	29
M_j	20.5	22.1	24.9
S_j^2	5.25	4.69	6.89

Das Beispiel aus dem Studienbrief

- Quadratsummen zwischen den Gruppen

$$SS_A = \sum_{j=1}^{3} n_j(M_j - 22.5)^2$$

$$= 10 \cdot (20.5 - 22.5)^2 + 10 \cdot (22.1 - 22.5)^2 + 10 \cdot (24.9 - 22.5)^2$$

$$= 99.2 \ .$$

Das Beispiel aus dem Studienbrief

- Quadratsummen innerhalb der Gruppen

$$SS_w = \sum_{j=1}^{3} n_j S_j^2 = 10 \cdot 5.25 + 10 \cdot 4.69 + 10 \cdot 6.89 = 168.30\,.$$

Das Beispiel aus dem Studienbrief

- Mittlere Quadratsummen zwischen den Gruppen

$$MS_A = \frac{SS_A}{J-1} \quad . \qquad\qquad MS_A = \frac{99.2}{3-1} = 49.6$$

- Mittlere Quadratsummen innerhalb der Gruppen

$$MS_w = \frac{SS_\omega}{N-J} \quad . \qquad\qquad MS_w = \frac{168.30}{30-3} = 6.23\,.$$

F-Wert

- F-Wert ist F-verteilt bei J-1 und N-J Freiheitsgraden, wobei J die Anzahl der Gruppen und N die Anzahl der Beobachtungen ist

$$\boldsymbol{F} \sim F_{J-1,N-J} \,.$$

- Die Nullhypothese wird verworfen, wenn der empirische F-Wert grösser oder gleich dem kritischen F-Wert ist

$$F_{\mathbf{empirisch}} \geq F_{\mathbf{krit}} \Leftrightarrow p \leq \alpha \,.$$

Das Beispiel aus dem Studienbrief

- Der F-Bruch

$$F = \frac{MS_A}{MS_w}.$$

$$F = \frac{49.6}{6.23} = 7.96\,.$$

Effektstärke eta-Quadrat

- Häufiges Mass für Effektstärke: Varianzaufklärung des Faktors
- Varianzaufklärung: Wie viel der gesamten Varianz wird durch den Faktor erklärt?
- Also: Der Anteil der erklärten Varianz an der kompletten Varianz

$$\eta^2 = \frac{SS_A}{SS_{\text{tot}}}.$$

Effektstärke eta-Quadrat

- Der Schätzwert für die Effektstärke überschätzt die Effektstärke systematisch um einen gewissen Wert, er ist nicht **erwartungstreu**
- Eine mögliche Korrektur sieht wie folgt aus:

$$\eta^2 = \frac{SS_A - (J-1)MS_w}{SS_{\text{tot}} + MS_w}.$$

- Für die allermeisten Fälle reicht aber die einfache Formel aus der vorigen Folie

Einordnung der Effektstärke

- Kleine Effekte: $\eta^2 \geq .01$
- Mittlere Effekte: $\eta^2 \geq .06$
- Grosse Effekte: $\eta^2 \geq .14$

(Cohen, 1988)

Übungsaufgaben

3.1) Welche zwei formalen Voraussetzungen hat eine Varianzanalyse? Eine Verletzung welcher Voraussetzung ist schwerwiegender?

3.2) Für was steht die Abkürzung *SS* im Kontext der Varianzanalyse?

3.3) Bei welcher bzw. welchen der folgenden Berechnungen hat sich ein Fehler eingeschlichen?

a) $SS_{tot} = 3279.0$, $SS_A = 0.2$, $SS_w = 3278.8$

b) $SS_{tot} = 0.827$, $SS_A = 0.430$, $SS_w = 0.397$

c) $SS_{tot} = 228.1$, $SS_A = 67.6$, $SS_w = 161.5$

d) $SS_{tot} = 450.3$, $SS_A = -12.5$, $SS_w = 462.8$

3.4) Welche(n) der folgenden empirischen *F*-Brüche kann es nicht geben?

a) $F = 12.30$

b) $F = -4.20$

c) $F = 0.98$

d) $F = 182942.02$

Übungsaufgaben

3.6) Sie lesen in einem Forschungsartikel, dass eine Studie mit drei Gruppen durchgeführt wurde und in jeder Gruppe 30 Versuchspersonen getestet wurden. Weiterhin wird in dem Artikel vermerkt, dass einzelne Versuchspersonen aus der Analyse ausgeschlossen wurden, aber es wird nicht berichtet, wie viele Personen dies waren. Lässt sich diese Information aus der berichteten Teststatistik – $F(2, 82) = 14.20$, $p < .001$ – ablesen?

Übungsaufgaben

3.7) Vervollständigen Sie die fehlenden Zellen, die in den folgenden ANOVA-Tabellen markiert sind.

a)

Q.d.V.	*df*	*SS*	*MS*	*F*	*p*
Faktor	2	99.20	49.60		<.001
Fehler	57	168.30			

b)

Q.d.V.	*df*	*SS*	*MS*	*F*	*p*
Faktor		99.20	16.53	5.60	<.001
Fehler	57	168.30	3.00		

c)

Q.d.V.	*df*	*SS*	*MS*	*F*	*p*
Faktor	6	99.20	16.53	0.07	.998
Fehler	12		230.00		

Übungsaufgaben

ANOVA

Miles per Gallon

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7984.957	2	3992.479	97.969	<.001
Within Groups	16056.415	394	40.752		
Total	24041.372	396			

ANOVA Effect Sizes[a]

		Point Estimate	95% Confidence Interval	
			Lower	Upper
Miles per Gallon	Eta-squared	.332	.258	.396
	Epsilon-squared	.329	.255	.393
	Omega-squared Fixed-effect	.328	.254	.393
	Omega-squared Random-effect	.196	.146	.244

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

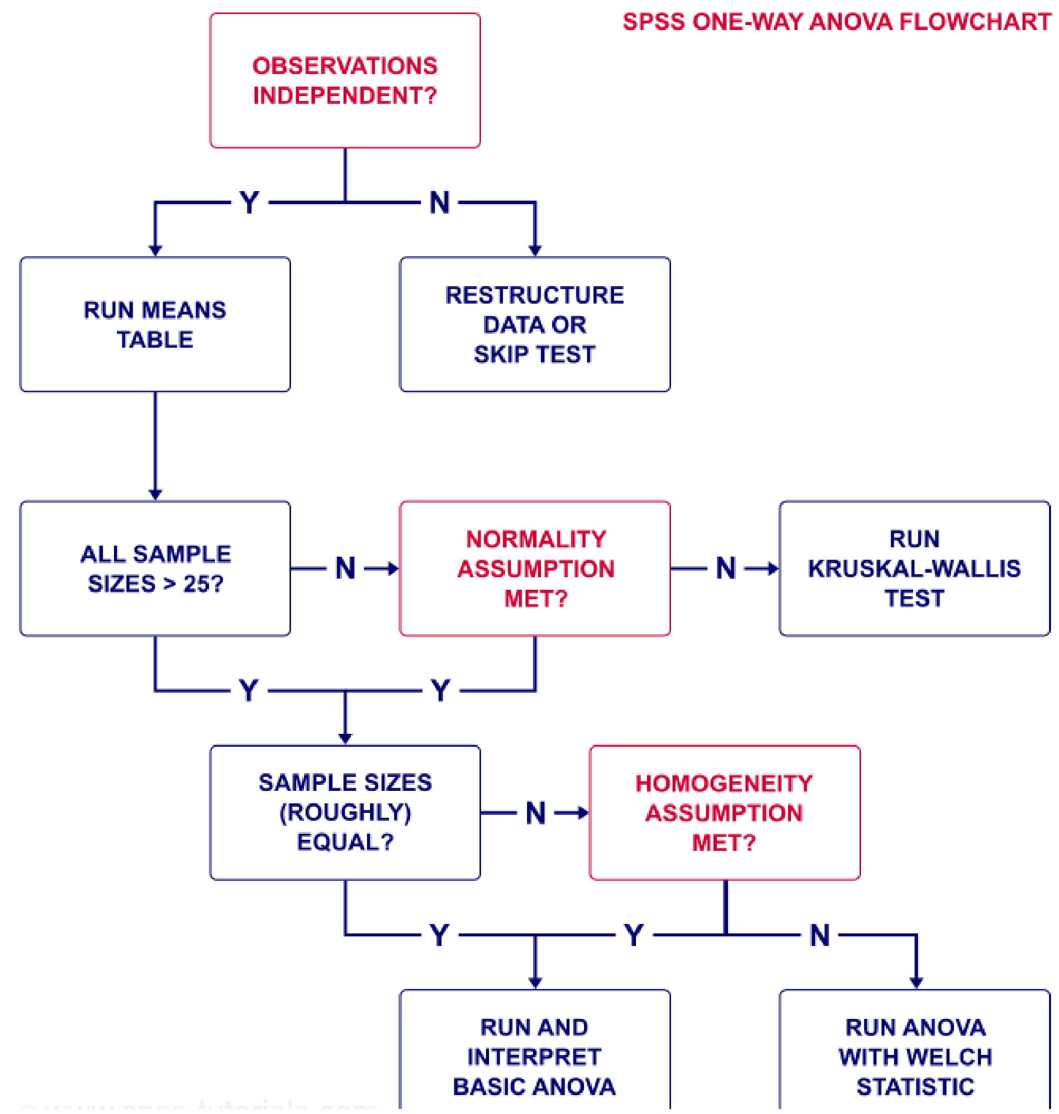
Einfaktorielle ANOVA

- Ein Faktor, mehrere Ausprägungen
- Testet auf Mittelwertsunterschiede
- Ähnlich dem t-Test
- Ziel: Überprüfen, ob die Mittelwerte in allen Gruppen identisch sind
- Grösster Unterschied zum t-Test: Vergleich mehrerer Gruppen in einem Test ist möglich

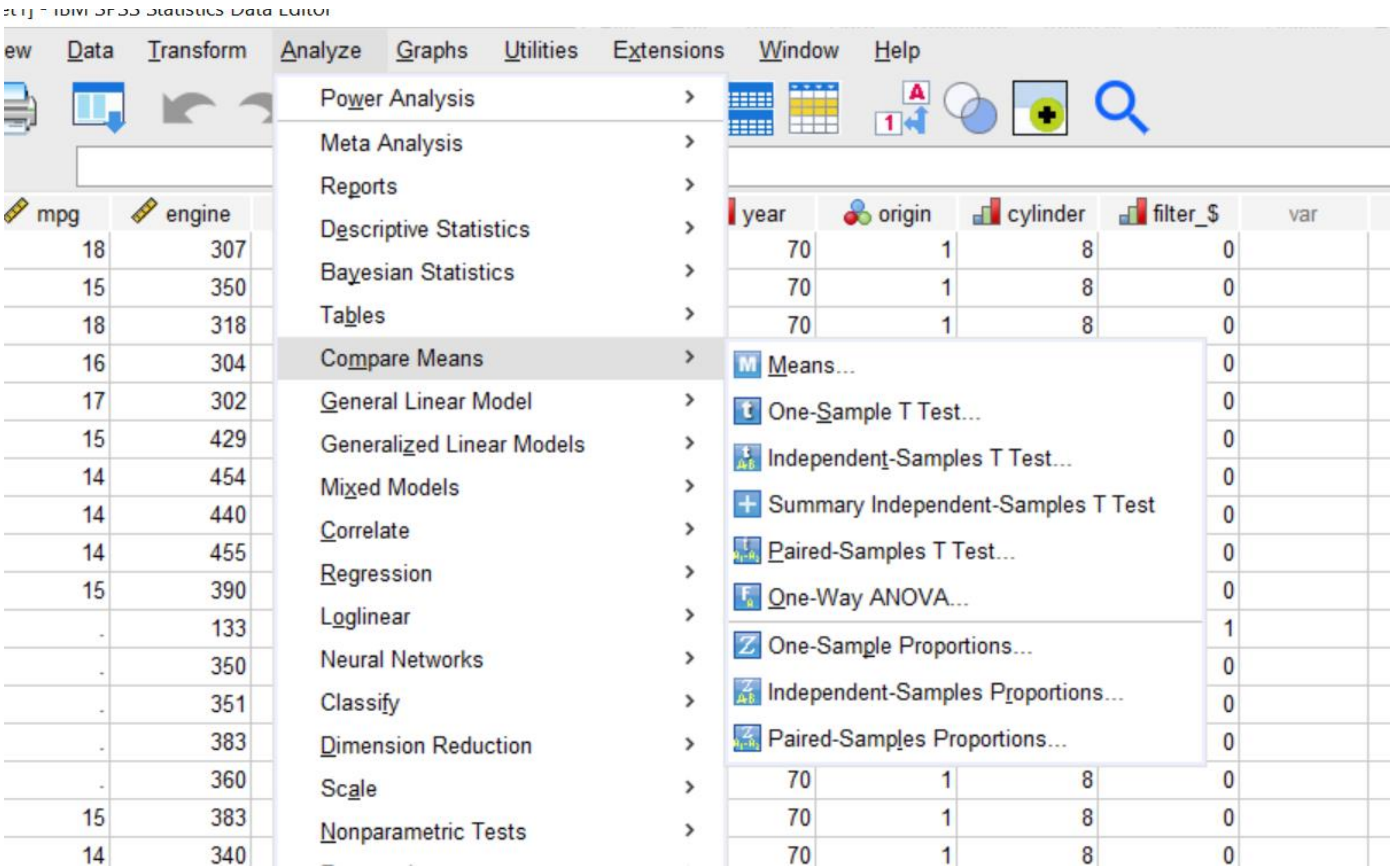
Einfaktorielle ANOVA

- Praxisbeispiel: Wir schauen uns wieder die verschiedenen Autos aus dem vorhergegangenen Datensatz an. Wir haben insgesamt drei verschiedene Herkünfte der Autos: Amerikanisch, europäisch, japanisch. Wir wollen untersuchen, ob sich die Durchschnittsverbräuche zwischen den Herkunftsländern unterscheiden
- Öffnet den Datensatz cars.sav

Einfaktorielle ANOVA



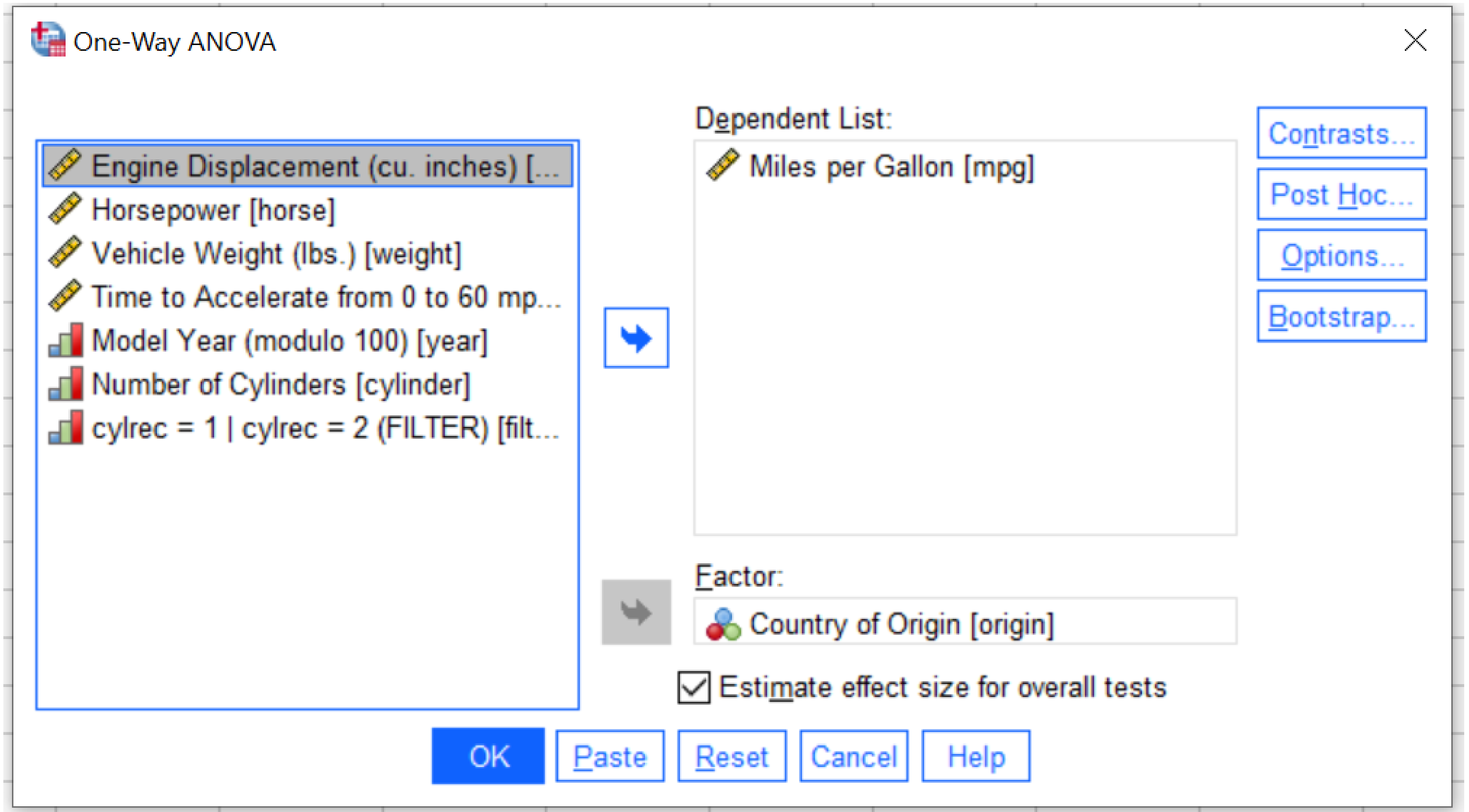
Einfaktorielle ANOVA



Einfaktorielle ANOVA

```
/* Auf gehts, ab gehts. ANOVA berechnen.

ONEWAY mpg BY origin
  /ES=OVERALL
  /MISSING ANALYSIS
  /CRITERIA=CILEVEL(0.95)
  /POSTHOC=BONFERRONI ALPHA(0.05).
```



Schritt 1 – Mittelwerte und Standardabweichungen betrachten

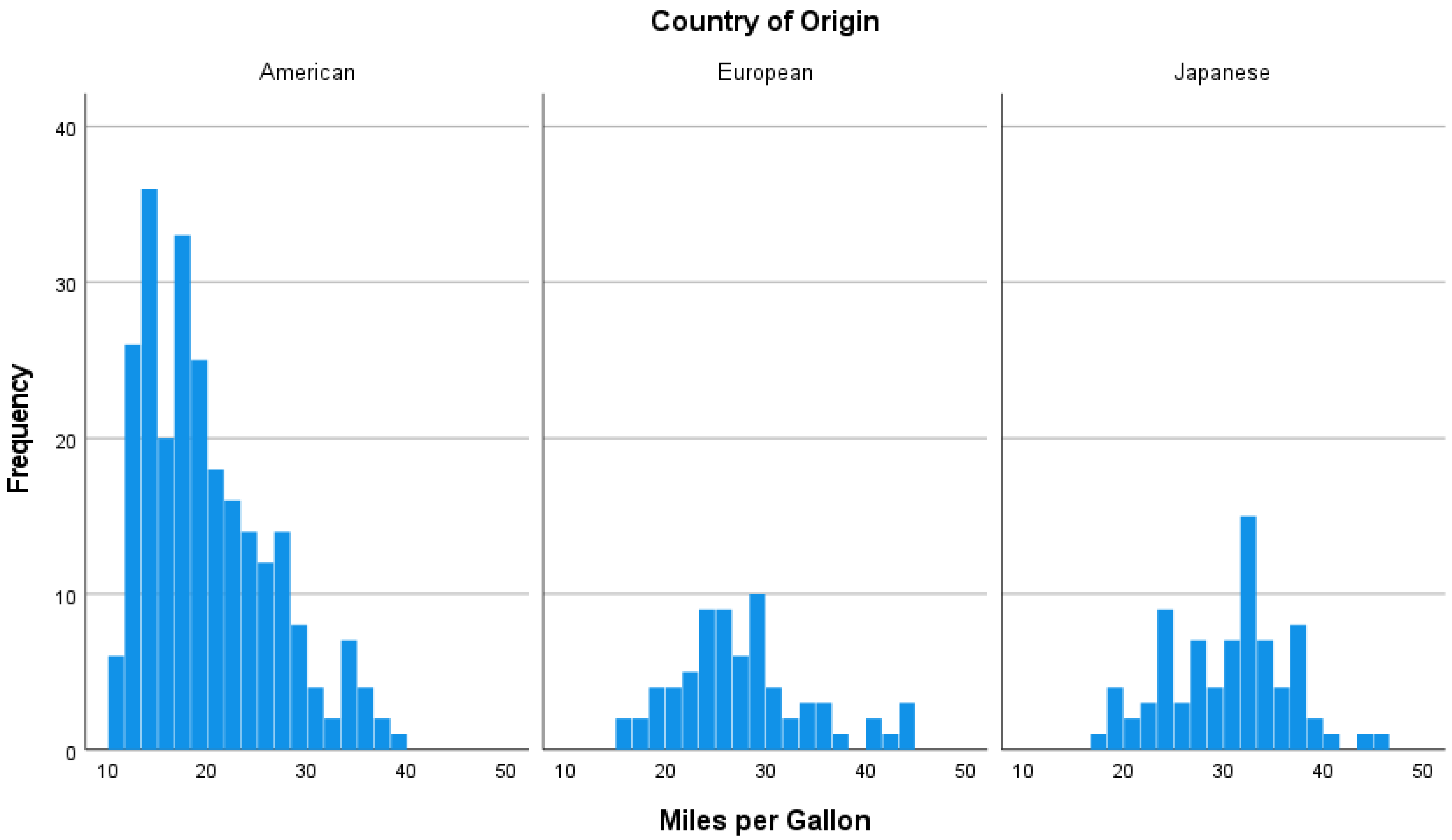
```
means mpg by origin
    /cells count mean stddev.
```

Report

Miles per Gallon

Country of Origin	N	Mean	Std. Deviation
American	248	20.13	6.377
European	70	27.89	6.724
Japanese	79	30.45	6.090
Total	397	23.55	7.792

Schritt 2 – ist Normalverteilung gegeben?



Schritt 3 – ANOVA berechnen - Ergebnisse

ANOVA

Miles per Gallon

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7984.957	2	3992.479	97.969	<.001
Within Groups	16056.415	394	40.752		
Total	24041.372	396			

Die Mittelwerte unterscheiden sich signifikant zwischen den Gruppen. $F(2, 394) = 97.97$, $p < .001$

Ergebnisse

ANOVA Effect Sizes[a]

		Point Estimate	95% Confidence Interval	
			Lower	Upper
Miles per Gallon	Eta-squared	.332	.258	.396
	Epsilon-squared	.329	.255	.393
	Omega-squared Fixed-effect	.328	.254	.393
	Omega-squared Random-effect	.196	.146	.244

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

Ergebnisse

Multiple Comparisons

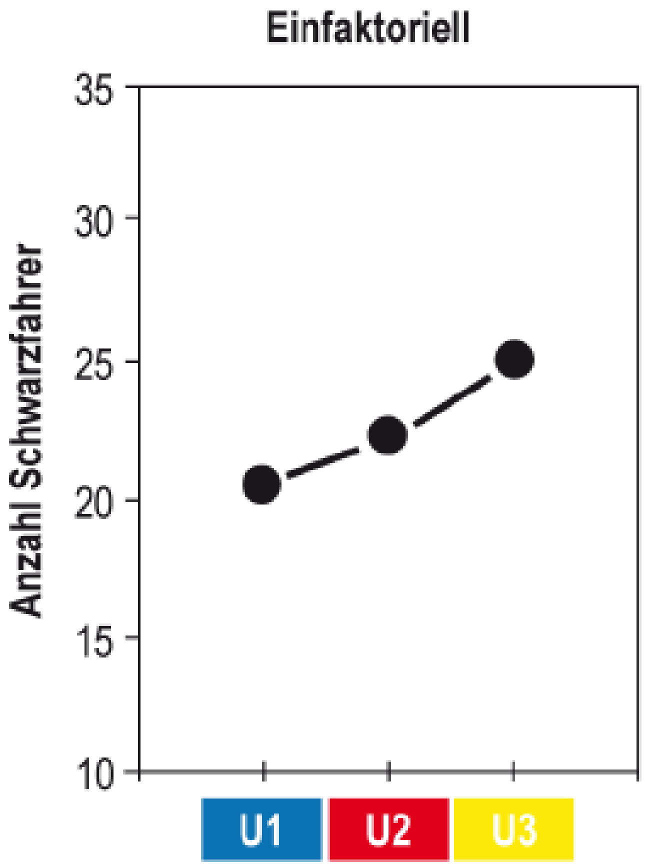
Dependent Variable: Miles per Gallon

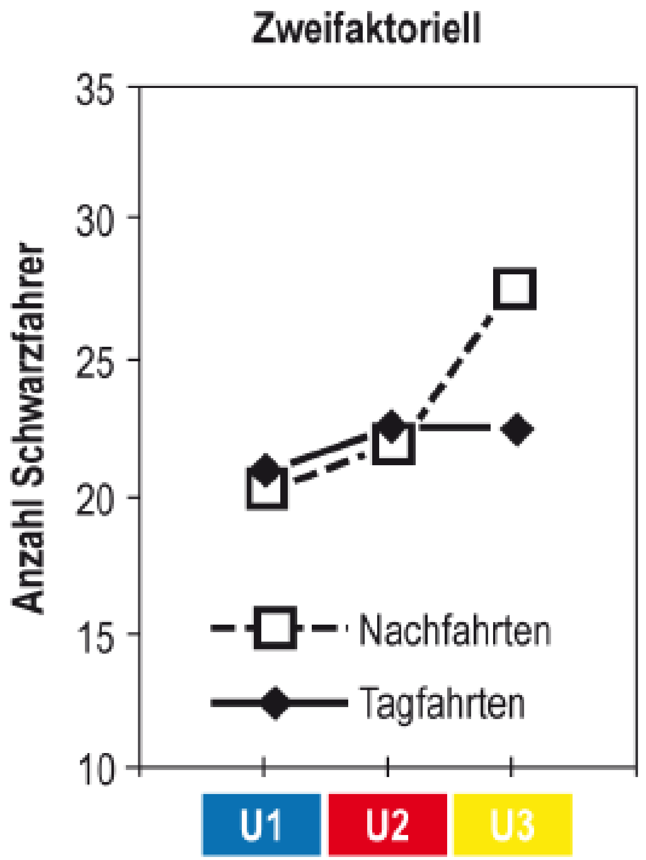
Bonferroni

(I) Country of Origin	(J) Country of Origin	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
American	European	-7.763*	.864	<.001	-9.84	-5.69
	Japanese	-10.322*	.825	<.001	-12.31	-8.34
European	American	7.763*	.864	<.001	5.69	9.84
	Japanese	-2.559*	1.048	.045	-5.08	-.04
Japanese	American	10.322*	.825	<.001	8.34	12.31
	European	2.559*	1.048	.045	.04	5.08

*. The mean difference is significant at the 0.05 level.

Einfaktoriell vs. Zweifaktoriell





Einfaktoriell vs. Zweifaktoriell

- Beispieldaten aus unserem Beispiel einer zweifaktoriellen ANOVA. Faktor 1: U-Bahn Linie, Faktor 2: Tageszeit

		U-Bahn-Linie		
		U1	U2	U3
		19	24	23
		20	23	22
	Tagfahrten	24	23	23
		20	22	22
		21	20	22
		25	25	26
		21	23	27
	Nachtfahrten	17	17	27
		19	23	28
		19	21	29
M_j	(Tagfahrten)	20.8	22.4	22.4
M_j	(Nachtfahrten)	20.2	21.8	27.4
S_j^2	(Tagfahrten)	2.96	1.84	0.24
S_j^2	(Nachtfahrten)	7.36	7.36	1.04

Globale Erwartungswerte

- Mittlerer Erwartungswert des Faktors über alle Stufen des anderen Faktors hinweg

		Faktor A: U-Bahn-Linie			
		U1	U2	U3	
Faktor B: Tageszeit	Tagfahrten	μ_{11}	μ_{21}	μ_{31}	$\mu_{\cdot 1}$
	Nachtfahrten	μ_{12}	μ_{22}	μ_{32}	$\mu_{\cdot 2}$
		$\mu_{1\cdot}$	$\mu_{2\cdot}$	$\mu_{3\cdot}$	μ

Für U-Bahn Linie: $$\mu_{1\cdot} = \frac{\mu_{11} + \mu_{12}}{2} \quad \text{und} \quad \mu_{2\cdot} = \frac{\mu_{21} + \mu_{22}}{2} \quad \text{und} \quad \mu_{3\cdot} = \frac{\mu_{31} + \mu_{32}}{2},$$

Für Tageszeit: $$\mu_{\cdot 1} = \frac{\mu_{11} + \mu_{21} + \mu_{31}}{3} \quad \text{und} \quad \mu_{\cdot 2} = \frac{\mu_{12} + \mu_{22} + \mu_{32}}{3}.$$

Haupteffekte

- Wir haben zwei Haupteffekte, also einen Haupteffekt pro Faktor

- Haupteffekt A (U-Bahn-Linie):

  $H_0^A\colon \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot}$

  $H_1^A\colon \mu_{m\cdot} \neq \mu_{n\cdot}$ für mindestens ein Paar $m, n \in \{1,2,3\}$

- Haupteffekt B (Tageszeit):

  $H_0^B\colon \mu_{\cdot 1} = \mu_{\cdot 2}$

  $H_1^B\colon \mu_{\cdot 1} \neq \mu_{\cdot 2}$

Haupteffekte

- Analog zur einfaktoriellen Varianzanalyse können wir die Effekte der beiden Faktoren als Differenz der Populationsmittelwerte zum Gesamterwartungswert auffassen

$$\alpha_j = \mu_{j.} - \mu \quad \forall j \in \{1,2,3\} \quad \text{und} \quad \beta_k = \mu_{.k} - \mu \quad \forall k \in \{1,2\}\,.$$

Für unser Beispiel:

$$\alpha_1 = -2.00 \quad \alpha_2 = -0.40 \quad \alpha_3 = 2.40 \quad \text{und} \quad \beta_1 = -0.63 \quad \beta_2 = 0.63\,.$$

Daher können wir auch eine vorläufige Modellgleichung aufstellen:

$$\mu_{jk} = \mu + \alpha_j + \beta_k\,.$$

Interaktionseffekt

- Wir haben zwei Haupteffekte alpha und beta, wir haben das Haupteffektmodell:

$$\mu_{jk} = \mu + \alpha_j + \beta_k \ .$$

- Theoretisch könnten wir jetzt alle Werte vorhersagen.. Oder?
- Nicht ganz, denn es gibt auch einen sog. Interaktionseffekt
- Was sagt dieser Effekt aus?

Interaktionseffekt

Tabelle 4.3: *Aus den Beispieldaten geschätzte Erwartungswerte μ_{jk} (obere Hälfte) im Vergleich zu den Schätzungen auf Basis des Haupteffektmodells (untere Hälfte)*

Geschätzte Erwartungswerte		**Faktor A: U-Bahn-Linie**			
		U1	U2	U3	
Faktor B: Tageszeit	Tagfahrten	20.80	22.40	22.40	21.87
	Nachtfahrten	20.20	21.80	27.40	23.13
		20.50	22.10	24.90	22.50

Vorhersagen des Haupteffektmodells		**Faktor A: U-Bahn-Linie**			
		U1	U2	U3	
Faktor B: Tageszeit	Tagfahrten	19.87	21.47	22.47	21.87
	Nachtfahrten	21.13	22.73	25.53	23.13
		20.50	22.10	24.90	22.50

Interaktionseffekt

- Offensichtlich gibt es systematische Abweichungen der empirischen Daten von den durch die beiden Haupteffekte vorhergesagten Daten
- Das liegt daran, dass die beiden Effekte unter Umständen voneinander abhängig sind. Ein Effekt A mag stärker auftreten, wenn Effekt B einen gewissen Wert hat und umgekehrt. Dann reicht es nicht mehr, sich nur die Haupteffekte anzuschauen.
- Der Interaktionseffekt berechnet sich wie folgt:

$$(\alpha\beta)_{jk} = \mu_{jk} - (\mu + \alpha_j + \beta_k) = \mu_{jk} - \mu - \alpha_j - \beta_k \ .$$

Interaktionseffekt

- Also kommt zu unseren beiden Haupteffekten noch ein dritter hinzu

$$(\alpha\beta)_{jk} = \mu_{jk} - \left(\mu + \alpha_j + \beta_k\right) = \mu_{jk} - \mu - \alpha_j - \beta_k \ .$$

- …der natürlich auch eine Null- und eine Alternativhypothese hat

- Interaktionseffekt AB (U-Bahn-Linie × Tageszeit):

  $H_0^{AB}: (\alpha\beta)_{jk} = 0 \qquad \forall j \in \{1,2,3\}, \ \forall k \in \{1,2\}$

  H_1^{AB}: es gibt mindestens ein $(\alpha\beta)_{jk} \neq 0 \ \ j \in \{1,2,3\}, \ k \in \{1,2\}$

Übersicht der Effekte

- Haupteffekt A und B:

$$\alpha_j = \mu_{j\cdot} - \mu \quad \forall j \in \{1,2,3\} \quad \text{und} \quad \beta_k = \mu_{\cdot k} - \mu \quad \forall k \in \{1,2\}\,.$$

- Haupteffekt A (U-Bahn-Linie):

  $H_0^A\colon \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot}$

  $H_1^A\colon \mu_{m\cdot} \neq \mu_{n\cdot}$ für mindestens ein Paar $m, n \in \{1,2,3\}$

- Haupteffekt B (Tageszeit):

  $H_0^B\colon \mu_{\cdot 1} = \mu_{\cdot 2}$

  $H_1^B\colon \mu_{\cdot 1} \neq \mu_{\cdot 2}$

Interaktionseffekt: $$(\alpha\beta)_{jk} = \mu_{jk} - (\mu + \alpha_j + \beta_k) = \mu_{jk} - \mu - \alpha_j - \beta_k\,.$$

- Interaktionseffekt AB (U-Bahn-Linie × Tageszeit):

  $H_0^{AB}\colon (\alpha\beta)_{jk} = 0 \quad \forall j \in \{1,2,3\},\ \forall k \in \{1,2\}$

  $H_1^{AB}\colon$ es gibt mindestens ein $(\alpha\beta)_{jk} \neq 0 \;\; j \in \{1,2,3\},\ k \in \{1,2\}$

Modellgleichung auf Parameterebene

- Interaktionseffekte sind die systematische Abweichung von der reinen Additivität der Haupteffekt. Sind alle Interaktionseffekte 0, liegt tatsächlich reine Additivität vor
- Daraus abgeleitet ergibt sich folgende Modellgleichung für die zweifaktorielle Varianzanalyse:

$$\mu_{jk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}\ .$$

Zellenmittelwert

Globaler Mittelwert

Haupteffekt A für die Faktorstufe j

Haupteffekt B für die Faktorstufe k

Interaktionseffekt AB für die Faktorstufen j und k

Quadratsummenzerlegung

- Analog zur einfaktoriellen ANOVA findet auch hier wieder die Quadratsummenzerlegung statt. Wir berechnen eine Quadratsumme für jeden Effekt (also A, B und den Interaktionseffekt)

Quadratsummenzerlegung

- Die Quadratsummenzerlegung erfolgt analog zur einfaktoriellen ANOVA, bloss dass hier eben für jeden einzelnen Effekt eine Quadratsumme berechnet wird

$$SS_{\text{tot}} = SS_A + SS_B + SS_{AB} + SS_W \ .$$

$$MS_A = \frac{SS_A}{J-1}, \qquad MS_B = \frac{SS_B}{K-1} \qquad MS_{AB} = \frac{SS_{AB}}{(J-1)(K-1)} \ ,$$

Quadratsummenzerlegung

$$SS_{\text{tot}} = \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{i=1}^{n}(y_{ijk} - M)^2 \quad \text{mit } df_{\text{tot}} = N - 1$$

$$SS_A = nK\sum_{j=1}^{J}(M_{j.} - M)^2 \quad \text{mit } df_A = J - 1$$

$$SS_B = nJ\sum_{k=1}^{K}(M_{.k} - M)^2 \quad \text{mit } df_B = K - 1$$

$$SS_{AB} = n\sum_{j=1}^{J}\sum_{k=1}^{K}(M_{jk} - M_{j.} - M_{.k} + M)^2 \quad \text{mit } df_{AB} = (J - 1)(K - 1)$$

$$SS_w = \sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{i=1}^{n}(y_{ijk} - M_{jk})^2 \quad \text{mit } df_w = JK(n - 1) = N - JK$$

Mittlere Quadratsummen

- Analog zur einfaktoriellen ANOVA berechnen wir die mittleren Quadratsummen durch die Teilung der Quadratsummen durch die Freiheitsgrade

- Effekte: $$MS_A = \frac{SS_A}{J-1}, \quad MS_B = \frac{SS_B}{K-1} \quad \text{und} \quad MS_{AB} = \frac{SS_{AB}}{(J-1)(K-1)},$$

- Fehler: $$MS_w = \frac{SS_w}{N-JK}.$$

F-Brüche

- Analog zur einfaktoriellen ANOVA können wir uns an dieser Stelle mehrere F-Brüche ausgeben lassen, für jeden Haupteffekt sowie für den Interaktionseffekt

$$F^{A} = \frac{MS_{A}}{MS_{w}} \qquad F^{B} = \frac{MS_{B}}{MS_{w}} \qquad F^{AB} = \frac{MS_{AB}}{MS_{w}}$$

Eta-Quadrat

- Hier wird üblicherweise die Quadratsumme des jeweiligen Effektes ins Verhältnis zur Summe aus dieser Quadratsumme und der Quadratsumme des Fehlerterms gesetzt
- Wir betrachten nicht mehr die gesamte Varianzaufklärung unseres Modells, sondern lediglich die Varianzaufklärung des jeweils betrachteten Effekts. Daher: **partielles Eta-Quadrat**

$$\eta_p^2 = \frac{SS_{\text{Effekt}}}{SS_{\text{Effekt}} + SS_w}$$

Übungsaufgaben

4.1) Welche der folgenden Fragestellungen kann über eine zweifaktorielle Varianzanalyse beantwortet werden? Wie würden Sie die jeweiligen Faktoren bezeichnen?

a) Unterscheidet sich die mittlere Geschwindigkeit junger Fahrer ($\leq$ 25 Jahre), mittelalter Fahrer (25-60 Jahre) und älterer Fahrer ($>$ 60 Jahre) in Spielstraßen, Tempo-30-Zonen und sonstigen Ortsbereichen?

b) Unterscheidet sich die mittlere Arbeitszufriedenheit zwischen den sechs unterschiedlichen Abteilungen eines Unternehmens?

c) Wie wirkt sich die Dauer von Schlafentzug (0 Tage vs. 1 Tag vs. 2 Tage) bei Frauen und Männern auf die Gedächtnisspanne aus?

4.2) Sowohl bei der einfaktoriellen als auch bei der zweifaktoriellen Varianzanalyse wird die Gesamtvarianz eines Datensatzes in unterschiedliche Anteile zerlegt. Welche Varianzquellen werden bei den beiden Verfahren jeweils unterschieden?

Übungsaufgaben

4.3) In den Beispielrechnungen zu einfaktoriellen und zweifaktoriellen Varianzanalysen kommen jeweils die beiden Varianzquellen „Faktor A: U-Bahn-Linie“ und „(Zufalls-)Fehler“ vor. Vergleichen Sie die entsprechenden Zeilen der beiden resultierenden ANOVA-Tabellen in Abschnitt 3.6.1 und Abschnitt 4.4: Welche Zellen sind gleich, welche unterscheiden sich?

Übungsaufgaben

4.4) Welche der folgenden Aussagen sind richtig (R), welche falsch (F)?

a) __ Eine signifikante Interaktion bedeutet, dass ein Faktor nur auf bestimmten Stufen des anderen Faktors wirkt, auf anderen Stufen hingegen keine Wirkung zeigt.

b) __ Wenn sich alle empirisch beobachteten Mittelwerte voneinander unterscheiden, dann ist das Haupteffektmodell unzureichend und es muss eine signifikante Interaktion vorliegen.

c) __ Wenn eine einfaktorielle Varianzanalyse keinen signifikanten Effekt eines Faktors A gezeigt hat, kann eine zweifaktorielle Varianzanalyse dennoch einen signifikanten Effekt eben dieses Faktors A zeigen.

d) __ Der zweite Faktor kann nicht mehr Stufen haben als der erste Faktor ($K \leq J$).

e) __ Wird ein gegebener Datensatz mit einer zweifaktoriellen Varianzanalyse ausgewertet, dann hat der Fehlerterm immer weniger Freiheitsgrade als wenn eine einfaktorielle Varianzanalyse mit einem der beiden Faktoren berechnet worden wäre.

Mehrfaktorielle ANOVA

- Manchmal vergleichen wir ein Ereignis, welches von mehreren Faktoren beeinflusst wird
- Beispiel: Wir beobachten Personen, die ihr Gewicht verringern wollen. Diese unterscheiden sich sowohl hinsichtlich der Diät, die sie machen (Faktor 1: Ernährung) als auch hinsichtlich der Menge an Sport, die sie treiben (Faktor 2: Sport)
- Wir wollen jetzt herausfinden, ob sich der Gewichtsverlust zwischen den verschiedenen Ernährungsformen unterscheidet, von der Menge des Sports abhängt oder ob es dort sogar eine Interaktion gibt

Mehrfaktorielle ANOVA

```
79  crosstabs diet by exercise
80    /statistics chisq.
81
```

Zunächst schauen wir an, ob alle Gruppen ähnlich viele Beobachtungen haben

Diet assigned to participant * Exercise level assigned to participant Crosstabulation

Count

		Exercise level assigned to participant			Total
		None	30 minutes per day	60 minutes per day	
Diet assigned to participant	None	20	20	20	60
	Atkins	20	20	20	60
	Vegetarian	20	20	20	60
Total		60	60	60	180

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	.000[a]	4	1.000
Likelihood Ratio	.000	4	1.000
Linear-by-Linear Association	.000	1	1.000
N of Valid Cases	180		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 20.00.

Mehrfaktorielle ANOVA

- Annahmen
  - Unabhängigkeit der Beobachtungen
  - Homoskedastizität
  - Normalverteilung

Mehrfaktorielle ANOVA

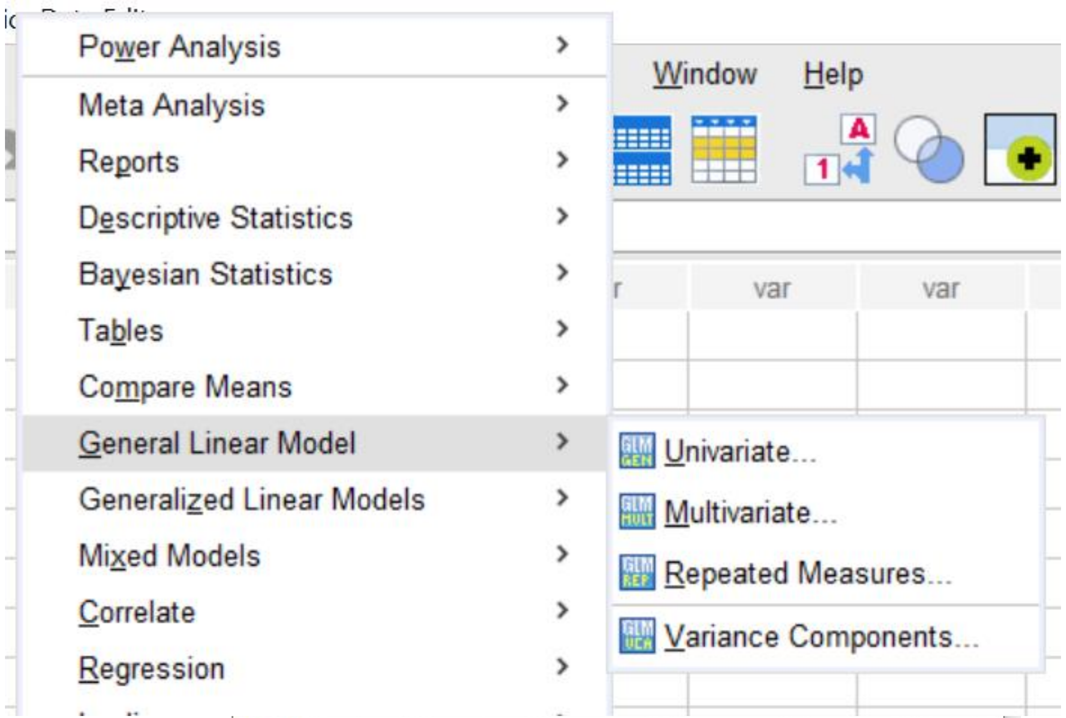
- Means
- Was sehen wir hier?
  - Sehen wir einen Effekt von Diät? Von Sport?

Report

Observed weigth loss in kilos over last 2 months

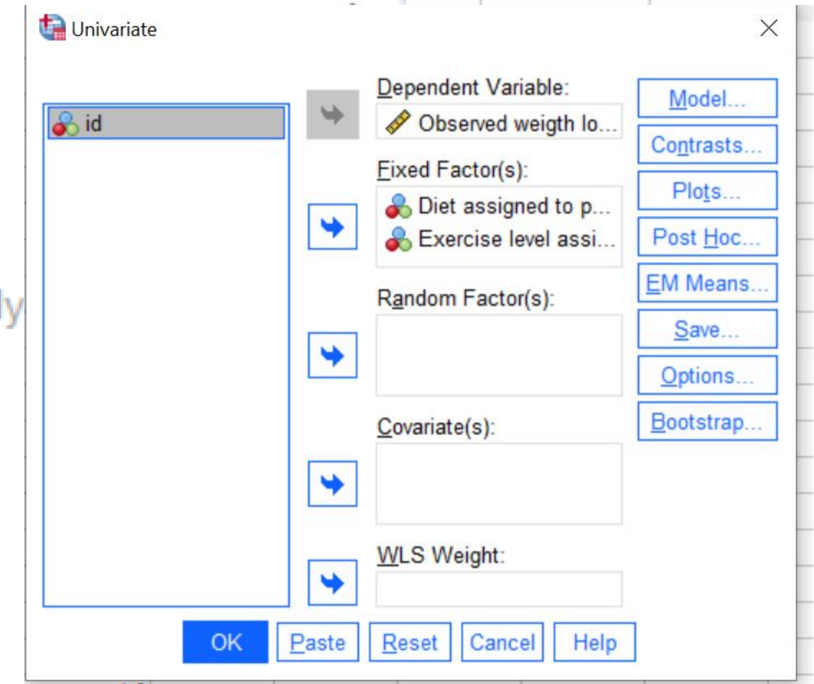
Diet assigned to participant	Exercise level assigned to participant	Mean	N	Std. Deviation
None	None	.100	20	2.7503
	30 minutes per day	1.332	20	3.0208
	60 minutes per day	7.092	20	2.4665
	Total	2.841	60	4.0961
Atkins	None	2.464	20	2.7789
	30 minutes per day	1.712	20	3.7043
	60 minutes per day	8.756	20	3.0127
	Total	4.311	60	4.4691
Vegetarian	None	4.328	20	2.7752
	30 minutes per day	4.716	20	2.4608
	60 minutes per day	9.888	20	3.2356
	Total	6.311	60	3.7861
Total	None	2.297	60	3.2322
	30 minutes per day	2.587	60	3.4114
	60 minutes per day	8.579	60	3.0975
	Total	4.488	180	4.3441

Mehrfaktorielle ANOVA




```
/* ANOVA

UNIANOVA wloss BY diet exercise /* Hier übergeben wir die Variablen
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /POSTHOC=diet exercise (BONFERRONI) /* Hier haben wir die Post-Hoc Analy
  /PLOT=PROFILE(exercise*diet) /* Hier der Plot
  /PRINT=ETASQ HOMOGENEITY /* Hier geben wir die Effektstärke aus
  /CRITERIA=ALPHA(.05)
  /DESIGN=diet exercise diet*exercise.
```



Mehrfaktorielle ANOVA

1. Levene-Test

(testet Varianzgleichheit)

Levene's Test of Equality of Error Variances[a,b]

		Levene Statistic	df1	df2	Sig.
Observed weigth loss in kilos over last 2 months	Based on Mean	.744	8	171	.653
	Based on Median	.718	8	171	.676
	Based on Median and with adjusted df	.718	8	157.543	.676
	Based on trimmed mean	.749	8	171	.648

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Dependent variable: Observed weigth loss in kilos over last 2 months

b. Design: Intercept + diet + exercise + diet * exercise

Mehrfaktorielle ANOVA - Ergebnisse

Hat Diät einen Effekt?

Hat Sport einen Effekt?

Gibt es eine Interaktion (und was heisst das überhaupt)?

Tests of Between-Subjects Effects

Dependent Variable: Observed weigth loss in kilos over last 2 months

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1905.037[a]	8	238.130	27.647	<.001	.564
Intercept	3624.868	1	3624.868	420.844	<.001	.711
diet	363.904	2	181.952	21.124	<.001	.198
exercise	1508.859	2	754.429	87.589	<.001	.506
diet * exercise	32.274	4	8.069	.937	.444	.021
Error	1472.879	171	8.613			
Total	7002.784	180				
Corrected Total	3377.916	179				

a. R Squared = .564 (Adjusted R Squared = .544)

Mehrfaktorielle ANOVA - Ergebnisse

Post-Hoc Vergleiche

Multiple Comparisons

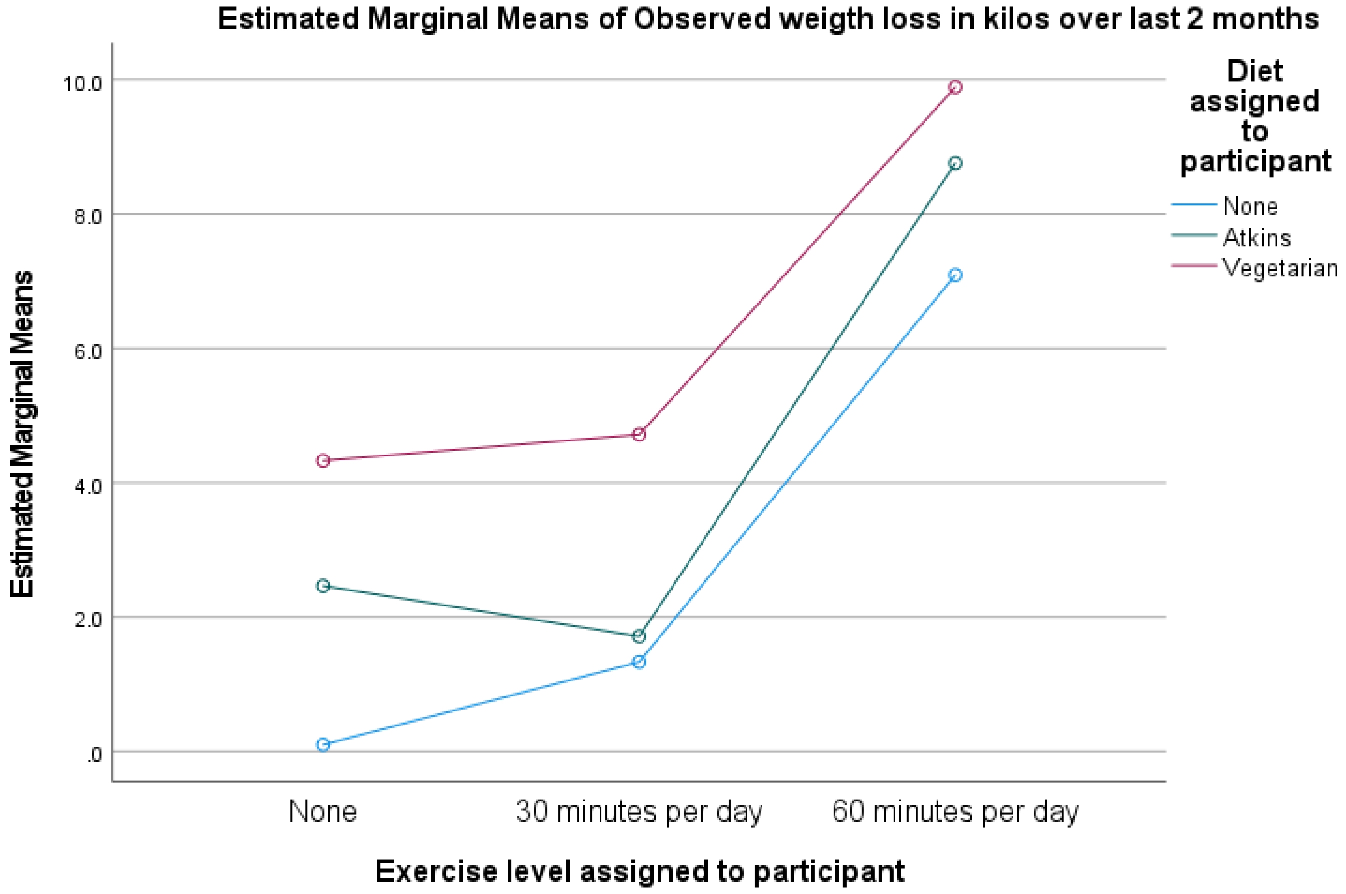
Dependent Variable: Observed weigth loss in kilos over last 2 months

Bonferroni

(I) Diet assigned to participant	(J) Diet assigned to participant	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
None	Atkins	-1.469*	.5358	.020	-2.765	-.174
	Vegetarian	-3.469*	.5358	<.001	-4.765	-2.174
Atkins	None	1.469*	.5358	.020	.174	2.765
	Vegetarian	-2.000*	.5358	<.001	-3.296	-.704
Vegetarian	None	3.469*	.5358	<.001	2.174	4.765
	Atkins	2.000*	.5358	<.001	.704	3.296

Based on observed means.
The error term is Mean Square(Error) = 8.613.

*. The mean difference is significant at the .05 level.



Mehrfaktorielle ANOVA - Ergebnisse

Übung an der Tabelle

Tests of Between-Subjects Effects

Dependent Variable: Observed weigth loss in kilos over last 2 months

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1905.037[a]	8	238.130	27.647	<.001	.564
Intercept	3624.868	1	3624.868	420.844	<.001	
diet	363.904	2	181.952		<.001	
exercise	1508.859	2	754.429		<.001	
diet * exercise	32.274	4	8.069		.444	
Error	1472.879	171	8.613			
Total	7002.784	180				
Corrected Total	3377.916	179				

a. R Squared = .564 (Adjusted R Squared = .544)

Und noch eine Übung

- Wir untersuchen den Effekt von Diät und Geschlecht auf Gewichtsverlust
- Öffnet dazu den Datensatz diet.sav

Und noch eine Übung

- Fragen:
  - Welche Voraussetzungen muss ich annehmen?
  - Wieviel Varianz klären die beiden Faktoren auf?
  - Was ist ein geeignetes Mass für die Effektstärke?
  - Welche Effektstärken kann ich berichten?
  - Welche Faktoren haben einen Effekt auf Gewichtsverlust?
  - Wie könnte ein Liniendiagramm dazu aussehen?
  - Benenne drei Nullhypothesen und drei Alternativhypothesen, die in diesem Modell getestet werden
  - Was wäre hier ein Alpha-Fehler, was wäre hier ein Beta-Fehler? Erkläre im Kontext der Aufgabenstellung

Tests of Between-Subjects Effects

Dependent Variable: Weight lost (kg)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	94.600[a]	5	18.920	3.519	.007
Intercept	1144.438	1	1144.438	212.874	<.001
Diet	49.679	2	24.840	4.620	.013
gender	.428	1	.428	.080	.779
Diet * gender	33.904	2	16.952	3.153	.049
Error	376.329	70	5.376		
Total	1654.350	76			
Corrected Total	470.929	75			

a. R Squared = .201 (Adjusted R Squared = .144)

Mehrfaktorielle ANOVA mit Messwiederholung

- Was ist der Unterschied zu einer «klassischen» mehrfaktoriellen ANOVA?
- Es gibt einen «within-subjects»-Effekt (Innersubjekteffekt) und einen «between-subjects»-Effekt (Zwischensubjekteffekt)

Within-Subjects Effect vs. Between-Subjects Effect

- Within-Subjects Effekt
  - Alles, was «innerhalb einer Versuchsperson» geschieht
    - Messwiederholung
    - Paneldaten
    - Prä/Post-Treatment Messungen
    - In den zu vergleichenden Gruppen befinden sich dieselben Personen
- Between-Subjects Effekt
  - Alles, was «zwischen den Versuchspersonen» geschieht
    - Vergleich zweier unabhängiger Gruppen
    - Experimental- vs. Kontrollgruppe
    - In den zu vergleichenden Gruppen befinden sich NICHT dieselben Personen

Mehrfaktorielle ANOVA

- Beispiel: Sie entwickeln ein Medikament. Um die Wirksamkeit des Medikaments zu prüfen, testen Sie vor der Einnahme und nach der Einnahme des Medikaments den Krankheitswert der Probanden.
- So haben wir es beim t-Test mit Messwiederholung gemacht. Dieselbe Gruppe wird vor und nach der Einnahme des Medikaments getestet. Aber: Ist das ausreichend? Was meint ihr?

Zusammenfassung

- t-Test mit unabhängigen Stichproben
- t-Test mit verbundenen Stichproben
- Einfaktorielle ANOVA
- Mehrfaktorielle ANOVA
- Mehrfaktorielle ANOVA mit Messwiederholung

Welcher Test ist angebracht?

- Sie begleiten eine Schulklasse. Sie haben etwas Sorge, dass die mathematischen Kompetenzen der Schüler sich im Laufe der Zeit nicht verbessern. Daher führen Sie einen Mathe-Test durch und wiederholen ihn im kommenden Jahr. Welchen Test wenden Sie an?

Welcher Test ist angebracht?

- Sie arbeiten bei der Polizei. Nachdem Sie die Ergebnisse der Polizeistreifen beobachtet haben, kommt Ihnen das Gefühl, dass einige Polizisten mehr Straftaten aufdecken als andere. Um das zu belegen, führen Sie einen statistischen Test durch. Welchen?

Welcher Test ist angebracht?

- Sie sind immer noch bei der Polizei. Die von Ihnen in der vorigen Aufgabe untersuchten Streifenpolizisten arbeiten jeweils an unterschiedlichen Standorten. Einige davon sind im ländlichen Raum angesiedelt, andere dagegen in der Stadt. Sie gehen davon aus, dass die Unterscheidung Stadt/Land ebenfalls einen Einfluss darauf haben kann, wieviele Straftaten eine Streife anzeigt. Daher führen Sie den Test erneut durch, ändern aber ein kleines Detail. Was ändern Sie und welchen Test führen Sie durch?

Welcher Test ist angebracht?

- Sie haben den leisen Verdacht, dass Franzosen im Mittel empathischer sind als Deutsche. Sicher sind Sie sich allerdings nicht, daher haben Sie sich einen psychometrischen Fragebogen besorgt, der die Empathiefähigkeit einer Person misst. Welches statistische Verfahren wenden Sie an?

Welcher Test ist angebracht?

- Sie erarbeiten eine Intervention, die Arbeitnehmer trainieren soll, effektiver im Homeoffice zusammenzuarbeiten. Dabei erwarten Sie, dass sich die Team Performance der Arbeitnehmer erhöhen soll. Ihr Auftraggeber möchte einen statistischen Nachweis, dass ihr Training die Team Performance auch tatsächlich erhöht. Daher haben Sie einen Fragebogen entwickelt, der die Team Performance zuverlässig misst. Jetzt ist nur noch die Frage nach dem statistischen Testverfahren offen. Welches wählen Sie und warum?

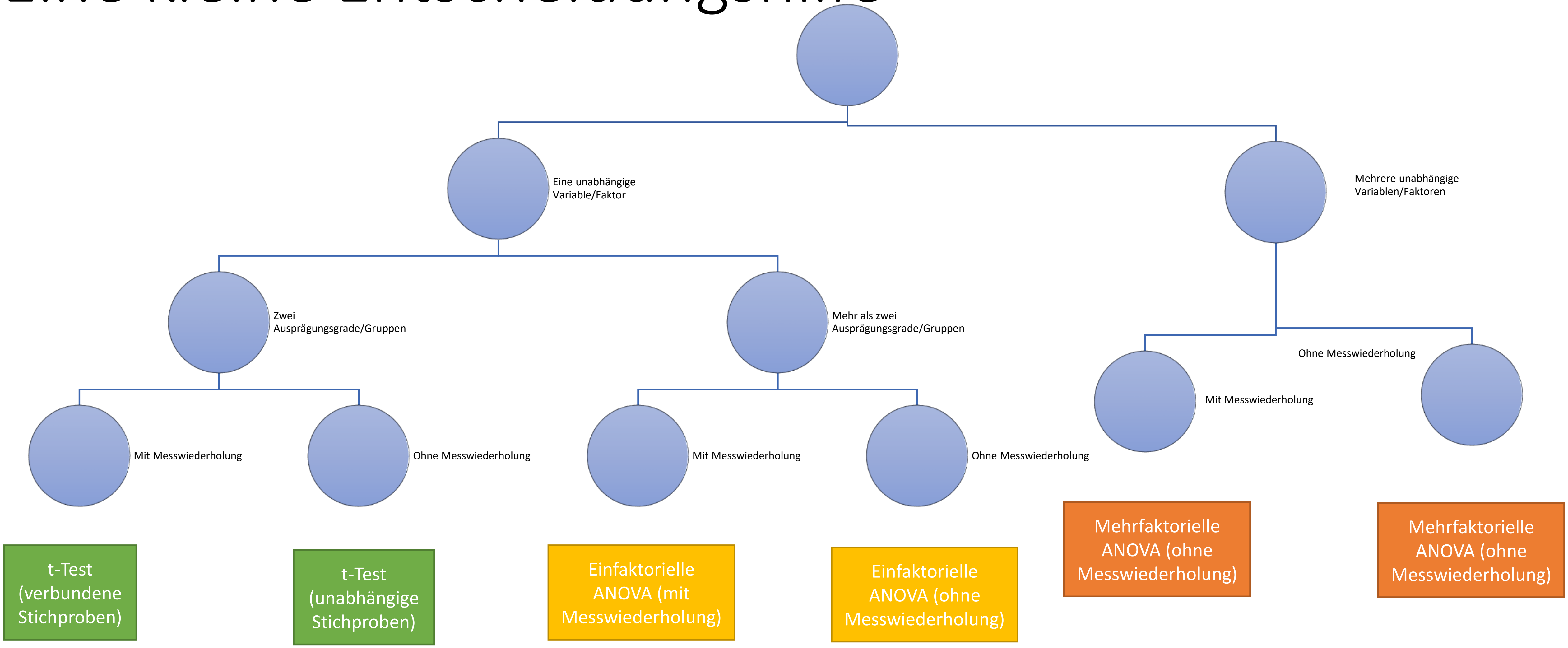
Wann wähle ich den t-Test (unabhängige Stichproben)?

- Zwei Gruppen
- Unabhängige Stichproben, d.h. kein Overlap zwischen den Versuchspersonen
- Klassisches Beispiel: Einfacher Mittelwertvergleich zwischen zwei Gruppen
  - Sind Mädchen kreativer als Jungen?
  - Hat Marke A eine positivere Aussenwirkung als Marke B?
  - …

Wann wähle ich den t-Test (verbundene Stichproben)?

- Zwei Gruppen
- Verbundene Stichproben (d.h. Messwiederholung, die Personen in den Gruppen sind dieselben)
- Klassisches Beispiel: Vorher-Nachher Mittelwertsvergleich innerhalb einer Gruppe
  - Hat sich Team Performance einer Gruppe innerhalb eines gewissen Zeitraums verändert?
  - Ist der Mittelwert für Depression innerhalb einer Gruppe im Sommer niedriger als im Winter?

Eine kleine Entscheidungshilfe



Ganz allgemein – was sind die Voraussetzungen für ANOVA und t-Test?

- Faktoren (abhängige Variable) kategorial
- Unabhängige Variable mindestens intervallskaliert
- Unabhängige Variable (annähernd) normalverteilt
- N zwischen den Gruppen annähernd gleich (wenn nicht, unbedingt auf Homoskedastizität prüfen!)
- Homoskedastizität (Varianz für alle Gruppen annähernd gleich)
- Bei Tests mit Messwiederholung: Sphärizität
- Grundsätzlich: ANOVA und t-Test sind gegenüber moderaten Verletzungen der Normalverteilung und Homoskedastizität relativ robust

Korrelationsmasse

- Manchmal wollen Sie den Zusammenhang zweier Variablen untersuchen
- Dafür eignet sich die Betrachtung einer Korrelation

Der **Korrelationskoeffizient** ist eine Maßzahl, die darüber Auskunft gibt, ob ein linearer Zusammenhang zwischen zwei Variablen besteht und berechnet sich über die Abweichung der Beobachtungen vom Mittelwert der beiden Stichproben. Er nimmt Werte zwischen –1 und 1 an.

Wenn der berechnete Wert 0 ergibt, existiert kein linearer Zusammenhang zwischen den beiden Variablen.

Bei –1 sagt man, die beiden Variablen sind negativ linear abhängig.

Bei 1 sind beide Variablen positiv linear abhängig.

Korrelationskoeffizient

- Korrelationskoeffizient nach Pearson
  - Linearer Zusammenhang
  - Stetige oder intervallskalierte Variablen
- Rangkorrelationskoeffizient nach Spearman
  - Linearer Zusammenhang muss nicht zwingend gegeben sein
  - Ordinale Variablen

Bestimmtheitsmass

- R-Quadrat

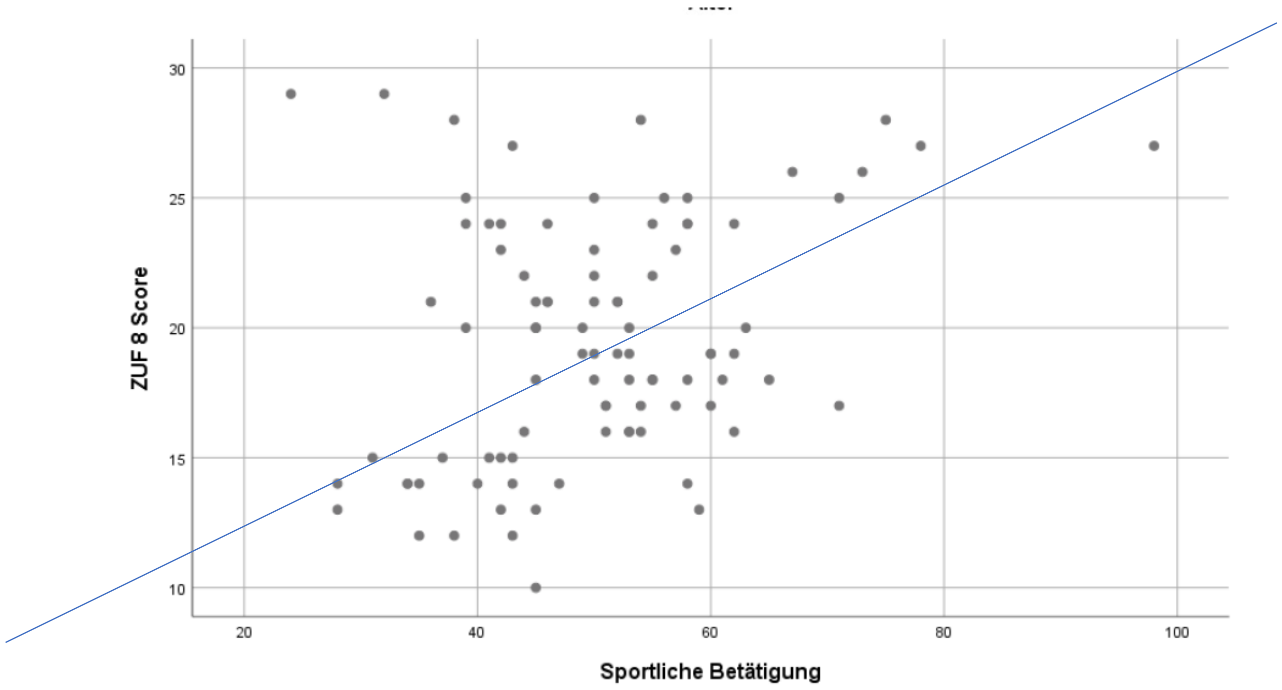
Das **Bestimmtheitsmaß** ist ein Maß zur Erklärung, wie viel Varianz der einen Variablen sich durch eine andere Variable erklären lässt (vgl. Janssen, Laatz 2007: 427).

Regressionsanalyse

- Das Prinzip der Regressionsanalyse ist, dass eine mathematische Gleichung aufgestellt wird, die den Zusammenhang zwischen zwei oder mehreren Variablen möglichst optimal widerspiegelt.
- Einfacher Regression:
  - Eine Einflussgrösse
- Multiple Regression
  - Mehrere Einflussgrössen

Lineare Regressionsanalyse

- Die lineare Regressionsanalyse ist anwendbar, wenn die Zielgröße metrisch ist, und ein linearer Zusammenhang zwischen Ziel- und Einflussgröße besteht.



Regressionsanalyse

- Regressionsgleichung

$$Y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i \ , \ i = 1, \ldots, n$$

- β_0= y-Achsenabschnitt («Wo trifft der Graph auf die y-Achse?»)
- β_p= Steigung der Regression («Um wieviel y steigt der Graph für 1 x?»)

Regression – praktische Anwendung

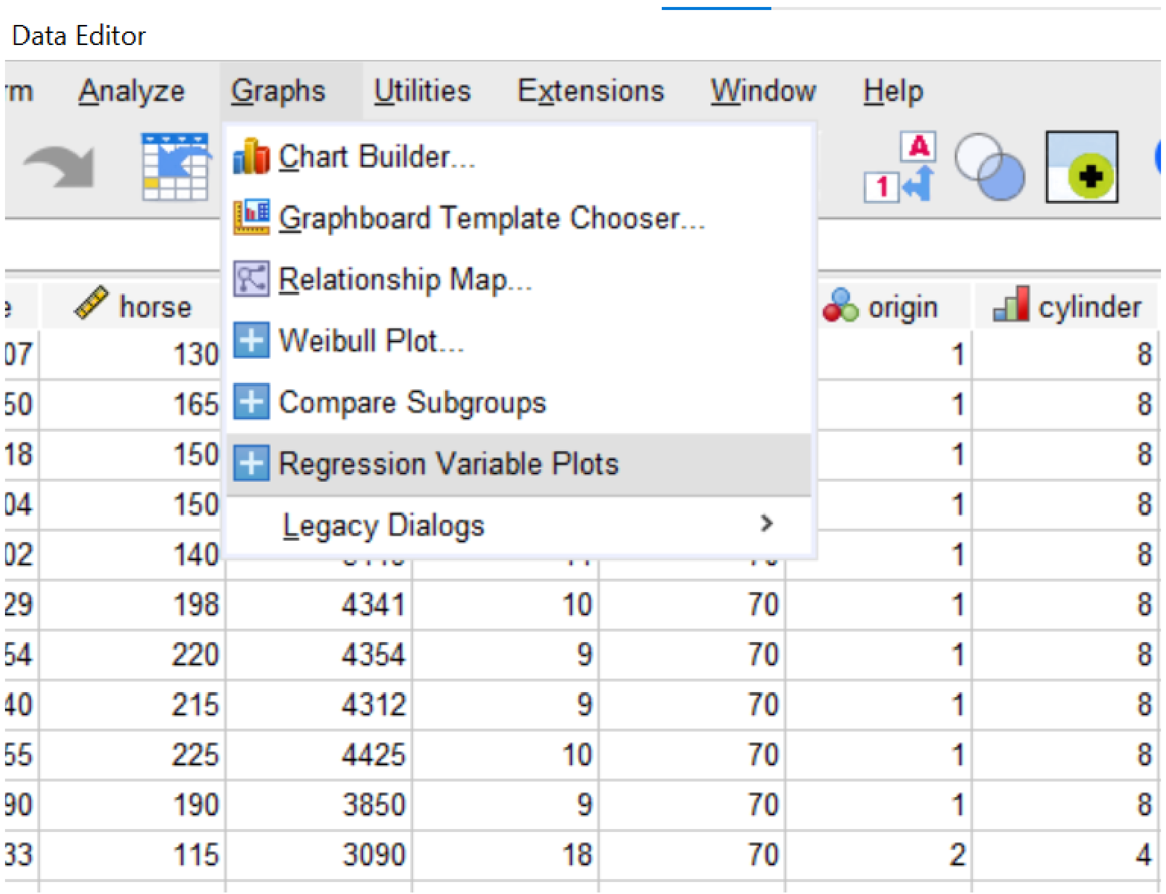
- Wir wollen herausfinden, ob Fahrzeuggewicht eine Auswirkung auf den Verbrauch hat
- Öffnet dazu wieder cars.sav

Regression – praktische Anwendung

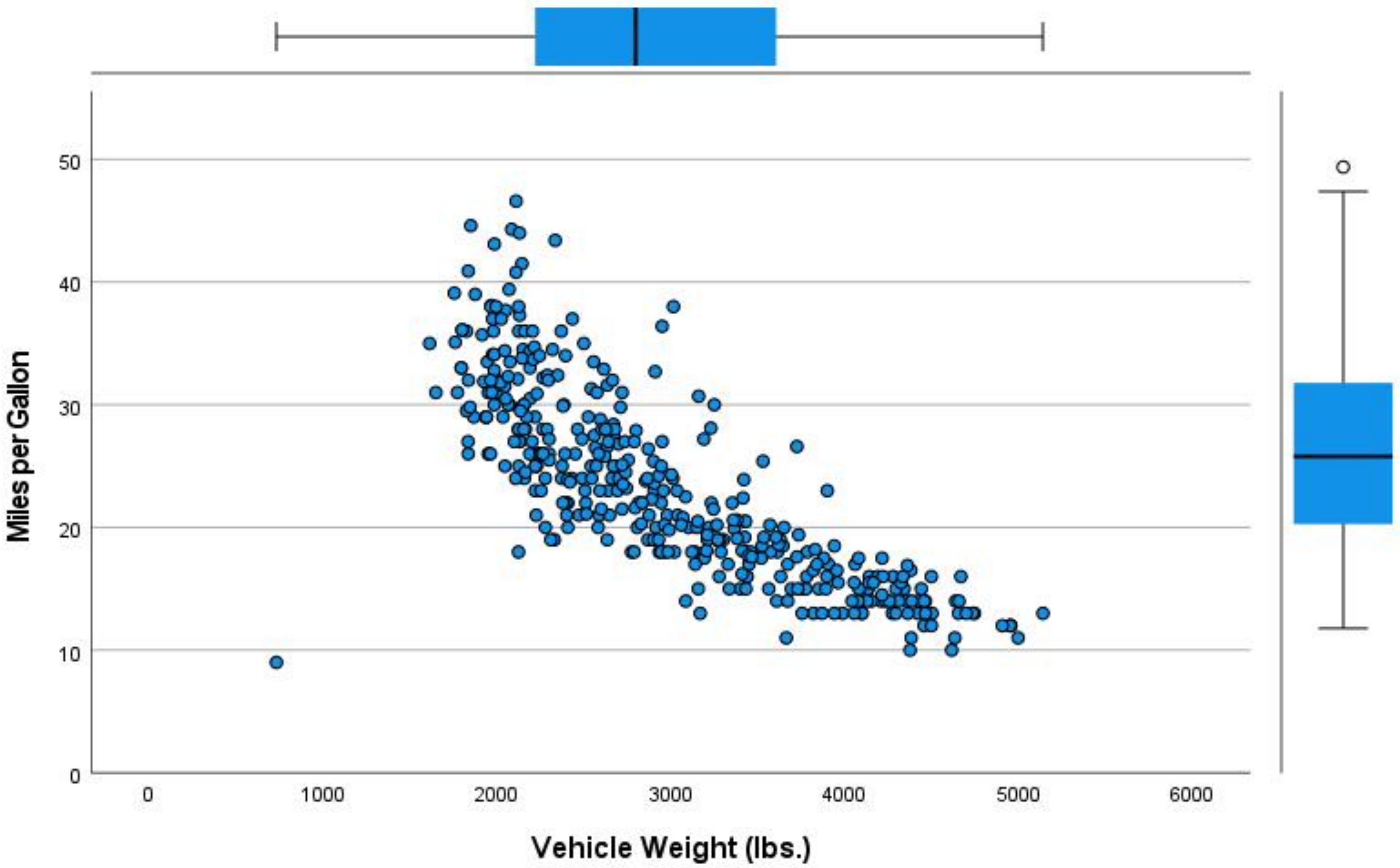
- Unabhängige Variable: Fahrzeuggewicht
- Abhängige Variable: Verbrauch
- Beide Variablen metrisch, intervallskaliert
- Pearson-Korrelationskoeffizient

Regression – praktische Anwendung

- Zunächst plotten wir den Zusammenhang

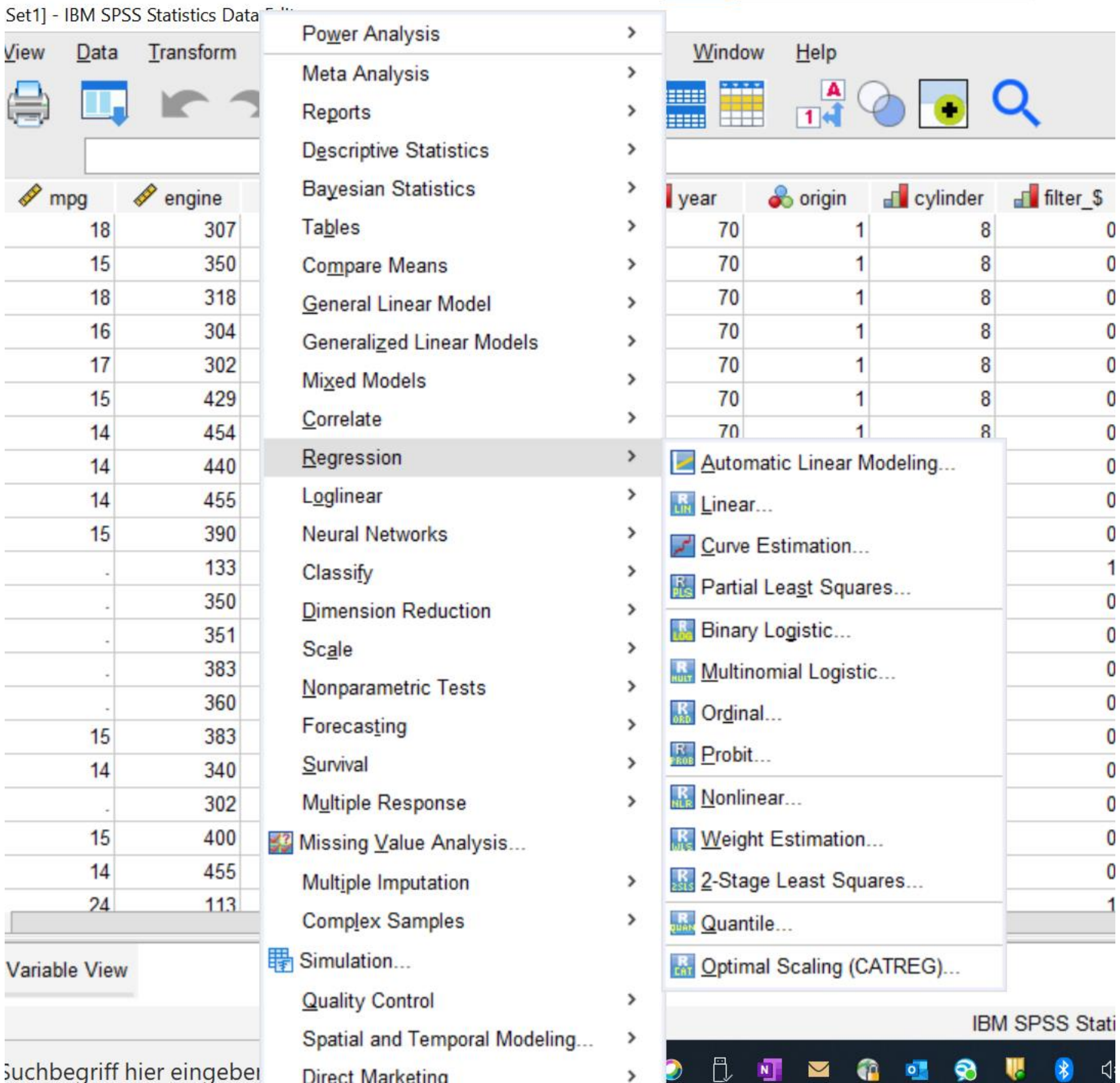


Regression – praktische Anwendung



Regression – praktische Anwendung

- Analysieren > Regression > Linear



Regression – praktische Anwendung

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.807[a]	.651	.650	4.622

a. Predictors: (Constant), Vehicle Weight (lbs.)

b. Dependent Variable: Miles per Gallon

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	45.492	.841		54.110	<.001
	Vehicle Weight (lbs.)	-.007	.000	-.807	-27.194	<.001

a. Dependent Variable: Miles per Gallon

- 65.1% Varianzaufklärung
- Gewicht und Verbrauch hängen signifikant zusammen
- Für jedes zusätzliche Pfund an Gewicht geht die Reichweite um 0.007 mpg zurück
- Der y-Achsenabschnitt liegt bei 45.5
- Das heisst, theoretisch würde ein Auto mit 0lbs Gewicht einen Verbrauch von 45.5 mpg haben