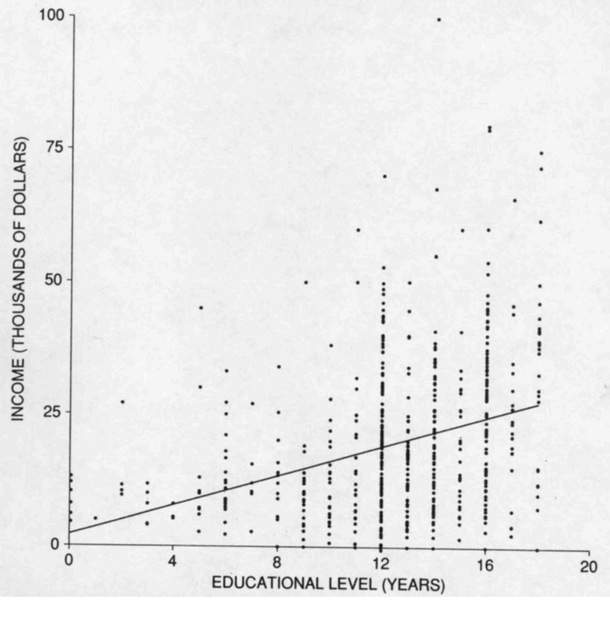
komplexe Forschungsmethoden

Das Zauberwort bei der Regression heißt **Vorhersage**!
↓
es geht darum, dass mit einer oder mehreren Variablen eine andere vorhergesagt werden soll, sofern ein linearer Zusammenhang gefunden wurde

21.Oktober - was ist lineare Regression?
statistisches Verfahren, mit dem versucht wird, eine beobachtete abhängige Variable durch eine oder mehrere unabhängige Variablen zu erklären,

- → Ziel: Vorhersage der Werte einer Variable (Kriterium, meist Y) bei Kenntnis der Werte der anderen Variablen (Prädiktor, meist X), z.B.
  - Vorhersage der Körpergröße aufgrund des Körpergewichts,
  - Vorhersage des Einkommens aufgrund des Bildungsniveaus,
  - Vorhersage der Intelligenz aufgrund der Intelligenz des Vaters ...
- → Y soll aufgrund von X vorhergesagt werden → Regression von Y auf X
- → Voraussetzung: der Zusammenhang zwischen X und Y ist linear,
- → Ergebnisse der Analyse:
  - Geradengleichung - liefert die Vorhersage,
  - Gütemaße der Vorhersage,
    - * Determinationskoeffizient,
      Maß für die Abweichungen der Vorhersagen eines Regressionsmodells von den empirischen Daten
    - * Standardschätzfehler
      ein Maß dafür, wie stark die wahren y-Werte von den vorhergesagten Werten abweichen
    - * (Residuen-Plot)
      das Residuum ist die Differenz zwischen einem vorhergesagten Wert und einem beobachteten Wert

Figure 1. The regression line. The scatter diagram shows income and education for a representative sample of 637 California men age 25–29 in 1988.



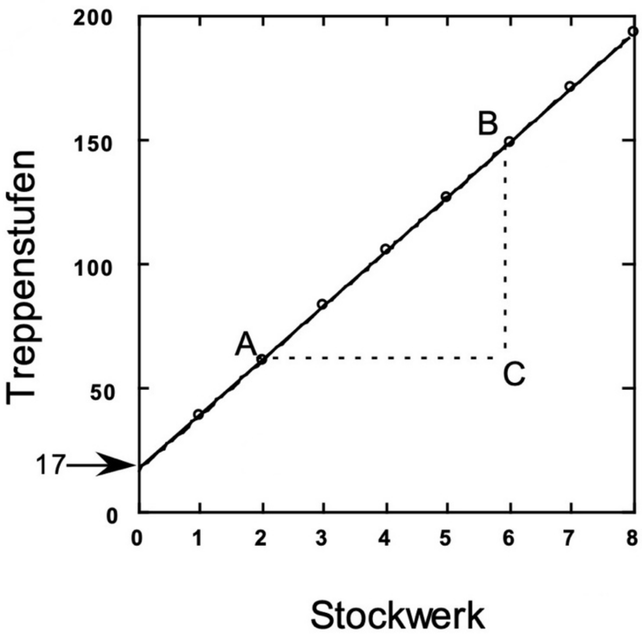
Bestimmen einer Geradengleichung für einen deterministischen linearen Zusammenhang
Beispiel: Vorhersage der Anzahl der Treppenstufen Y bei Kenntnis des Stockwerks X

$$y_i = b_{yx}\, x_i + a_{yx}$$

↓ Steigung der Regressionsgeraden ↓ BC/AC
Anzahl der Stufen pro X-Stockwerke
22Stufen pro Stockwerk
17Stufen zum Eingang

↘ Ordinatenabschnitt - gibt an, welchen Wert Y am Punkt x=0 annimmt ↓ Y-Wert für X=0
Anzahl der Stufen bis zum Erdgeschoss

- Geradengleichung: $y_i = 22x_i + 17$
- Treppen zum 4.Stock: $y_4 = 22 * 4 + 17 = 105$



Bestimmen einer Geradengleichung für einen probabilistischen linearen Zusammenhang I
tatsächlicher Wert = vorhergesagter Wert + Fehler

$$y_i \quad = \quad \hat{y}_i \quad + \quad e_i$$

(Häubchen auf y → Schätzung)

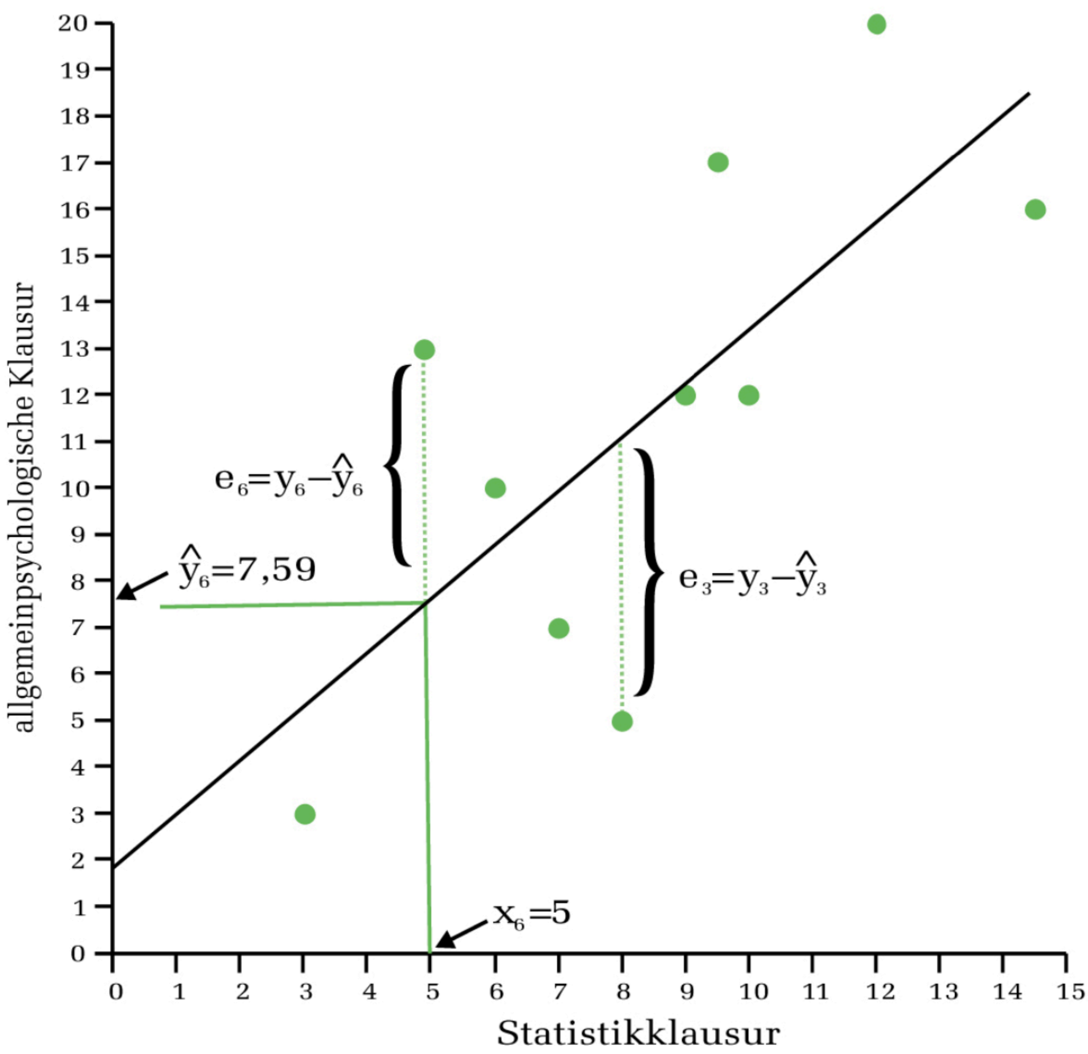
vorhergesagter Wert: $\hat{y}_i = b_{yx}\, x_i + a_{yx}$
Fehler (Residuum): $e_i = y_i - \hat{y}_i$
nach Einsetzen: $y_i = b_{yx}\, x_i + a_{yx} + e_i$

→ Problem: wie kann der Fehler möglichst klein gehalten werden?
→ Lösung bei der linearen Regression: Methode der kleinsten Quadrate → minimieren der Summe der Abweichungsquadrate → Quadrate der Differenzen zwischen tatsächlichen und vorhergesagten Werten,

$$\Rightarrow \Sigma e_i^2 = \Sigma(y_i - \hat{y}_i)^2 = \text{min!}$$

Abweichungsquadrate sind minimal, wenn $\hat{y}_i = b_{yx} x_i + a_{yx}$ mit:

$$b_{yx} = \frac{\mathrm{cov}(x,y)}{s_x^2} = \frac{s_{xy}}{s_x^2} \quad \text{und} \quad a_{yx} = \bar{y} - b_{yx}\bar{x}$$



<u>Beziehung zwischen Korrelation und Regression</u>

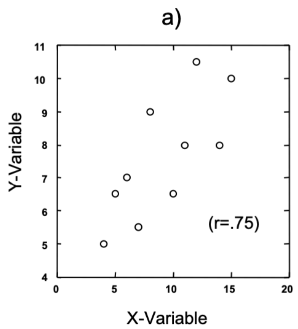
→ Korrelation:
- hier wird die Stärke des Zusammenhangs zwischen zwei verschiedenen Variablen berechnet,
- die Aussage, die bei der Korrelation getroffen werden kann, ist also, dass bestimmte Werte auf der einen Variable mit bestimmten Werten auf der anderen Variable zusammenhängen - dadurch wird es möglich, eine Vorhersage zu treffen, ohne jedoch eine Kausalbeziehung herzustellen,

→ Regression:
- sie basiert auf der Korrelation und ermöglicht die bestmögliche Vorhersage für eine Variable,
- im Gegensatz zur Korrelation muss hierbei festgelegt werden, welche Variable durch eine andere Variable vorhergesagt werden soll,

- die Variable, die vorhergesagt werden soll nennt man bei der Regression *Kriterium*,
- die Variable, die für die Vorhersage eingesetzt wird, bezeichnet man als *Prädiktor*,
- anhand des Prädiktors wird demzufolge das Kriterium vorhergesagt,

<u>Rechenbeispiel</u>

Datenpunkt (i)	Verbaler Test	Numerischer Test
1	4,0	5,0
2	5,0	6,5
3	6,0	7,0
4	7,0	5,5
5	8,0	9,0
6	10,0	6,5
7	11,0	8,0
8	12,0	10,5
9	14,0	8,0
10	15,0	10,0



→ Vorhersage des numerischen Tests Y aufgrund des verbalen Tests X

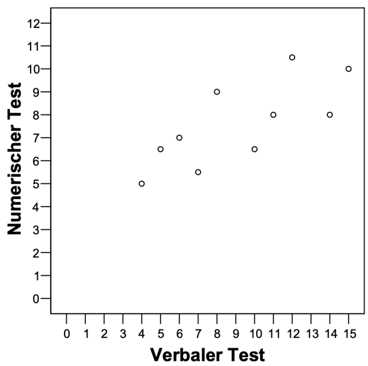
Datenpunkt (i)	Verb. (x)	Num. (y)
1	4,0	5,0
2	5,0	6,5
3	6,0	7,0
4	7,0	5,5
5	8,0	9,0
6	10,0	6,5
7	11,0	8,0
8	12,0	10,5
9	14,0	8,0
10	15,0	10,0
N = 10	Σ = 92	Σ = 76
	AM = 9,2	AM = 7,6

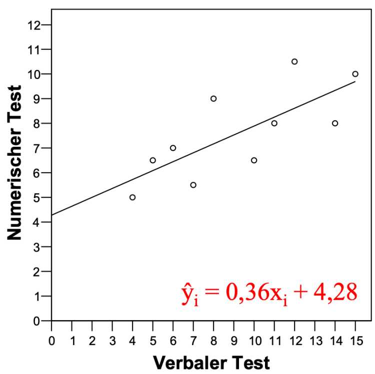
$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$	$(x_i-\bar{x})(y-\bar{y})$
-5,2	-2,6	27,04	6,76	13,52
-4,2	-1,1	17,64	1,21	4,62
-3,2	-0,6	10,24	0,36	1,92
-2,2	-2,1	4,84	4,41	4,62
-1,2	1,4	1,44	1,96	-1,68
0,8	-1,1	0,64	1,21	-0,88
1,8	0,4	3,24	0,16	0,72
2,8	2,9	7,84	8,41	8,12
4,8	0,4	23,04	0,16	1,92
5,8	2,4	33,64	5,76	13,92
Σ = 0	Σ = 0	Σ = 129,60	Σ = 30,40	Σ = 46,8
		$s^2 = 12,96$	$s^2 = 3,04$	$s_{xy} = 4,68$
		s = 3,6	s = 1,74	r = .75

$$b_{yx} = \frac{\operatorname{cov}(x,y)}{s_x^2} = \frac{4,68}{12,96} = 0,36 \quad \text{oder} \quad b_{yx} = r \cdot \frac{s_y}{s_x} = 0.75 \cdot \frac{1,74}{3,60} = 0,36$$

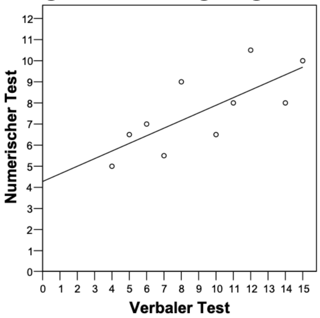
$$a_{yx} = \bar{y} - b_{yx}\bar{x} = 7,6 - 0,36 \cdot 9,2 = 4,28 \qquad \hat{y}_i = b_{yx} \cdot x_i + a_{yx} = 0,36x_i + 4,28$$

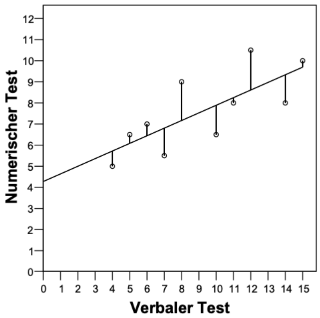
→ die Regressionsgerade im Streudiagramm:





→ Bestimmung der vorhergesagten Werte und Residuen:





Datenpunkt (i)	X	Y	$X-\overline{X}$	$Y-\overline{Y}$	$(X-\overline{X})(Y-\overline{Y})$	$\hat{y}_i$	Residuen
1	4,0	5,0	-5,2	-2,6	13,52	5,72	-0,72
2	5,0	6,5	-4,2	-1,1	4,62	6,08	0,42
3	6,0	7,0	-3,2	-,6	1,92	6,44	0,56
4	7,0	5,5	-2,2	-2,1	4,62	6,80	-1,30
5	8,0	9,0	-1,2	1,4	-1,68	7,16	1,84
6	10,0	6,5	,8	-1,1	-0,88	7,88	-1,38
7	11,0	8,0	1,8	,4	0,72	8,24	-0,24
8	12,0	10,5	2,8	2,9	8.12	8,60	1,90
9	14,0	8,0	4,8	,4	1,92	9,32	-1,32
10	15,0	10,0	5,8	2,4	13,92	9,68	0,32
Mittelwert	9,2	7,6			**COV(X,Y) = 4,68**		

$$\hat{y}_i = 0{,}36x_i + 4{,}28$$

→ zum Vergleich: Regression von X auf Y

$$r \cdot \frac{s_y}{s_x} = \frac{\mathrm{cov}(x,y)}{s_x \cdot s_y} \cdot \frac{s_y}{s_x} = b_{yx}$$

b_{yx} und a_{yx}: Vorhersage des y-Werts aufgrund des x-Werts

$$a_{yx} = \bar{y} - b_{yx}\bar{x}$$

$$r \cdot \frac{s_x}{s_y} = \frac{\mathrm{cov}(x,y)}{s_x \cdot s_y} \cdot \frac{s_x}{s_y} = b_{xy}$$

b_{xy} und a_{xy}: Vorhersage des x-Werts aufgrund des y-Werts

$$a_{xy} = \bar{x} - b_{xy}\bar{y}$$

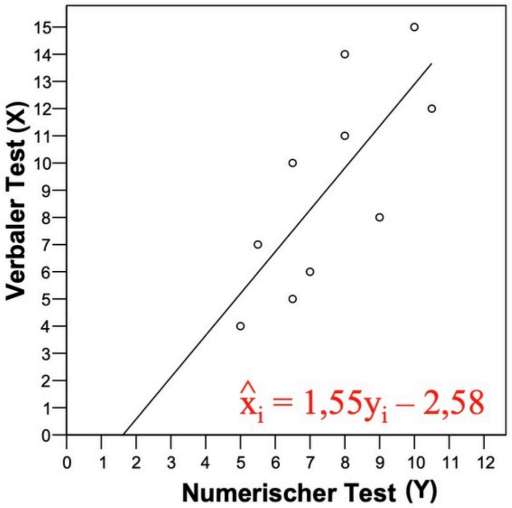
$$\hat{x} = b_{xy} \cdot y_i + a_{xy}$$

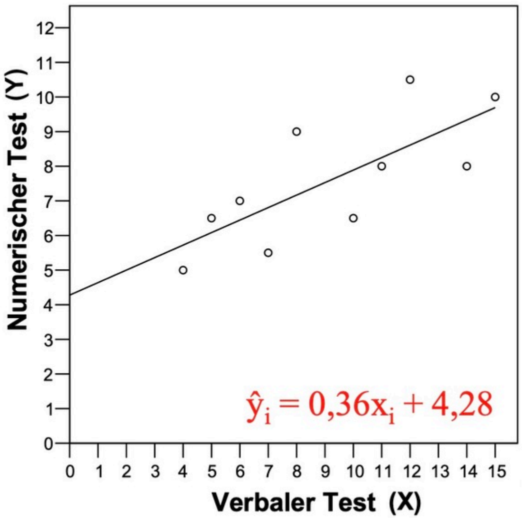
→ in unserem Fall: Vorhersage des verbalen Tests X aufgrund des numerischen Tests Y

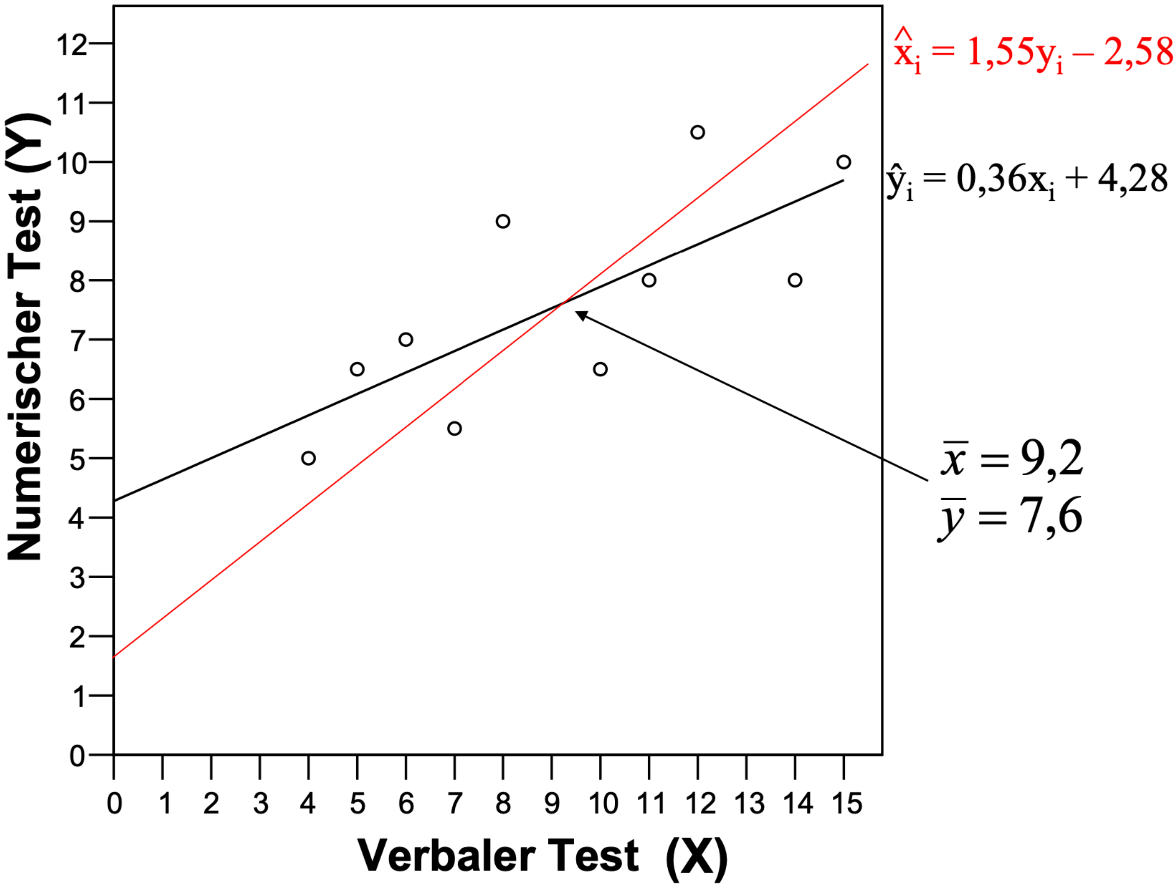
$r = 0{,}75 \qquad s_x = 3{,}6 \qquad s_y = 1{,}74 \qquad \text{Mittelwert } x = 9{,}2 \qquad \text{Mittelwert } y = 7{,}6$

$b_{xy} = r * s_x/s_y \qquad b_{xy} = 0{,}75 * 3{,}6/1{,}74 = 1{,}55$

$a_{xy} = \text{Mittelwert } x - b_{xy} * \text{Mittelwert } y = 9{,}2 - 1{,}55 * 7{,}6 = -2{,}58$







→ einige Eigenschaften der Regression:

- $\Sigma(y_i - \hat{y}_i)^2 = \Sigma e_i^2 = \min$ gemäß Methode der kleinsten Quadrate

- $\Sigma(y_i - \hat{y}_i) = \Sigma e_i = 0$ Summe und Mittelwert der Fehler sind also 0

- $x_i = \bar{x} \;\Rightarrow\; \hat{y}_i = \bar{y}$ Am Mittelwert der x-Werte wird also der Mittelwert der y-Werte vorhergesagt.
$\Rightarrow$ Die Regressionsgerade verläuft durch den Schwerpunkt der bivariaten Verteilung.

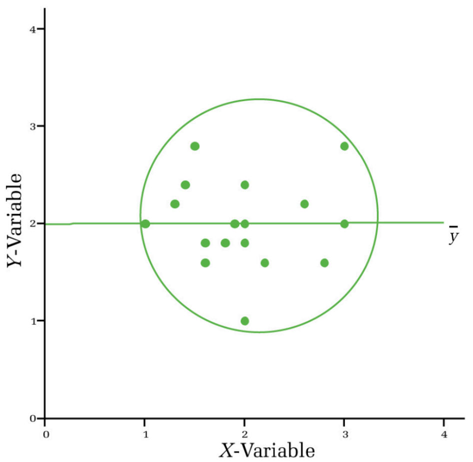
→ die Regressionsgerade bei einer Korrelation von 0

$$b_{yx} = r \cdot \frac{s_y}{s_x} = 0 \cdot \frac{s_y}{s_x} = 0$$

• Bei r = 0 verläuft die Regressionsgerade also parallel zur x-Achse.

$$a_{yx} = \bar{y} - b_{yx}\bar{x} = \bar{y} - 0 \cdot \bar{x} = \bar{y}$$

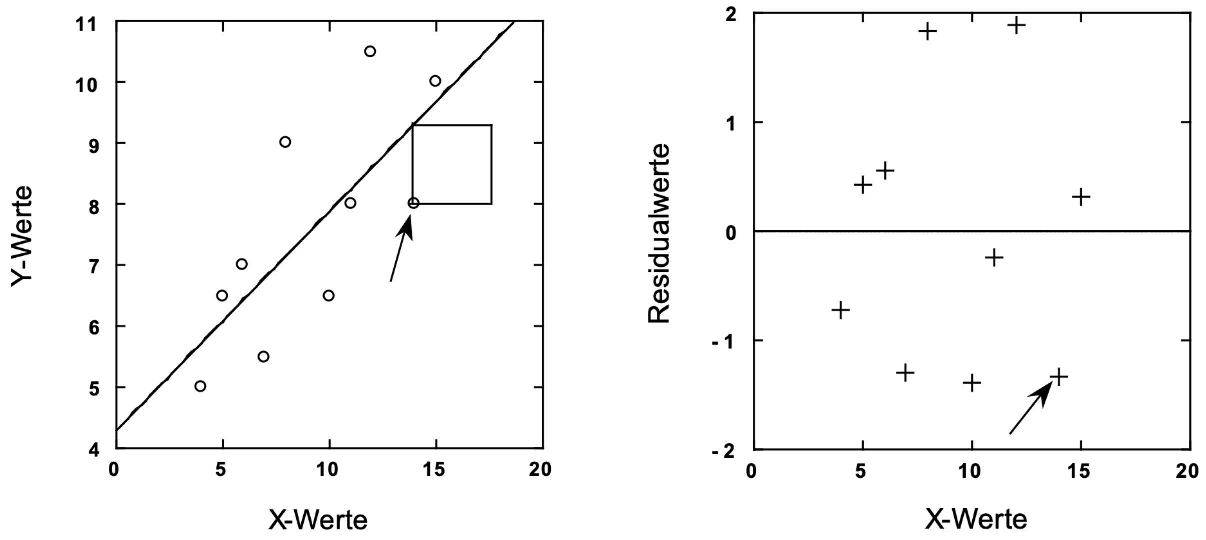
• Bei r = 0 liegt die Regressionsgerade auf der Höhe des Mittelwerts von y.



besteht zwischen den Variablen kein Zusammenhang, so wird stets der Mittelwert von y vorhergesagt,

<u>wie gut ist die Vorhersage?</u>
1. Residuen-Plot
   - das Konzept der Varianzzerlegung
2. Determinationskoeffizient
3. Standardschätzfehler

1. Residuen-Plot - die Residuen werden gegen die x-Werte abgetragen
Optimum: alle Residualwerte haben den Wert 0



Varianzzerlegung

→ die Regressionsrechnung zielt darauf, Unterschiede zwischen Personen vorherzusagen oder aufzuklären,

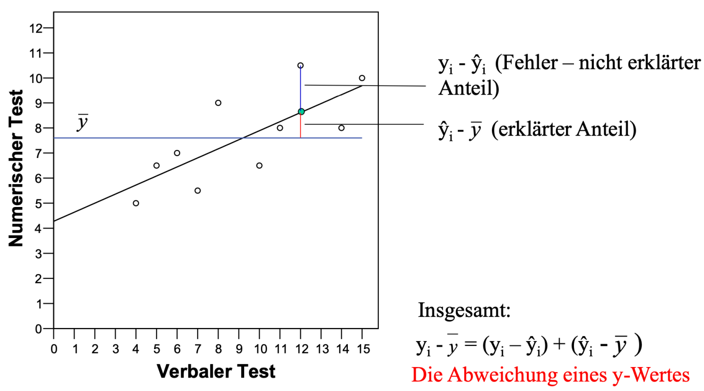
→ dies ist generell ein wesentliches wissenschaftliches Ziel, das natürlich nicht nur mit der Regressionsrechnung verfolgt wird,

→ Maß für die Unterschiedlichkeit hinsichtlich einer Variablen: **Varianz**

$$s_y^2 = \frac{\sum_i^n (y_i - \bar{y})^2}{n}$$

→ die Varianz wird in *vorhersagbare* und *nicht-vorhersagbare* Anteile zerlegt,

Grundlage der Varianz: $y_i - \bar{y}$



Insgesamt:

$y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$

Die Abweichung eines y-Wertes vom Mittelwert kann also in einen erklärten und einen nicht-erklärten Anteil zerlegt werden! 20

• Vorhergesagter Wert $\hat{y}_i$

→ auf der Grundlage der vorhergesagten und nicht-vorhergesagten Abweichungen können weitere Varianzen berechnet werden,

→ Varianz der vorhergesagten Werte (Regressionsvarianz): $s_{\hat{y}}^2 = \frac{\sum_i^n (\hat{y}_i - \bar{y})^2}{n}$

→ Fehlervarianz: $s_e^2 = \frac{\sum_i^n (y_i - \hat{y}_i)^2}{n}$

$$\frac{\sum (y_i - \bar{y})^2}{n} = \frac{\sum (\hat{y}_i - \bar{y})^2}{n} + \frac{\sum (y_i - \hat{y}_i)^2}{n}$$

$$s_y^2 \quad = \quad s_{\hat{y}}^2 \quad + \quad s_e^2$$

Gesamtvarianz = Regressionsvarianz + Fehlervarianz

Die Gesamtvarianz kann also in die erklärte Varianz und die Fehlervarianz zerlegt werden!

<

2. Determinationskoeffizient (r2): Anteil der Varianz der y-Werte, der durch die x-Werte aufgeklärt (determiniert) werden kann,

Erklärte Varianz: $s_{\hat{y}}^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{n} = r^2 \cdot s_y^2$

Anteil erklärter Varianz (*Determinationskoeffizient*): $r^2 = \frac{s_{\hat{y}}^2}{s_y^2}$

Optimum: alle tatsächlichen Werte liegen auf der Regressionsgeraden → $r^2 = 1$.

Anteil der Fehlervarianz: $1 - r^2 = \frac{s_e^2}{s_y^2}$

r-Quadrat → r2

- gibt an, wie gut die Vorhersage mit dem Regressionsmodell ist,
- wie gut kann mit den Prädiktoren die aV erklärt werden?
- Anteil an der GesamtStreuung der aV, der durch die Regression erklärt werden kann,
- Anteil an der Gesamtvariabilität der werte, der durch die Regression erklärt werden kann,
- Determinationskoeffizient,
- Bestimmtheitsmaß,
- Maß für die Güte der Anpassung,
- Maß dafür, wie gut die Regressionsgerade in diese Punktewolke hineinpasst,
- je größer r2 desto besser die Vorhersage von y,
- immer zwischen 0 und 1,
- 0 → nichts wird vorhergesagt,
- 1 → perfekte Vorhersage,
- bei der einfachen Regression entspricht r2 immer dem quadrierten Korrelationskoeffizienten r,
- Basis für die Berechnung der Effektstärke f,
- basiert auf der Zerlegung der Streuung der Werte um den Mittelwert von y,

3. Standardschätzfehler (s_e): Maß für die Streuung (die typische Abweichung) der tatsächlichen y-Werte um die Regressionsgerade - er entspricht der Wurzel aus der Fehlervarianz:

Da $1-r^2 = \frac{s_e^2}{s_y^2}$ gilt zudem: $s_e = s_y\sqrt{1-r^2}$

→ Optimum: alle tatsächlichen Werte liegen auf der Regressionsgeraden $s_e = 0$

Vorhersage vs. Kausalität

→ wenn die Prädiktorvariable eine gute Vorhersage für die Kriteriumsvariable liefert, muss das nicht heißen, dass eine Kausalbeziehung zwischen Prädiktor und Kriterium besteht!

Regression mit z-standardisierten Variablen

→ in der Psychologie gibt es idR keine etablierten Maßeinheiten. Variablen wie Aggression, Befinden, Intelligenz (usw.) können auf Skalen von 0- 7, 0- 10, 20- 80, 55- 145 (usw.) gemessen werden → Untersuchungen mit identischen Fragestellungen führen in der Regressionsrechnung unter Umständen zu nicht direkt vergleichbaren Ergebnissen,

→ Vergleich der Ergebnisse aus Regressionsanalysen kann auch dann wünschenswert sein, wenn unterschiedliche Variablen untersucht werden,

- verändert sich die Reaktionszeit stärker als die Konzentrationsleistung, wenn der Schlafentzug um eine Stunde verlängert wird?

→ Lösung: z-Standardisierung der Variablen

- alle Variablen erhalten unabhängig von ihrer ursprünglichen Maßeinheit dieselbe neue Maßeinheit: ihre Standardabweichung

der Regressionseffekt

→ ein vorhergesagter y-Wert wird weniger stark vom Mittelwert der y-Werte abweichen als der

→ entsprechende x-Wert vom Mittelwert der x-Werte abweicht,

→ einzige Ausnahme: X und Y korrelieren perfekt

→ r = 1 oder -1

→ die größten Väter haben nicht die größten Söhne!

→ Illustration: Regression mit z-standardisierten Variablen

- $z_{xi} = 2$

$$\hat{z}_{y_i} = r_{xy} \cdot z_{x_i}$$
$$= 1{,}0 \cdot 2 = 2$$
$$= 0{,}8 \cdot 2 = 1{,}6$$
$$= 0{,}5 \cdot 2 = 1$$
$$= 0{,}2 \cdot 2 = 0{,}4$$
$$= 0{,}0 \cdot 2 = 0$$

→ aufgrund der Größe der Väter X wird die Größe der Söhne Y vorhergesagt,

- Addition der Mittelwerte = 175cm
- Standardabweichungen → $s_x + s_y = 8$cm
- Korrelation → r = 0,6
- x = 183cm → $z_x = 1$
- vorhergesagter z_y-Wert → $\hat{z} = r * z_x = 0{,}6 * 1 = 0{,}6$
- vorhergesagter Wert in cm → 175cm + 0,6 * 8cm = 179,8cm

→ veränderte Ausgangswerte:

- Mittelwert für x → 175cm, Mittelwert für y → **178cm**
- Standardabweichungen → $s_x = 8$cm, $s_y =$ **10cm**
- Korrelation → r = 0,6

– $x = 183cm \rightarrow z_x = 1$
– vorhergesagter z_y-Wert → $\hat{z} = r * z_x = 0{,}6 * 1 = 0{,}6$
– vorhergesagter Wert in cm → 178cm + 0,6 * 10cm = 184cm

→ auch hier gilt aber, dass die relative Abweichung vom Mittelwert beim Sohn geringer ist als beim Vater!

Literatur
Sedlmeier P. & Renkewitz, F. (2018). *Forschungsmethoden und Statistik*. München: Pearson Studium
→ Kapitel 8

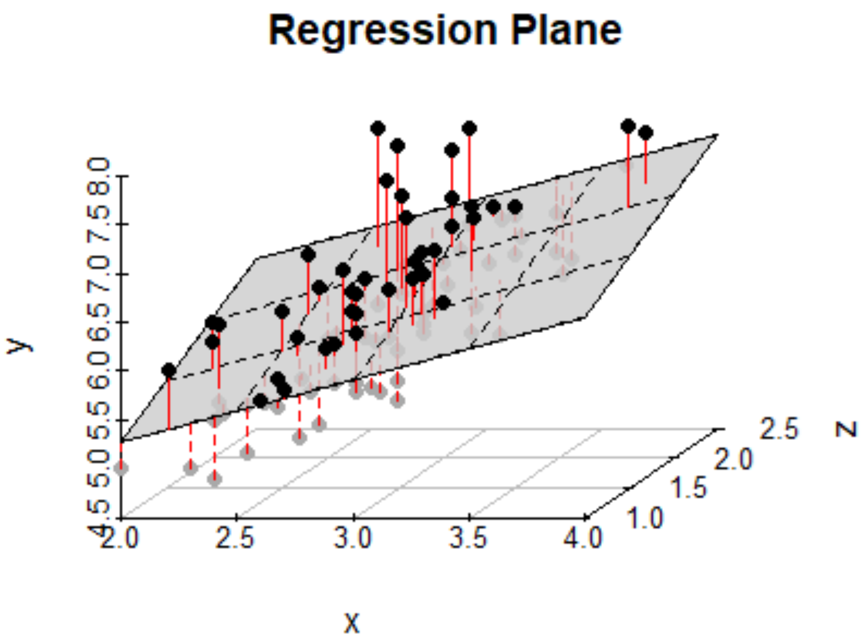
28.Oktober → multiple Regression
wir erweitern nun das *Datenmodell*, indem wir die Zahl der Prädiktoren erhöhen,

Modell mit zwei Prädiktoren: $$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i}$$

Modell mit beliebig vielen Prädiktoren: $$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki}$$

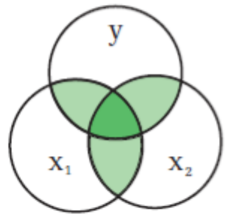
- de facto ist die sinnvolle Zahl der Prädiktoren durch die Stichprobengröße begrenzt!

→ mit zwei Prädiktoren wird aus der Regressionsgeraden eine Regressionsebene:



→ mit mehr als zwei Prädiktoren: Hyperebene
- für *k* Prädiktoren benötigen wir einen (*k*+1)-dimensionalen Raum
- passt schlecht ins Bild...
- ändert aber auch nichts an den Prinzipien - an den Basics,

der zentrale (zusätzliche) Aspekt: die Prädiktoren werden (in der Regel) untereinander korrelieren,



→ Kreis: Varianz einer Variable

- die Kreise sind gleich groß, weil wir (für den Moment) davon ausgehen, dass die Variablen (z-)standardisiert sind → Varianz von 1!
- Schnittmenge zweier Kreise: gemeinsame Varianz → Determination (r^2)
- gesamte Schnittmenge von x_1 **oder** x_2 mit y: Varianz in y, die durch x_1 **und** x_2 aufgeklärt wird → multiple Determination (R^2)
- offensichtlich gilt im Beispiel:

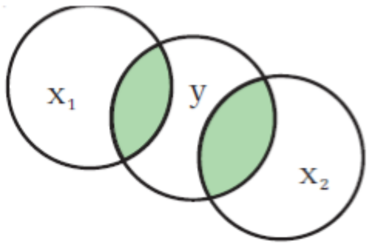
$$R^2_{Y.12} < r^2_{Y1} + r^2_{Y2}$$

in Worten:

- die multiple Determination $R^2_{Y.12}$ (die Varianzaufklärung) aus der Regression von y auf x_1 und x_2 ist kleiner als die Summe der Determinationen aus den beiden einfachen Regressionen von y auf x_1 und y auf x_2
- inhaltlich: x_1 und x_2 enthalten redundante Information (die Schnittmenge aller 3 Kreise), die zur Vorhersage von y nicht zweimal genutzt werden kann
- gilt immer, wenn Prädiktoren untereinander (und mit dem Kriterium) korrelieren,

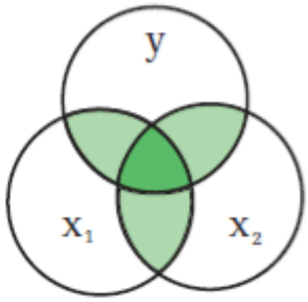
Gegenbeispiel: die Prädiktoren korrelieren nicht

$$R^2_{Y.12} = r^2_{Y1} + r^2_{Y2}$$



→ die multiple Determination entspricht der Summe der Determinationen aus den einfachen Regressionen,

- der Varianzanteil in y, der nur durch x_2 erklärt werden kann,
- oder: der eigenständige Beitrag von x_2 zur
- Varianzaufklärung von y,
- technisch: Semipartialdetermination
- Wurzel daraus: Semipartialkorrelation



→ das Regressionsgewicht eines Prädiktors hängt von seinem eigenständigen Beitrag zur Varianzaufklärung ab!

→ was bedeutet also das Regressionsgewicht b_1?
   - wenn x_1 um eine Einheit steigt und alle anderen Prädiktoren konstant bleiben, erwarten wir eine Änderung in y um b_1 -Einheiten,
   - tatsächlich bleiben die anderen Prädiktoren nicht konstant - die Prädiktoren korrelieren ja untereinander,
   - in diesem Sinn eine rein *theoretische* oder *statistische* Größe

einige Implikationen:

→ im besten Fall sind Prädiktoren untereinander nicht (nur schwach) korreliert,

→ das Regressionsgewicht eines Prädiktors (und sein Beitrag zur Varianzaufklärung) hängt davon ab, welche anderen Prädiktoren in das Modell aufgenommen werden (und davon, wie der fragliche Prädiktor mit diesen übrigen Prädiktoren korreliert),
   - sehr guter Grund, sich gründlich Gedanken darüber zu machen, welche Prädiktoren einbezogen werden,

→ die Reihenfolge, mit der Prädiktoren aufgenommen werden, ist relevant,
   - Möglichkeiten im *SPSS-Sprech*:
      * forced entry - so viel wie *gemeinsam* oder *gleichzeitig*
      * hierarchisch
      * schrittweise - das ist böse! - siehe unten...

→ sind Prädiktoren untereinander stark korreliert, entsteht das Problem der **Multikollinearität**:
   - die Prädiktoren erklären überwiegend dieselbe Varianz und können die Determination also nicht erhöhen,
   - zudem (und wichtiger): der eigenständige Beitrag jedes Prädiktors (und damit sein Regressionsgewicht) kann schlechter (ungenauer) geschätzt werden,
   - das heißt: der Standardfehler der Regressionsgewichte wächst!
      * exzellente Gelegenheit, sich zu fragen, was ein Standardfehler nochmal war!
   - Tipp: **nicht** mit dem Standard**schätz**fehler verwechseln!
   - Multikollinearität ist also unbedingt zu vermeiden...

Interpretation von Regressionskoeffizienten
- wir sagen Albumverkäufe (in 1000Stück) aus den folgenden drei Prädiktoren vorher - Beispiel nach Field:
   * advertising budget - Werbe-Etat in 1000Pfund,
   * No. of plays in the radio - wie oft wurden Songs des Albums im Radio gespielt?
   * Image - als wie attraktiv wurde das Band-Foto auf dem Album-Cover im Mittel beurteilt?

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-26,613	17,350		-1,534	,127
	Advertsing budget (thousands)	,085	,007	,511	12,261	,000
	No. of plays on radio	3,367	,278	,512	12,123	,000
	Band image rating (0-10)	11,086	2,438	,192	4,548	,000

a. Dependent Variable: Al usands)

Wenn ein Albumsong ein zusätzliches Mal im Radio gespielt wird und sich die anderen Prädiktoren nicht ändern, erhöhen sich sich die Albumverkäufe um 3,367 Einheiten (also 3367 Stück).

standardisierte Gewichte:

Die unstandardisierten Gewichte sind untereinander nicht vergleichbar – sie haben ja unterschiedliche Einheiten. 1000 zusätzliche Pfund für Werbung sind nur schwerlich mit „ein weiteres Mal im Radio gespielt werden“ oder „ein Punkt bessere Bewertung des Bandfotos“ zu vergleichen.

Wenn sich die Häufigkeit, mit der ein Albumsong im Radio gespielt wird, um **eine Standardabweichung** erhöht und sich die anderen Prädiktoren nicht ändern, erhöhen sich die Albumverkäufe um **0,512 Standardabweichungen**.
In diesem Modell hat die „Radio-Häufigkeit“ also den stärksten eigenständigen Beitrag zur Varianzaufklärung der Albumverkäufe.

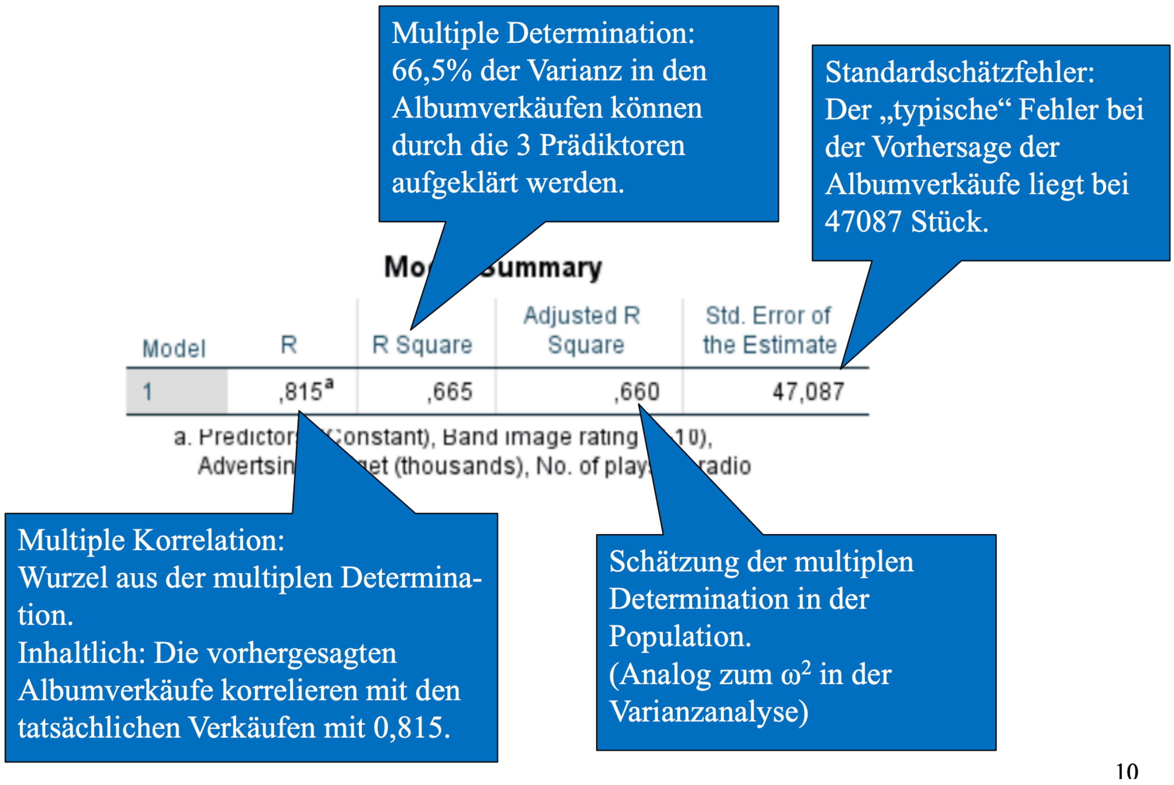
Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-26,613	17,350		-1,534	,127
	Advertsing budget (thousands)	,085	,007	,511	12,261	,000
	No. of plays on radio	3,367	,278	,512	12,123	,000
	Band image rating (0-10)	11,086	2,438	,192	4,548	,000

a. Dependent Variable: Album sales (thousands)

- um eine Vergleichbarkeit der Regressionsgewichte zu erreichen, müssen die Variablen zunächst eine gemeinsame Einheit erhalten - dazu werden sie (z-) standardisiert,
- das Resultat der Regression mit den standardisierten Variablen sind die (standardisierten) β-Gewichte!

<u>Gütemaße am Beispiel</u>



Signifikanztests

→ drei relevante Tests:
- Test der Nullhypothese, dass die Varianzaufklärung (r-Quadrat) des gesamten Regressionsmodells in der Population 0 beträgt,
- erfolgt über *F*-Statistik und analog zur einfaktoriellen ANOVA,

$$F = \frac{MS_{\hat{y}}}{MS_e} = \frac{\sum_i^N (\hat{y}_i - \bar{y}_i)^2 / k}{\sum_i^N (y_i - \hat{y}_i)^2 / (N-k-1)}$$

k → Anzahl der Prädiktoren

alternative Formel:

$$F = \frac{(N-k-1)R^2}{k(1-R^2)}$$

Freiheitsgrade:

$$df_{Zähler} = k \text{ und } df_{Nenner} = N - k - 1$$

→ Test der Nullhypothese, dass das Regressionsgewicht eines Prädiktors in der Population 0 beträgt,
- also H_0 - in der Population beträgt $b_i = 0$
- wir fragen also, ob wir annehmen können, dass ein Prädiktor in der Population überhaupt einen eigenständigen Beitrag zur Varianzaufklärung leistet,

→ Test erfolgt über eine t-Statistik

$$t = \frac{b_j}{SE_{b_j}}$$

- Freiheitsgrade: $df = N - 2$
- SE_{b_j} entspricht dem Standardfehler des Regressionsgewichts,
- allgemein: jeder *t*-Wert entspricht dem Bruch aus interessierender Kenngröße (hier: das Regressionsgewicht *b*) und dem entsprechenden Standardfehler,

Signifikanztests im Beispiel

→ Test der Varianzaufklärung gegen 0:

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	861377,418	3	287125,806	129,498	,000[b]
	Residual	434574,582	196	2217,217		
	Total	1295952,000	199			

a. Dependent Variable: Album sales (thousands)

b. Predictors: (Constant), Band image rating (0-10), Advertsing budget (thousands), No. of plays on radio

→ ANOVA-Tabelle wie sie aus der Varianzanalyse vertraut ist,

→ die Tabelle enthält die Quadratsummen der vorhergesagten Werte und der Fehler, die entsprechenden Freiheitsgrade und Varianzschätzungen und den daraus resultierenden *F*-Wert,

→ unser Ergebnis ist signifikant, wir schließen also, dass die Varianzaufklärung des gesamten Modells auch in der Population größer als 0 ist → das ist hier (und öfter mal) kein besonderer Gewinn!

- wenn die Prädiktoren einigermaßen sinnvoll und begründet ausgewählt sind, wäre 0 Varianzaufklärung schon sehr wenig...

→ zur Illustration: Test der Veränderung in der Varianzaufklärung

- wir sagen die Albumverkäufe zunächst anhand des Werbe-Etats vorher - Modell 1,
- wir fügen dann die Attraktivität des Band-Fotos hinzu - Modell 2,

Varianzaufklärung im Modell 1.

Dieser Zuwachs in der Varianzaufklärung ist signifikant. Wir können also annehmen, dass der Zuwachs auch in der Population größer als 0 ist.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,578[a]	,335	,331	65,991	,335	99,587	1	198	,000
2	,643[b]	,413	,407	62,129	,079	26,380	1	197	,000

a. Predictors: (Constant), dvertsing budget (thousands)

b. Predictors: (Constan dvertsing budget (thousands), Band image rating

Varianzaufklärung im Modell 2.

Durch die Hinzunahme des Prädiktors „Attraktivität d. Band-Fotos“ hat sich die Varianzaufklärung also um 7,9% erhöht..

→ Signifikanztests der Prädiktoren

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-26,613	17,350		-1,534	,127
	Advertsing budget (thousands)	,085	,007	,511	12,261	,000
	No. of plays on radio	3,367	,278	,512	12,123	,000
	Band image rating (0-10)	11,086	2,438	,192	4,548	,000

a. Dependent Variable: Album sales (thousands)

- zur Erinnerung:

$$t = \frac{b_j}{SE_{b_j}}$$

- also z.B.

$$t = \frac{0{,}085}{0{,}007} = 12{,}261$$

- in diesem Beispiel sind offensichtlich alle 3 Prädiktoren signifikant,
- wir können also annehmen, dass die Regressionsgewichte auch in der Population nicht 0 sind und dass (folglich) jeder der Prädiktoren auch in der Population einen eigenständigen Beitrag zur Varianzaufklärung leistet,

<u>Voraussetzungen</u>

→ lineare Beziehung zwischen dem Kriterium und *allen* Prädiktoren,

  - anderenfalls sind die Schätzungen von Regressionsgewichten und Varianzaufklärung falsch,
- → Homoskedastizität: bei allen Ausprägungen der Prädiktorvariablen sollten die Residuen die gleiche Varianz aufweisen - entspricht der Voraussetzung der Varianzhomogenität in der ANOVA,
- → die Residuen sollten normalverteilt sein
  - sofern die letzten beiden Voraussetzungen verletzt sind, sind die Resultate von Signifikanztests (und Konfidenzintervallen) potentiell falsch; die Schätzungen von Regressionsgewichten und Varianzaufklärung bleiben korrekt,
- → unabhängige Residuen: die Residuen von beliebig ausgewählten Datenpunkten sollten unkorreliert sein,
  - diese Voraussetzung ist in der Regel nur dann sinnvoll (und vor allem prüfbar), wenn die Datenpunkte in irgendeiner Form *sortiert* sind, z.B. weil die Datenpunkte zeitlich nacheinander erhoben wurden - dann lässt sich prüfen, ob die Residuen benachbarter Datenpunkte korrelieren,
- → wünschenswert: die Prädiktoren korrelieren nicht mit externen (also im Modell nicht berücksichtigten) Variablen - das wird in der Psychologie meist nicht realisierbar sein,
- → **die Konsequenz ist, dass die Regressionsgewichte der Prädiktoren nur innerhalb des betrachteten Modells gültig sind – würden andere (korrelierende) Prädiktoren in das Modell aufgenommen, würden sich die Gewichte ja ändern,**
- → Wichtig! - die in der Psychologie übliche Interpretation berücksichtigt diesen Umstand eher nicht - zumindest nicht sonderlich explizit,

Regressionsdiagnostik

- → Ausreißer können einen deutlichen Einfluss auf die Regressionsanalyse ausüben,
  - unsere Ergebnisse hängen dann also von einzelnen Fällen ab, was die Generalisierbarkeit (die Übertragbarkeit auf andere Stichproben) infrage stellt,
- → wie bemerken wir das?
  - Field stellt eine Vielzahl von Methoden vor…
- → zentrale Idee:
  - wir entfernen einen Datenpunkt und führen anschließend die Regressionsanalyse erneut durch,
  - damit erhalten wir neue Regressionsgewichte (und also eine neue Regressionsgleichung), neue vorhergesagte Werte und neue Residuen,
  - je stärker sich diese Ergebnisse ändern, desto größer ist der Einfluss des ausgeschlossenen Datenpunkts,
  - **Cook‘s distance** ist eine von mehreren Möglichkeit zusammenzufassen, wie stark sich die Residuen **aller Werte** durch den Ausschluss eines Datenpunktes ändern,
    - Werte größer 1 gelten per Konvention als problematisch,
  - grundsätzlich wird natürlich jeder Datenpunkt einmal ausgeschlossen, sonst können wir den Einfluss des Punktes nicht bestimmen,
    - klingt vermutlich aufwendig, lässt sich aber in SPSS mit wenigen Mausklicks umsetzen,
- → Multikollinearität verschlechtert die Schätzeigenschaften des Modells: die Varianzaufklärung erhöht sich kaum, die Regressionsgewichte werden aber ungenauer geschätzt,
- → wie bemerken wir das?
- → zentrale Idee:

- jeder Prädiktor wird auf alle anderen Prädiktoren regrediert,
- wir bestimmen 1 – r-Quadrat, also den Varianzanteil des Prädiktors, den er nicht mit den anderen Prädiktoren teilt und der daher überhaupt zur Vorhersage des Kriteriums *zur Verfügung steht*,
- diese Größe wird auch als **Toleranz** bezeichnet
    - Werte kleiner 0,1 sind sicher problematisch,
- hohe Multikollinearität eines Prädiktors erhöht den Standardfehler des Regressionsgewichts dieses Prädiktors bei der Vorhersage des Kriteriums,
- genauer: eine Toleranz von 0,1 bedeutet, dass sich die Varianz des Regressionsgewichts (das Quadrat des Standardfehlers, SE^2) um den Faktor 10 erhöht (im Vergleich zu einer einfachen Regression, in die nur dieser Prädiktor eingeht),
- dies wird als **Variance Inflation Factor** (**VIF**) bezeichnet

<u>Generalisierbarkeit</u>

→ die Ergebnisse der Regression beziehen sich zunächst auf die erhobene Stichprobe,
→ idR werden wir an Aussagen über die Population interessiert sein,
→ sind die Ergebnisse auch in der Population gültig, dann sollten sich in neuen Stichproben ähnliche Ergebnisse finden,
→ mit den Signifikanztests haben wir bislang ausschließlich geprüft, ob Varianzaufklärung und Regressionsgewichte in der Population *nicht 0* sind,
- Generalisierbarkeit würde aber einschließen, dass Varianzaufklärung und Regressionsgewichte in der Population *ähnlich ausfallen* wie in der Stichprobe,
- zur Erinnerung: wir haben zudem geprüft, ob die Generalisierbarkeit durch einzelne einflussreiche Werte gefährdet ist,

→ direkte Prüfung der Generalisierbarkeit:
Kreuzvalidierung:
- wir nutzen nur einen Teil der vorliegenden Stichprobe, um das Regressionsmodell zu schätzen,
- im verbliebenen Teil der Stichprobe wird geprüft, wie gut die Vorhersage des Regressionsmodells ausfällt,
- sind die Residuen im noch nicht genutzten Teil (also bei einer echten Vorhersage) ähnlich groß wie in der Schätzstichprobe?
- technisch gibt es eine Reihe unterschiedlicher Möglichkeiten, eine Kreuzvalidierung umzusetzen,

→ wichtiger Hintergrund:
- innerhalb einer Stichprobe wird sich die Vorhersage durch die Hinzunahme von Prädiktoren immer weiter verbessern (solange die zusätzlichen Prädiktoren nur irgendwie mit dem Kriterium oder den anderen Prädiktoren korrelieren,
- wenn genügend Prädiktoren zur Erprobung zur Verfügung stehen, wird in der Stichprobe also immer eine gute Vorhersage gelingen,
- dabei steigt allerdings die Wahrscheinlichkeit, dass dies nur an Spezifika der vorliegenden Stichprobe liegt,
- die Generalisierbarkeit ist daher zunehmend gefährdet,
- Problem des **overfitting**
- die Kreuzvalidierung kann daher eine sehr wichtige Maßnahme zur Prüfung von Regressionsergebnissen sein!

<u>Reihenfolge der Eingabe von Prädiktoren</u>

drei Methoden:

→ Forced entry/ Enter: alle betrachteten Prädiktoren gehen gemeinsam und gleichzeitig in das Modell ein,
  - das ist vermutlich der Regelfall - oder sollte es zumindest sein,
  - es gibt eine Reihe von theoretisch als relevant erachteten Prädiktoren, von denen man annimmt, dass sie gemeinsam in der Lage sind, das Kriterium gut vorherzusagen,

→ hierarchisch: Prädiktoren gehen nacheinander und in festgelegter Reihenfolge ein,
  - dies ist sinnvoll, wenn es bereits eine etablierte Theorie mit bestimmten Prädiktoren gibt,
  - diese Prädiktoren gehen zunächst in die Regression ein,
  - in einem nächsten Schritt wird dann geprüft, ob sich die Varianzaufklärung verbessert, wenn ein neuer Prädiktor/ neue Prädiktoren zusätzlich in das Modell aufgenommen wird/ werden,

→ schrittweise Regression: hier wird allein anhand von statistischen Kriterien über die Aufnahme von Prädiktoren entschieden,
  - dies lässt sich nochmal in unterschiedlichen Varianten umsetzen,
  - als Beispiel: forward method
  - zunächst wird diejenige Variable als Prädiktor aufgenommen, die die höchste einfache (bivariate) Korrelation mit dem Kriterium aufweist und die eine signifikante Varianzaufklärung erzielt - wenn dies keine Variable tut, ist das Verfahren (ohne Prädiktoren) beendet,
  - anschließend wird jeweils diejenige Variable gesucht, die die Varianzaufklärung am stärksten verbessert, und geprüft, ob der Zuwachs in der Varianzaufklärung signifikant ist,
  - wenn sich keine Variable mehr findet, die die Varianzaufklärung signifikant verbessert, ist das Verfahren beendet,
  - es sollte inzwischen klar sein, dass das Vorgehen *böse* ist, weil es die Gefahr des overfitting maximiert - in aller Regel sollte man das schlicht lassen,

<u>Beispiel</u>

→ die Datei **supermodel.sav** enthält Daten, die von 231 Models gesammelt wurden,

→ das Ziel besteht darin, das Tageseinkommen der Models vorherzusagen,

→ dazu stehen die Variablen Alter (in Jahren), Berufserfahrung (in Jahren) und beauty (ein Experten-Rating, das den Models einen Prozentrang im Vergleich zu anderen Models zuweist – der Wert 100 würde dem attraktivsten Model zugeordnet) zur Verfügung,
  - bestimmen Sie zunächst die Regressionsgleichung!
  - welches Einkommen erwarten Sie bei einem 17jährigen Model mit 3Jahren Berufserfahrung und dem beauty-Wert 85?
  - beurteilen Sie die Güte der Vorhersage anhand geeigneter Kennzahlen!
  - interpretieren Sie die Regressionsgewichte der einzelnen Prädiktoren!
  - von welchen Prädiktoren kann angenommen werden, dass sie auch in der Population einen eigenständigen Beitrag zur Varianzaufklärung leisten?
  - welche Variable ist in diesem Modell der stärkste Prädiktor des Einkommens?
  - zudem: entdecken Sie eine Auffälligkeit an den Regressionsgewichten? - falls ja: worauf könnte diese Auffälligkeit zurückgehen?
  - finden Sie Hinweise auf Multikollinearität oder Fälle mit kritischem Einfluss in den Daten?
  - sie sollten inzwischen Hinweise auf Multikollinearität gefunden haben...

- wie genau entsteht diese Multikollinearität hier?
- wie beeinflusst sie Ihre Interpretation der Ergebnisse?
- welche Änderung des Modells könnte Ergebnisse erbringen, die vertrauenswürdiger erscheinen?
- um dies einmal probiert zu haben, auch wenn es hier inhaltlich wenig sinnvoll ist: führen Sie eine hierarchische Regression durch, in der Sie im ersten Schritt das Alter in das Regressionsmodell aufnehmen und im zweiten Schritt die Variable beauty als Prädiktor hinzufügen,

moderation, mediation and multicategory predictors

! Moderation und Mediation sind zwei sehr unterschiedliche Konzepte,

! bei beiden Analysen geht es um ein besseres Verständnis des Zusammenhangs zwischen einer unabhängigen und einer abhängigen Variablen - hier soll überprüft werden, inwieweit eine dritte Variable in diesen Zusammenhang passt,

→ Moderation tritt auf, wenn sich die Beziehung zwischen zwei Variablen als Funktion einer dritten Variablen ändert, z.B. könnte die Beziehung zwischen dem Anschauen von Horrorfilmen und dem Gefühl von Angst vor dem Schlafengehen in Abhängigkeit davon zunehmen, wie lebhaft die Vorstellungskraft einer Person ist,

→ die Moderation wird anhand eines linearen Modells getestet, bei dem das Ergebnis (Angst vor dem Schlafengehen) von einem Prädiktor (wie viele Horrorfilme angesehen werden), dem Moderator (Imagination) und dem Zusammenspiel der Pädiktorvariablen vorhergesagt wird,

→ Prädiktoren sollten vor der Analyse zentriert werden,

→ die Interaktion zweier Variablen ist die Multiplikation ihrer Werte,

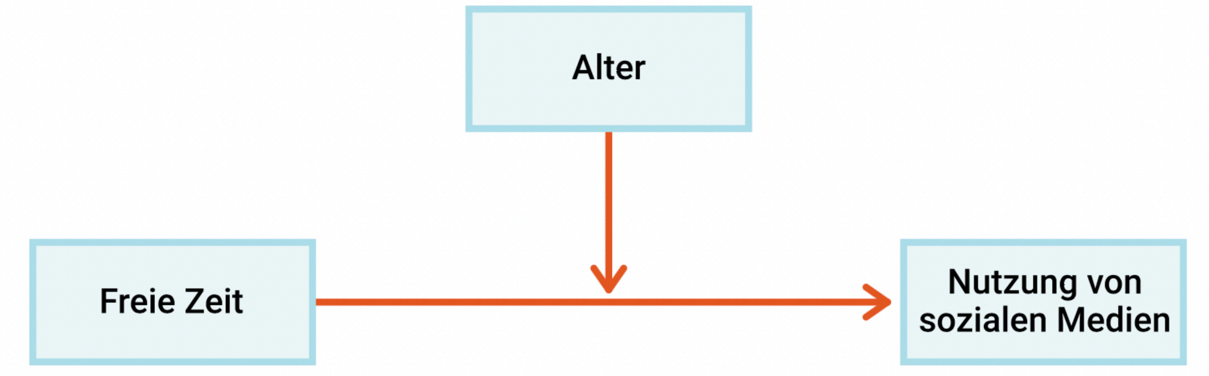
→ wenn die Interaktion signifikant ist, ist auch der Moderationseffekt signifikant,

→ wenn eine Moderation festgestellt wird, folgen Sie der Analyse mit einer einfachen Steigungsanalyse, die die Beziehung zwischen dem Prädiktor und dem Ergebnis bei niedrigem, mittlerem und hohem Niveau des Moderators untersucht,

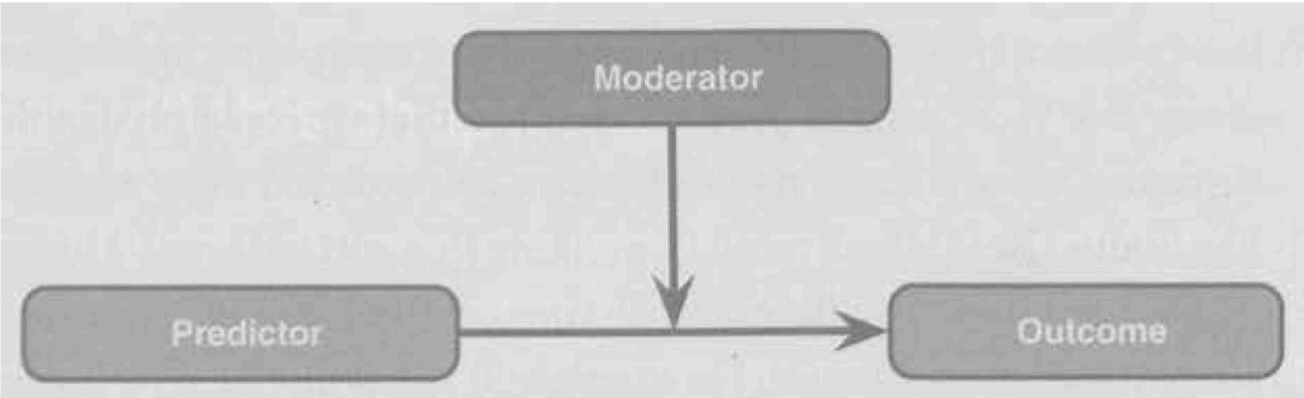
→ eine dritte Variable beeinflusst die **Stärke** zwischen aV und uV,

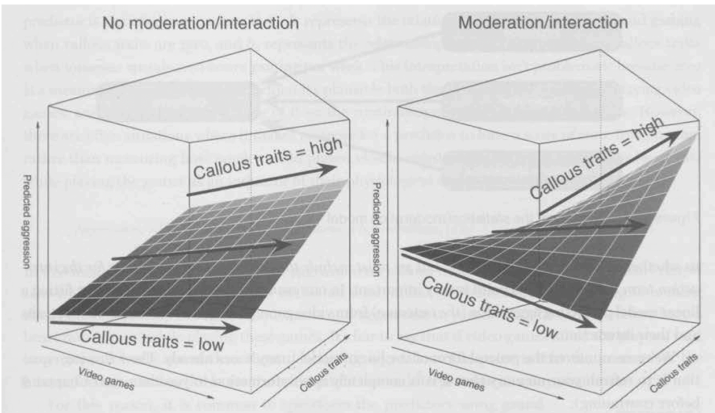
→ ein Moderator ist damit wie ein Wasserhahn, der entweder aufgedreht oder geschlossen wird, um die Stärke einer Beziehung zu regulieren,

→ Beispiel:



- das Alter einer Person moderiert die Beziehung zwischen der Menge an Freizeit und der Nutzung von sozialen Medien,
- eine Person mit viel Freizeit wird mit einer gewissen Wahrscheinlichkeit einen Teil davon mit der Nutzung von sozialen Medien verbringen, allerdings ist die Menge der Nutzung von sozialen Medien abhängig von dem Alter der Person,
- Alter wirkt wie ein Regler, der die Beziehung zwischen *Freizeit* und *Nutzung von sozialen Medien* beeinflusst,
- man kann bspw. herausfinden, dass je jünger eine Person ist, desto mehr Zeit wird sie mit der Nutzung von sozialen Medien verbringen, während ältere Personen ihre Freizeit anderweitig verbringen,
- Moderation beantwortet die Fragen: *wann?* und *unter welchen Bedingungen?*



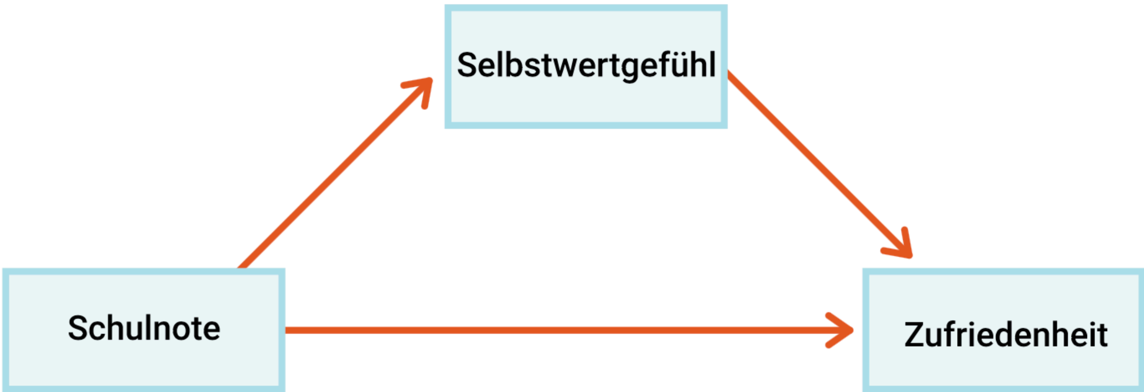


→ ein Mediator *mediiert* die Beziehung zwischen den uV und aV und erklärt den Grund für die Existenz dieser Beziehung,

→ damit ist die Mediationsanalyse auch gleichzeitig eine Analyse von kausalen Effekten,

→ als solche muss auch die zeitliche Beziehung zwischen den einzelnen Variablen logisch korrekt sein,

→ Beispiel:

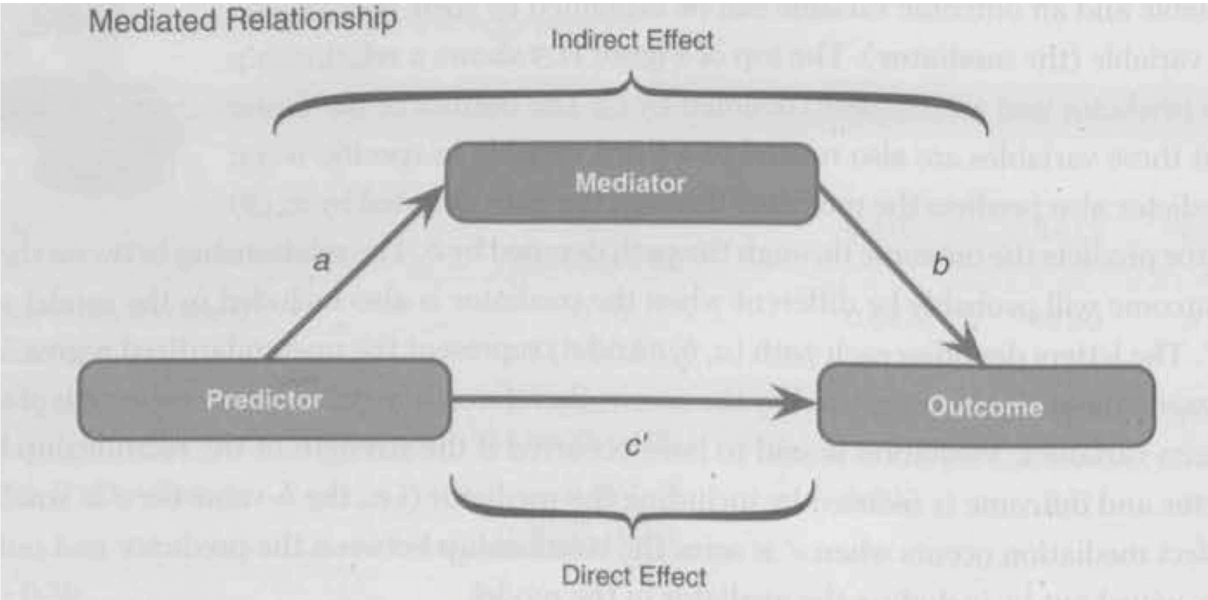


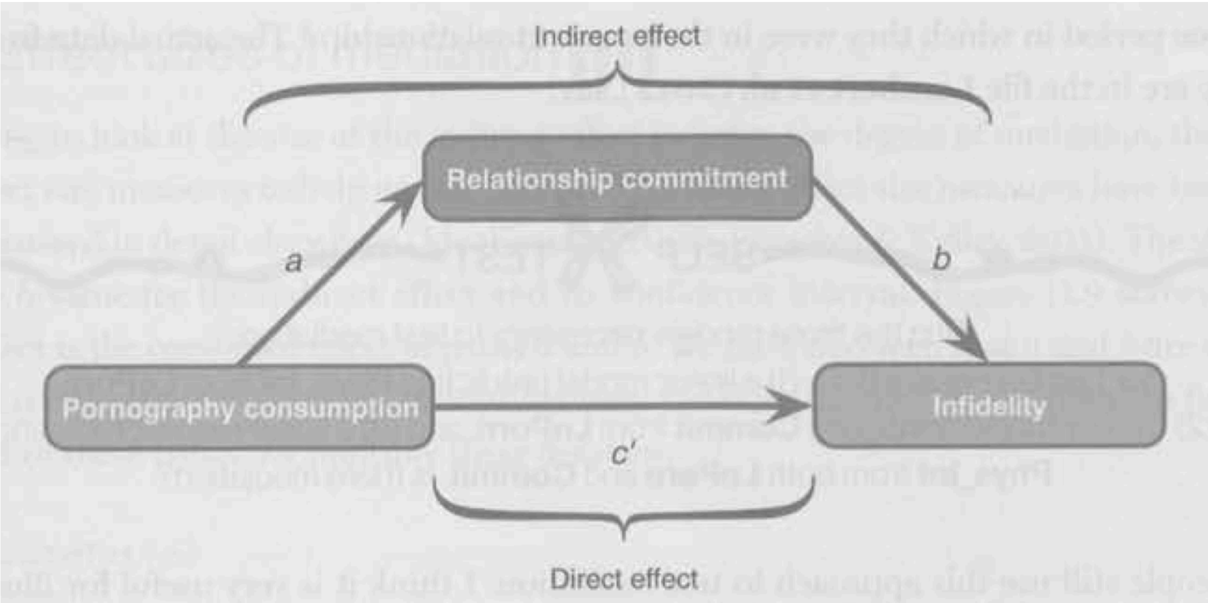
- das Selbstwertgefühl mediiert den Zusammenhang zwischen Schulnote und Zufriedenheit,
- im ersten Schritt könnte man vielleicht herausfinden, dass der Zusammenhang zwischen Schulnote und Zufriedenheit statistisch signifikant ist - das ist aber noch nicht die ganze Wahrheit, weil die Beziehung komplexer ist,
- gute Noten steigern das Selbstwertgefühl, welches sich wiederum positiv auf die Zufriedenheit der Schüler auswirkt,
- der Mediator ist ein erklärendes Bindeglied in dem Zusammenhang von uV und aV,
- Mediation erklärt, **warum** ein Effekt eintritt

→ Mediation liegt vor, wenn die Stärke der Beziehung zwischen einer Prädiktorvariablen und einer Ergebnisvariablen verringert wird, indem eine andere Variable als Prädiktor einbezogen wird,

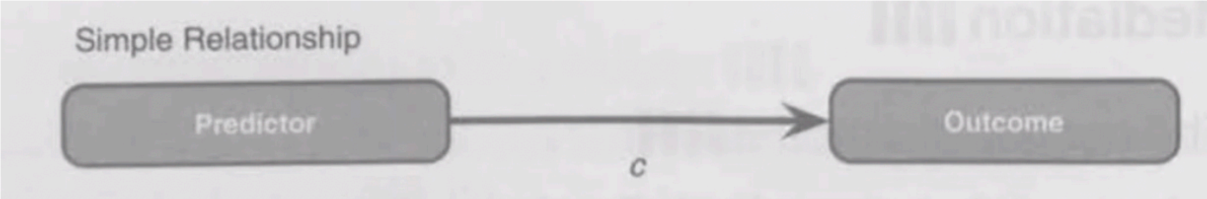
→ im Wesentlichen bedeutet das, dass die Beziehung zwischen zwei Variablen durch eine dritte erklärt wird, z.B. könnte die Beziehung zwischen dem Anschauen von Horrorfilmen und dem Gefühl der Angst vor dem Schlafengehen durch gruselige Bilder in Ihrem Kopf erklärt werden,

→ die Mediation wird getestet, indem die Größe des indirekten Effekts und sein Konfidenzintervall bewertet werden - wenn das Konfidenzintervall 0 enthält, neigen wir zu der Annahme, dass kein echter Vermittlungseffekt vorliegt - wenn das Konfidenzintervall keine 0 enthält, neigen wir zu dem Schluss, dass eine Mediation stattgefunden hat,





- Multiplikation von Regressionsgewichten - a/b → indirekter Effekt
- Beispiel: wenn Pornokonsum um eine Einheit steigt, sinkt das BeziehungsEngagement um 0,9 Einheiten und die UntreueWahrscheinlichkeit steigt wiederum um eine Einheit
- damit das irgendwie an den Start kommt, brauchen wir eine Beziehung zw einem Prädiktor und einem Kriterium, z.B. zw Pornokonsum und Untreue ↗
- unstrittig ist, dass die Menschen sich nach dem Ansehen eines Pornos nicht gleich auf den Weg machen, um untreu zu sein - irgendwas ist da noch zwischen - und das könnte eine Verminderung des BeziehungsEngagements sein,
- wie untersuchen wir?
  - lineare Regression zw Untreue und Pornokonsum
  - wir erhalten dabei aber noch keinen der drei Kennwerte - weder a noch b noch c‘ - sondern c



- neue Regression von BeziehungsEngagement auf Untreue und von BeziehungsEngagement auf Pornokonsum,
- **man braucht jeweils eine Regression pro Variable, auf die Pfeile zeigen,**
- **und in diesen Regressionen sind all die Variablen Prädiktoren, von denen die Pfeile ausgehen,**

- c‘ sollte kleiner als c werden,
- H_0 - $a * b = 0$
- Sobel-Test: Methode zum Testen der Signifikanz eines Mediationseffekts, interessierende Kenngröße - hier $\frac{a * b}{\text{Standardfehler von } a * b}$
- Ergebnis ist z als Prüfgröße,
- die Stickprobenkennwerteverteilung von a * b ist eine Normalverteilung → Problem,
- das Produkt zweier normalverteilter Variablen ist grundsätzlich nicht wieder normalverteilt!
- der Test beruht auf einer fehlerhaften Verteilungsrate,
- potentiell beruht jeder Test auf fehlerhaften Annahmen,

→ das Bootstrapping-Verfahren ist eine Methode des Resampling - dabei werden wiederholt Statistiken auf der Grundlage lediglich einer Stichprobe berechnet,

- das Konfidenzintervall für a * b ist dann signifikant, wenn es die 0 nicht mit einschließt,
- **Konfidenzintervalle sind keine Wahrscheinlichkeiten!**

<u>praktische Interpretationen</u>

```
**************************************************************
Model  = 4
    Y  = Phys_Inf
    X  = LnPorn
    M  = Commit
Sample size
       239
**************************************************************
Outcome: Commit

Model Summary
       R      R-sq     MSE       F       df1       df2        p
   .1418    .0201   .5354  4.8633    1.0000  237.0000    .0284

Model
              coeff       se         t         p      LLCI      ULCI
constant     4.2027    .0545   77.1777     .0000    4.0954    4.3100
LnPorn       -.4697    .2130   -2.2053     .0284    -.8892    -.0501
```

Output 11.4

! in welchem Zusammenhang stehen hier Standardfehler, t-Wert und p-Wert?
  ! zu sehen ist Output von Regressionsmodellberechnung
  ! coeff: Regressionskoeffizienten - b_0/b_1
  ! die angegebenen Werte 4.2027/.4697 sind deren Punktschätzungen
  ! se besagt die Präzision der Punktschätzungen
  ! Standardfehler - wie genau ist dieses Regressionsgewicht geschätzt worden?
  ! wenn man davon ausgehen würde, aus der gleichen Population mit der gleichen Stichprobengröße ganz viele von den Regressionskoeffizienten zu berechnen, wie

sehr würden wir dann erwarten, dass diese Regressionskoeffizienten schwanken - dafür ist der Standardfehler ein Ausdruck,

! Frage der Tests: ist der Regressionskoeffizient signifikant verschieden von 0? - EinStichproben-t-Test →
Mittelwert minus Mittelwert der H_0 geteilt durch Standardfehler des Mittelwerts

$$t = \frac{\bar{x} - \mu_0}{\hat{\sigma}_{\bar{x}}}$$

! p-Wert → Regressionskoeffizent/coeff/b_i geteilt durch Standardfehler

* Regression des Mediators *BeziehungsEngagement* auf den Prädiktor *Pornokonsum*,
* Pornokonsum korreliert negativ mit BeziehungsEngagement
* se: Streuung der Regressionsgewichte, Standardfehler
* p: Signifikanz,
* LLCI/ULCI: Konfidenzintervalle,
* LnPorn: logarithmierter Pornokonsum
* sinngemäß: wenn Pornokonsum um eine Einheit steigt, dann sinkt das BeziehungsEngagement um ,4697 Einheiten,
* -,4697 ÷ ,2130 = t-Wert,
* es gibt zumindest einen Zusammenhang zw Prädiktor und Mediator,

Outcome: Phys_Inf

Model Summary

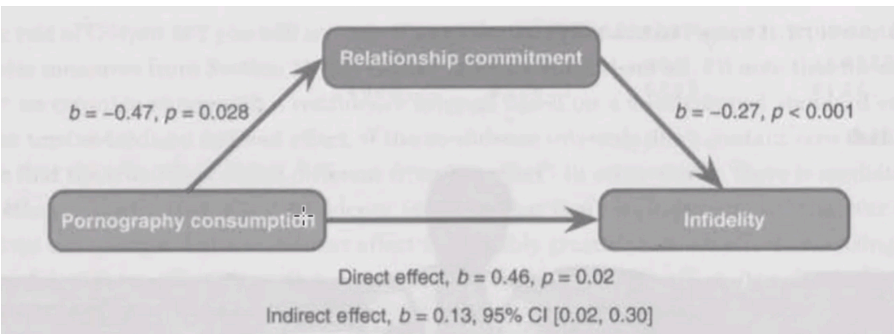
R	R-sq	MSE	F	df1	df2	p
.3383	.1144	.4379	15.2453	2.0000	236.0000	.0000

Model

	coeff	se	t	p	LLCI	ULCI
constant	1.3704	.2518	5.4433	.0000	.8744	1.8663
Commit	-.2710	.0587	-4.6128	.0000	-.3867	-.1552
LnPorn	.4573	.1946	2.3505	.0196	.0740	.8407

Output 11.5

* die einfache Regression von Pornokonsum auf Untreue ist signifikant,
* es müssen nicht alle direkten Bezüge signifikant sein,



* direkter Effekt + indirekter Effekt = Gesamteffekt aus einfacher Regression

Beispiel zum Thema:

- jeder mag von Zeit zu Zeit guten Klatsch, aber anscheinend hat er eine evolutionäre Funktion,
- eine Denkweise ist, dass Klatsch dazu verwendet wird, sexuelle Konkurrenten abzuwerten, insbesondere indem ihr Aussehen und ihr Sexualverhalten in Frage gestellt werden,
- wenn man z.B. einen Typen im Auge hat, er aber Jane im Auge hat, dann ist es eine gute Strategie, Klatsch zu verbreiten, dass Jane eine massive Eiterbeule auf ihrem Bauch hat und dass sie einen stinkenden Landstreicher namens XXX geküsst hat,
- anscheinend bewerten Männer Klatsch über Frauen als weniger attraktiv, und sie werden stärker vom Klatsch beeinflusst, wenn er von einer Frau mit einem hohen Partnerwert (d.h. attraktiv und sexuell begehrenswert) stammt,
- Karlijn Massar und ihre Kollegen stellten die Hypothese auf, dass, wenn diese Theorie zutrifft,
    a) jüngere Frauen mehr klatschen, weil es in jüngeren Jahren mehr Partnerkonkurrenz gibt und
    b) diese Beziehung durch den Partnerwert der Person vermittelt wird, weil für diejenigen mit hohem Partnerwert Tratsch zum Zwecke des sexuellen Wettbewerbs effektiver ist,
- 83 Frauen im Alter von 20 bis 50 füllten Fragebogenmessungen zu ihrer Neigung zum Klatsch (→ Gossip) und ihrer sexuellen Erwünschtheit (→Mate-Value) aus,

MediationsTest anhand von drei Regressionsmodellen:

1. eine Regression, die das Ergebnis (Gossip) aus der Prädiktorvariablen (Alter) vorhersagt,
2. eine Regression, die den Mediator (Mate-Value) aus der Prädiktorvariablen (Alter) vorhersagt,
3. eine Regression, die das Ergebnis (Gossip) sowohl aus der Prädiktorvariablen (Alter) als auch aus dem Mediator (Mate-Value) vorhersagt,

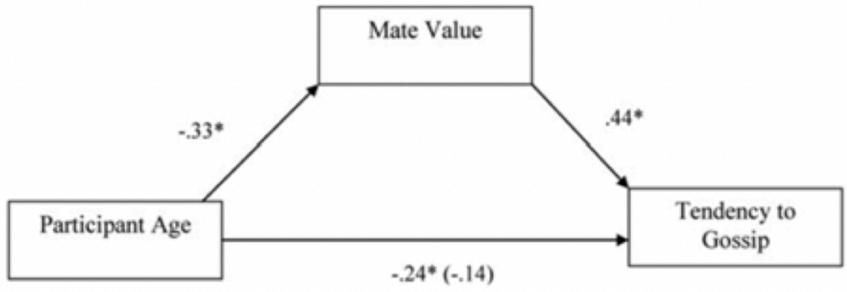
diese Modelle testen die vier Bedingungen der Mediation:

a) die Prädiktorvariable (Alter) muss die Ergebnisvariable (Gossip) in Modell 1 signifikant vorhersagen,
b) die Prädiktorvariable (Alter) muss den Mediator (Mate-Value) in Modell 2 signifikant vorhersagen,
c) der Mediator (Mate-Value) muss die Ergebnisvariable (Gossip) in Modell 3 signifikant vorhersagen und
d) die Prädiktorvariable (Alter) muss die Ergebnisvariable (Gossip) in Modell 3 weniger stark vorhersagen als in Modell 1,

! man kann sehen, dass die dritte Bedingung der Mediation erfüllt ist, da der Mate-Wert die Neigung zum Klatsch signifikant vorhersagte, während das Alter der Teilnehmer kontrolliert wurde, $t(78) = 3{,}5$, $p < 0{,}01$,

! schließlich wurde auch die vierte Bedingung der Mediation erfüllt, indem der standardisierte Regressionskoeffizient zwischen dem Alter der Teilnehmer und der Klatschtendenz unter Kontrolle des Partnerwerts deutlich abnahm, t(78) = □ 1,28, ns,

! daher können wir den Schluss ziehen, dass die Vorhersage des Autors unterstützt wird und der Zusammenhang zwischen dem Alter der Teilnehmer und der Neigung zum Klatschen durch den Partnerwert vermittelt wird,




```
**************************************************************************
Model = 4
    Y = Gossip
    X = Age
    M = Mate_Val

Sample size
         81

**************************************************************************
Outcome: Mate_Val

Model Summary
          R       R-sq          F        df1        df2          p
      .3815      .1455    13.4522     1.0000    79.0000      .0004

Model
              coeff         se          t          p
constant     3.7981      .2366    16.0558      .0000
Age          -.0266      .0073    -3.6677      .0004
```

- ! man kann sehen, dass das Alter den Partnerwert signifikant vorhersagt, b = -0.03, t = -3.67, p = .000,
- ! der R2-Wert sagt uns, dass das Alter 14,6% der Varianz des Partnerwerts erklärt,
- ! die Tatsache, dass b negativ ist, sagt uns, dass die Beziehung auch negativ ist: mit zunehmendem Alter nimmt der Partnerwert ab (und umgekehrt),

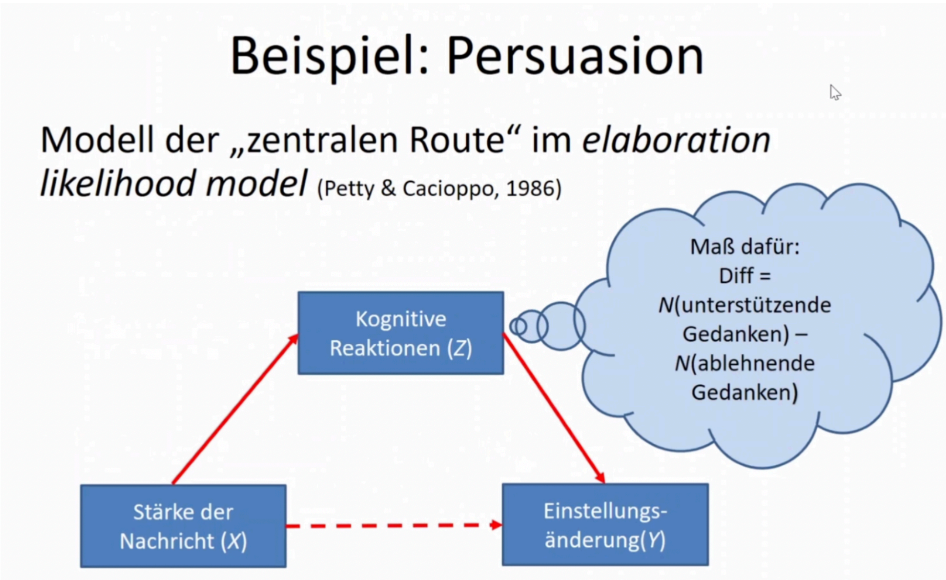
```
Outcome: Gossip

Model Summary
          R       R-sq          F        df1        df2          p
      .4614      .2129    10.5468     2.0000    78.0000      .0001

Model
              coeff         se          t          p
constant     1.1963      .5495     2.1771      .0325
Mate_Val      .4546      .1266     3.5921      .0006
Age          -.0113      .0088    -1.2753      .2060
```

- ! hier die Ergebnisse der Regression der Klatschtendenz, die sowohl aus dem Alter als auch aus dem Partnerwert vorhergesagt wurde,

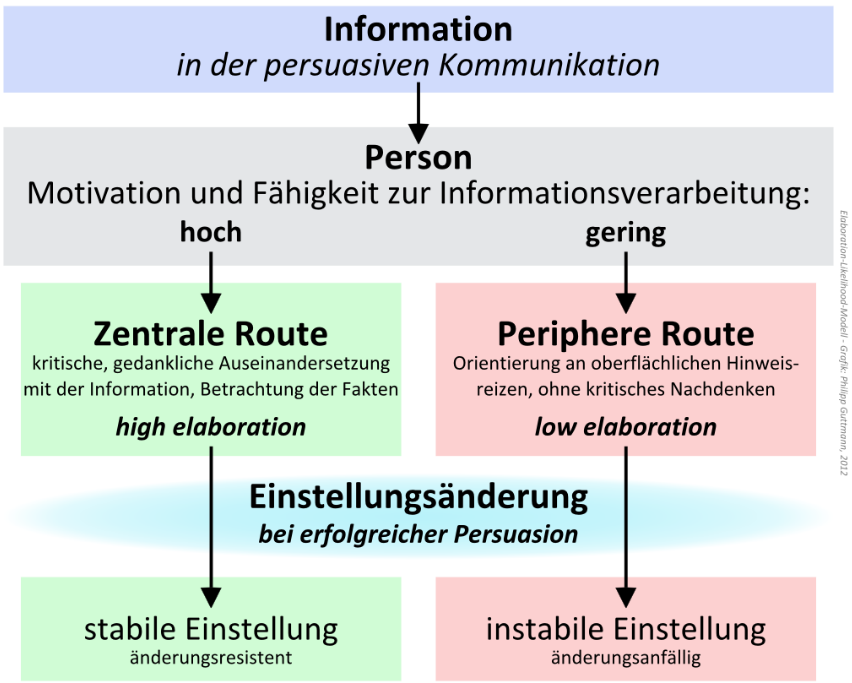
- ! man sieht, dass, während das Alter die Klatschneigung mit dem Partnerwert im Modell nicht signifikant vorhersagt, $b = -0,01$, $t = -1,28$, $p = .21$, der Partnerwert die Klatschneigung signifikant vorhersagt, $b = 0,45$, $t = 3,59$, $p < 0,01$
- ! der R2-Wert sagt uns, dass das Modell 21,3% der Varianz der Klatschneigung erklärt,
- ! das negative b für das Alter sagt uns, dass die Neigung zum Klatsch mit zunehmendem Alter abnimmt (und umgekehrt),
- ! das positive b für den Partnerwert zeigt an, dass mit zunehmendem Partnerwert auch die Neigung zum Klatsch zunimmt,
- ! diese Beziehungen gehen in die vorhergesagte Richtung,

<u>Exkurs: was bedeutet ein signifikanter indirekter Effekt?</u>

→ äquivalente Modelle,
→ empirische Unterscheidbarkeit verschiedener Modelle an einem Beispiel,
→ wie bewertet man die Leistungsfähigkeit statistischer Verfahren?



- *elaboration likelihood model* - wie kommt es zu Einstellungsänderungen?

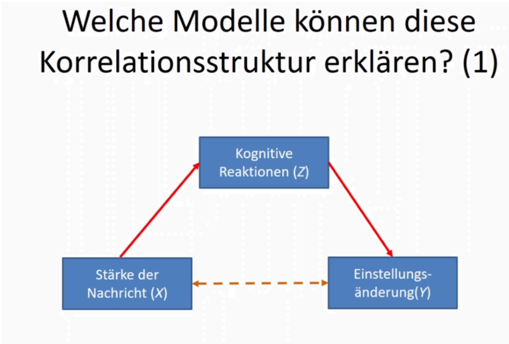


Korrelationsmatrix:

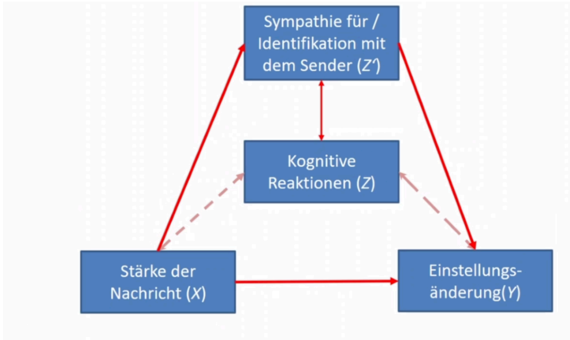
	Nachrichten-stärke (X)	Kognitive Bewertung (Z)	Einstellungs-änderung (Y)
Nachrichten-stärke (X)			
Kognitive Bewertung (Z)	$r(X, Z) > 0$		
Einstellungs-änderung (Y)	$r(X, Y)$	$r(Y, Z) > 0$	

- Bei vollständiger Mediation (und im besten Fall): $r(X, Y) = r(X, Z) \bullet r(Y, Z)$
- In jedem Fall: $r(X, Y) > 0$

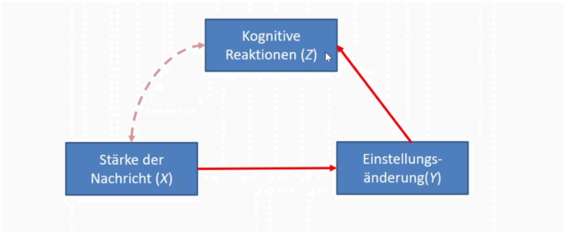
- Das Modell ist genau identifiziert ($df = 0$)!



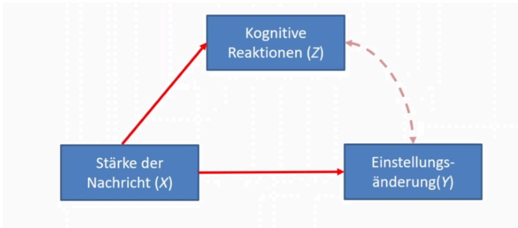
Variante 1: Mediation



Variante 2: „Schein-Mediation“



Variante 3: Korrelat der AV (Z ist ein alternatives Maß für Y)



Variante 4: Manipulation check (Z ist ein Maß für X)

→ offensichtlich gibt es bei den Modellen 2 - 4 keinen bzw. nicht den vermuteten indirekten Effekt,

→ kann der Signifikanztest des indirekten Effektes das Modell 1 von den Modellen 2 - 4 unterscheiden?
- findet der Test bei Gültigkeit des Modell 1 idR signifikante Ergebnisse und bei Gültigkeit eines der anderen Modelle idR nicht signifikante Ergebnisse?

→ wie finden wir das raus?

→ Monte-Carlo-Simulation
- es werden Daten simuliert, die bei Gültigkeit der verschiedenen Modelle entstehen würden,
- wir betrachten die Effekte als gegeben, die lt. jeweiligem Modell existieren,
- die übrigen Korrelationen werden geschätzt,

Modell	Gegebene Effekte	Zu schätzende Korrelation
1. Mediation	r(X, Z); r(Z, Y)	r(X, Y)
2. Schein-Mediation	r(X, Z'); r(Z', Y); r(Z', Z) = .7	r(X, Y)
3. Korrelat der AV	r(X, Y); r(Y, Z)	r(X, Z)
4. Manipulation Check	r(X, Y); r(X, Z)	r(Y, Z)

→ unter jedem der Modelle werden *k* Stichproben simuliert,

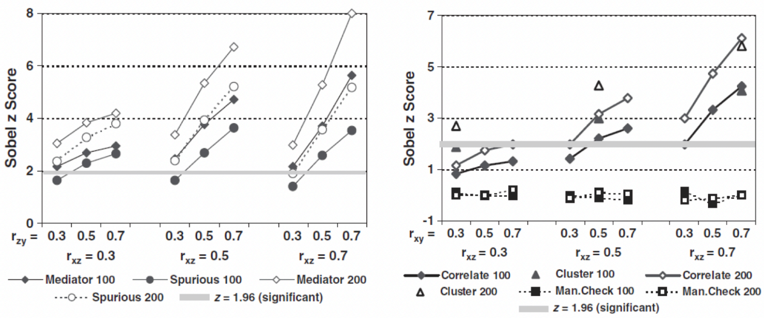
→ in jeder Stichprobe wird die zu schätzende Korrelation bestimmt,

→ in jeder Stichprobe wird der indirekte Effekt (von X über Z auf Y) auf Signifikanz geprüft,
- wir ermitteln, in welchem Anteil der Stichproben der indirekte Effekt signifikant ist,

Ergebnisse - Fiedler et al. (2011)

Table 1
Mean and standard deviation (SD) of mediation analysis results (Sobel z) as a function of different causal models and two sample sizes (across 100 simulation trials).

$n=100$						$n=200$						
r_{xy}	r_{xz}	r_{zy}	$r_{partial}$	Sobel z Mean	Sobel z SD	r_{xy}	r_{xz}	r_{zy}	$r_{partial}$	Sobel z Mean	Sobel z SD	
0.08	0.30	0.30	−0.01	2.15	0.12	*0.09*	0.30	0.30	0.00	3.04	0.11	Mediator
0.16	0.30	0.50	0.01	2.70	0.04	*0.15*	0.30	0.50	0.00	3.84	0.04	
0.21	0.30	0.70	−0.01	2.95	0.01	*0.22*	0.30	0.70	0.01	4.19	0.01	
0.14	0.50	0.30	−0.01	2.45	0.41	*0.16*	0.50	0.30	0.01	3.38	0.32	
0.23	0.50	0.50	−0.02	3.76	0.20	*0.24*	0.50	0.50	−0.01	5.33	0.20	
0.36	0.50	0.70	0.01	4.71	0.09	*0.35*	0.50	0.70	0.00	6.71	0.09	
0.21	0.70	0.30	0.00	2.15	0.61	*0.22*	0.70	0.30	0.01	2.99	0.60	
0.35	0.70	0.50	0.00	3.72	0.53	*0.35*	0.70	0.50	0.01	5.27	0.55	
0.49	0.70	0.70	−0.01	5.64	0.35	*0.49*	0.70	0.70	0.00	7.99	0.36	
0.07	0.30	*0.21*	0.01	1.64	0.39	*0.07*	0.30	*0.21*	0.01	2.37	0.46	Spurious mediator
0.11	0.30	*0.35*	0.01	2.30	0.29	*0.11*	0.30	*0.35*	0.00	3.29	0.27	
0.16	0.30	*0.49*	0.01	2.67	0.10	*0.16*	0.30	*0.49*	0.01	3.80	0.10	
0.12	0.50	*0.21*	0.02	1.65	0.62	*0.11*	0.50	*0.21*	0.00	2.40	0.60	
0.18	0.50	*0.34*	0.01	2.70	0.51	*0.17*	0.50	*0.35*	0.00	3.95	0.57	
0.24	0.50	*0.49*	0.00	3.63	0.39	*0.26*	0.50	*0.49*	0.01	5.21	0.41	
0.16	0.70	*0.21*	0.01	1.42	0.72	*0.15*	0.70	*0.20*	0.01	1.91	0.80	
0.25	0.70	*0.36*	0.00	2.60	0.81	*0.26*	0.70	*0.36*	0.01	3.56	0.80	
0.34	0.70	*0.48*	0.01	3.54	0.79	*0.35*	0.70	*0.49*	0.01	5.17	0.65	
0.30	*0.10*	0.30	0.29	0.82	0.75	0.30	*0.09*	0.30	0.29	1.16	0.66	Correlate
0.30	*0.15*	0.50	0.26	1.44	0.84	0.30	*0.15*	0.50	0.27	1.98	0.75	
0.30	*0.20*	0.70	0.23	2.00	0.67	0.30	*0.21*	0.70	0.22	3.00	0.68	
0.50	*0.15*	0.30	0.48	1.15	0.45	0.50	*0.15*	0.30	0.48	1.77	0.43	
0.50	*0.25*	0.50	0.45	2.22	0.49	0.50	*0.25*	0.50	0.45	3.15	0.48	
0.50	*0.35*	0.70	0.38	3.33	0.58	0.50	*0.35*	0.70	0.38	4.75	0.54	
0.70	*0.21*	0.30	0.68	1.33	0.20	0.70	*0.20*	0.30	0.69	1.98	0.28	
0.70	*0.36*	0.50	0.65	2.61	0.21	0.70	*0.35*	0.50	0.65	3.79	0.15	



! ein signifikantes Ergebnis für den indirekten Effekt bedeutet, dass es die vermutete Mediation geben **könnte**,
! leider bedeutet ein nicht-signifikantes Ergebnis nicht unbedingt, dass es dieses Mediation nicht geben kann → Powerproblem,
! das signifikante Ergebnis allein ist ein schwaches Argument für die Mediation,
! es fehlt eine Betrachtung der alternativen Erklärungen für das Befundmuster,
! diese müssten mit zusätzlichen theoretischen Begründungen oder anderen empirischen Befunden ausgeschlossen werden,

Tutorium

1. was genau sind VIF und Toleranz? …bzw. cook's cistance?
   → VIF klärt, wie stark sich der Standardfehler eines Regressionskoeffizienten aufgrund vorliegender Multikollinearität vergrößert

$$\widehat{\operatorname{var}}(\hat{\beta}_j) = \frac{s^2}{(n-1)\widehat{\operatorname{var}}(X_j)} \cdot \boxed{\frac{1}{1-R_j^2}}$$

- Hilfsmittel, um Multikollinearitäten zwischen den unabhängigen Variablen eines Modells zu entdecken,

- wie stark nimmt die Varianz eines geschätzten Regressionskoeffizienten zu, wenn eine Korrelation zwischen den Prädiktoren (uVs) besteht?

→ cook's distance - Datenpunkte mit großen Residuen (Ausreißern) könnten das Ergebnis und die Präzision einer Regression beeinflussen - Messung des Effekts der Auslassung einer gegebenen Beobachtung - Datenpunkte mit einem großen Cook-Abstand sollte man bei der Datenanalyse näher betrachten

- zeigt, wie stark sich die Residuen aller Werte durch den Ausschluss eines Datenpunktes ändern,
- Werte > 1 → problematisch

2. was ist bootstrapping?

→ wiederholte Berechnung von Statistiken auf der Grundlage lediglich einer Stichprobe,

→ Einsatz immer dann, wenn man eine statistische Größe hat, von dem man gern die Stichprobenkennwerteverteilung hätte, aber nicht weiß, wie sie ausschaut,

→ mit Standardfehler, Varianz und Mittelwert wird die entsprechende Stichprobenkennwerteverteilung (SKV) normalerweise angemessen beschrieben,

→ wenn man einen Kennwert hat, bei dem unbekannt ist, wie er verteilt ist, nimmt man die Formel für diesen Kennwert, bastelt/simuliert sich einen riesigen Datensatz, aus dem Stichprobenziehungen durchgeführt werden - diese Werte ergeben dann die simulierte SKV

3. stellen Sie sich vor, wir wollten die Beziehung zwischen der Anzahl der Stunden, die Sie pro Woche Gitarre üben, und dem Können untersuchen - wenn wir Grund zu der Annahme hätten, dass die Stärke oder Richtung der Beziehung zwischen diesen Variablen durch den Grad der Freude beeinflusst wird, welche Art von Analyse sollten wir dann mit diesen Daten durchführen?

→ Moderationsanalyse

! **der kombinierte Effekt zweier Variablen auf eine andere wird konzeptionell als Moderation und statistisch als Interaktionseffekt bezeichnet !**

4. die Mittelwertzentrierung für eine gegebene Variable wird erreicht durch:

→ jede Punktzahl - Mittelwert aller Punktzahlen (für diese Variable)

5. warum ist es eine gute Idee, die Variablen bei der Durchführung einer Moderationsanalyse zu zentrieren?

→ weil es die Multikollinearität zwischen Haupteffekten und dem Interaktionseffekt reduziert

6. stellen Sie sich vor, wir wollten untersuchen, ob die Freude am Gitarrenspiel das Verhältnis zwischen Übungszeit und Können beeinflusst - wie würden wir wissen, ob wir einen signifikanten Moderationseffekt hätten?

→ wenn die Wechselwirkung von Zeit zum Üben und Vergnügen ein signifikanter Prädiktor für das Können war,

DummyCodierung

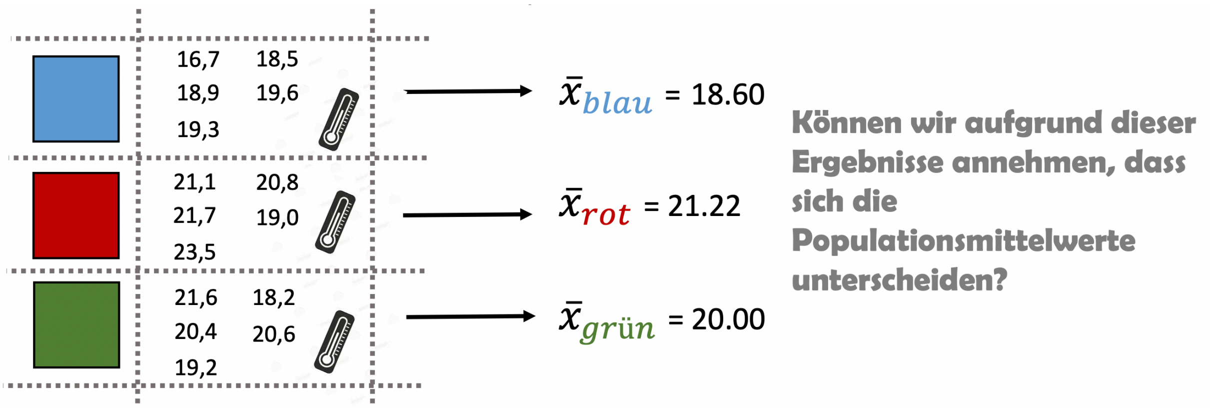
- → eine Umfrage zur Religiösität führt zu vielen Kategorien: Muslime, Juden, Hindus, Katholiken, etc. → keine Unterscheidung möglich,
- → Verwendung einer einzigen Variablen, die mit 0 und 1 codiert ist,
- → bei der Verwendung dieser Variablen in einem linearen Modell müssten wir DummyVariablen erstellen, die Personengruppen darstellen
- → man kreiert versch. Variablen, eine weniger als man Anzahl von Gruppen hat,
- → 8grundlegende Schritte:
  1. Gruppenanzahl zählen, eins abziehen,
  2. erstellen der neuen Variablen → DummyVariablen,
  3. eine der Gruppen als Basis, mit der alle anderen Gruppen verglichen werden - idR wählt man die Gruppe, die als Kontrolle angesehen werden könnte oder, wenn man keine bestimmte Hypothese hat, die Gruppe mit den meisten Menschen,
  4. nach Wahl der Basisgruppe, Zuweisung der 0 für alle DummyVariablen,
  5. erste DummyVariable der ersten Gruppe, die mit Basisgruppe verglichen werden soll, erhält den Wert 1,
  6. zweite DummyVariable der zweiten Gruppe, die mit Basisgruppe verglichen werden soll, erhält den Wert 1 - restliche Variablen erhalten jeweils Wert 0,
  7. diesen Vorgang wiederholen, bis man keine DummyVariablen mehr hat,
  8. Platzierung der DummyVariablen im linearen Modell im selben Block,

- → eine Variable mit den Ausprägungen 1 und 0 → ja-nein-Variable, die als Indikator für das Vorhandensein einer Ausprägung einer mehrstufigen Variablen dient,
- → diese der Dummy-Variable zugrunde liegende Variable kann ein beliebiges Skalenniveau haben,
- → Beispiel:
  - bei statistischen Auswertungen kann es hilfreich sein zu wissen, ob eine Untersuchungseinheit eine bestimmte Ausprägung einer kategorialen Variablen aufweist oder nicht - zu diesem Zweck bildet man eine DummyVariable mit den Ausprägungen 1 und 0:
    - 1 → Ausprägung liegt vor
    - 0 → Ausprägung liegt nicht vor
  - die Überführung einer kategorialen Variable in eine künstliche numerische Variable nennt man Kodierung,
  - bei einer Wahlumfrage gibt eine kategoriale Variable an, welche Partei der Befragte wählen würde - um den Anteil der CDU-Wähler zu ermitteln, benutzt man eine Dummy-Variable mit den Ausprägungen *1 → CDU-Wähler* und *0 → kein CDU-Wähler*,
  - **bei intervallskalierten Variablen werden Dummys oft benutzt, um anzuzeigen, ob ein Wert dichotom unter oder über einer bestimmten Grenze liegt,**
    - die DummyVariable bekommt den Wert 1, wenn die befragte Person jünger als 50 Jahre ist, und ansonsten den Wert 0,

Varianzanalyse - ANOVA

- → z.B. haben Farben Einfluss auf das Temperaturempfinden? ☆ ☆ ☆
- → Experiment
  - Ablauf: konstant temperierter Raum wird blau, rot oder grün ausgeleuchtet - jeder dieser Bedingungen werden 5 Probanden zufällig zugewiesen - Temperaturschätzung nach 10min Aufenthalt,

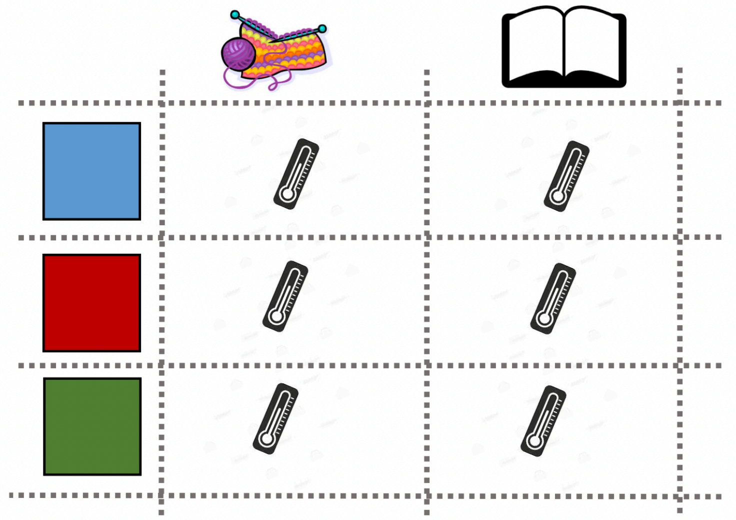
- uV: Farbe - in der Varianzanalyse auch **Faktor** - Ausprägungen (blau, rot, grün): Faktorstufen
- aV: Temperaturschätzung in °C - metrisch - mind. intervallskaliert



↘

einfaktoriell

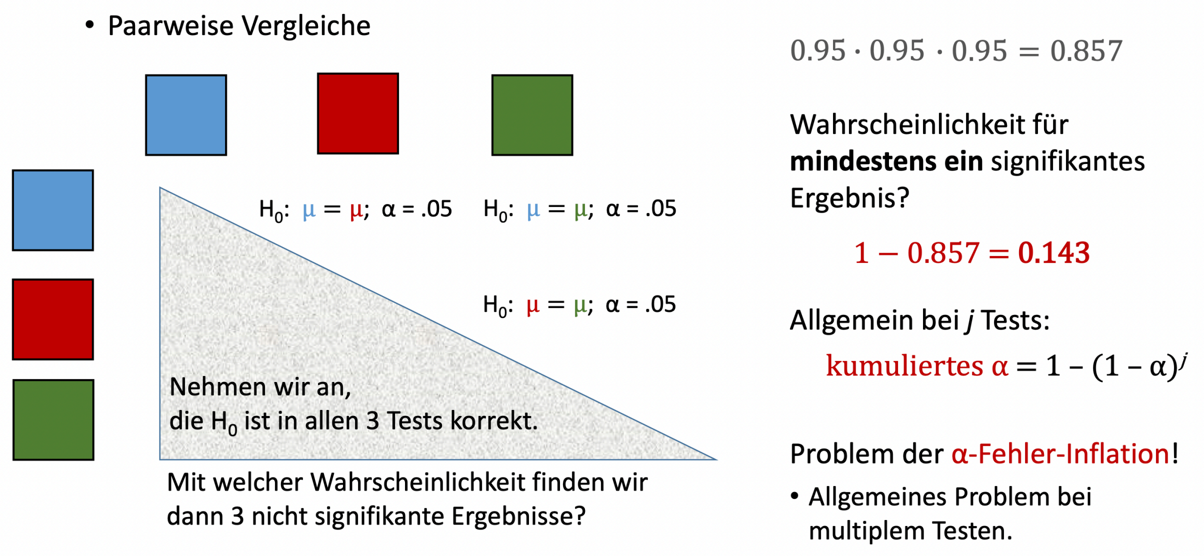
→ Ziel der Varianzanalyse:
- Vergleich der Populationsmittelwerte aus 3 oder mehr Gruppen/Bedingungen
- Erweiterung des t-Tests



zweifaktoriell/mehrfaktoriell

→ unser Ziel...
- zunächst: Verständnis der zentralen Idee der Varianzanalyse
- welche Konzepte sind neu (z.B. im Vergleich zum *t*-Test)?
- eher nebenbei: rechnerische Durchführung
- **Leitfragen:**
  - ? warum wird die Varianzanalyse benötigt?
  - ? wie kann eine Analyse der Varianz etwas über Mittelwertsunterschiede aussagen?

→ warum nicht mehrere t-Tests?



→ α-Fehler-Inflation

→ Lösungen

→ Korrektur von α für alle Einzeltests

- Adjustierung des α-Niveaus bei den Einzeltests, so dass das kumulierte α in der Gesamtstudie 5% beträgt
- dies leistet näherungsweise die **Bonferoni-Korrektur**: α‘ = kumuliertes α/j
- im Beispiel: α‘=0,05/3=0,0167
- **Problem: Verlust an Teststärke**

→ gleichzeitiger Vergleich aller Gruppen in *einem* Test

- dies leistet die **Varianzanalyse**
- α bleibt bei 0,05.

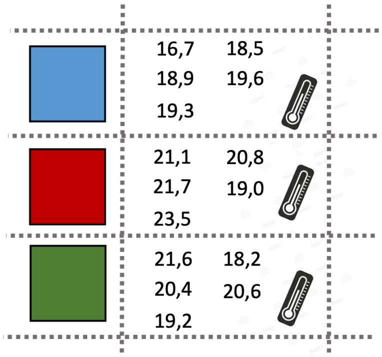
Varianzanalyse

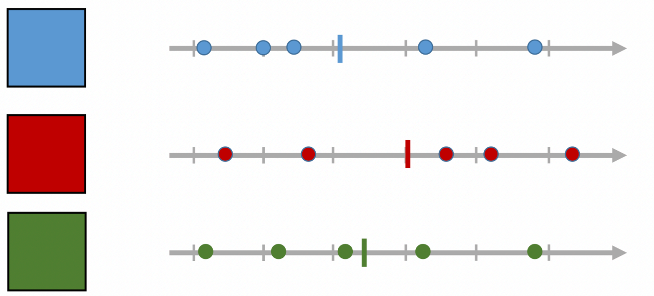
→ Nullhypothese:

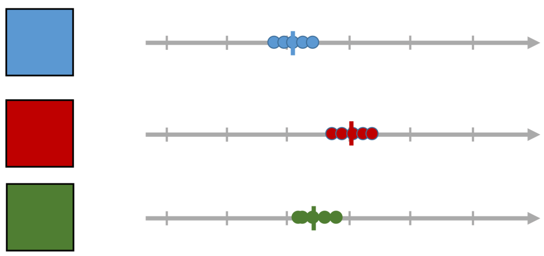
- **alle** betrachteten Populationsmittelwerte sind gleich
$H_0: \mu_{blau} = \mu_{rot} = \mu_{grün}$

→ Alternativhypothese:

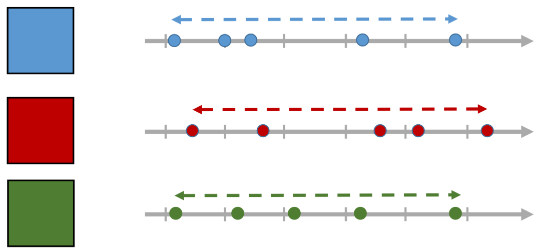
- nicht alle betrachteten Populationsmittelwerte sind gleich
- $\mu_{blau} \neq \mu_{rot}$ $\mu_{blau} \neq \mu_{grün}$ $\mu_{rot} \neq \mu_{grün}$
- nach einem signifikanten Ergebnis ist also **nicht** klar, welche 2 Populationsmittelwerte sich unterscheiden oder ob sich alle 3 unterscheiden,
- Omnibushypothese
- eine spezifischere Alternativhypothese kann in der gewöhnlichen Varianzanalyse nicht getestet werden,
- post-hoc Tests - nach einem signifikanten Ergebnis in der ANOVA
- falls vorab spezifischere Hypothesen bestehen: Kontraste



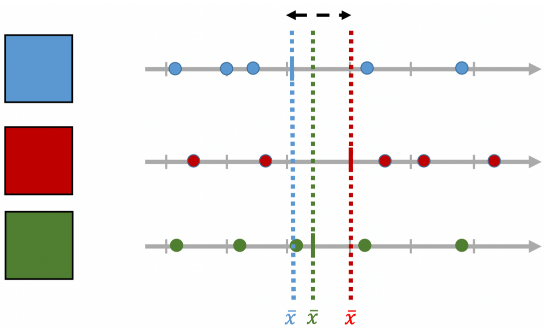




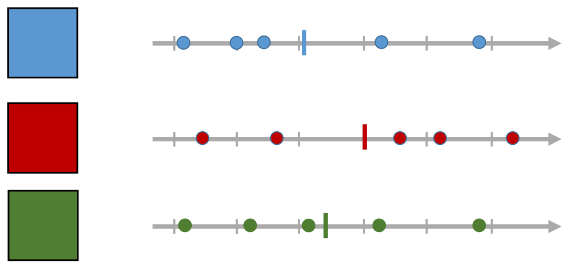
zwei *Arten* von Varianz!



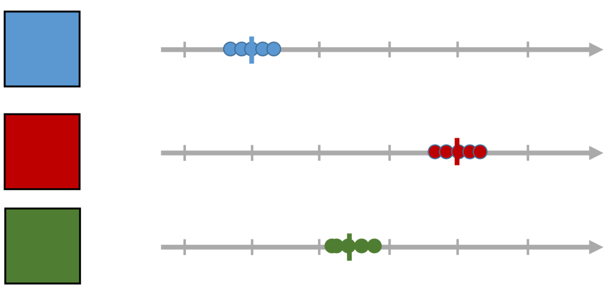
große Varianz **innerhalb** von Gruppen



kleine Varianz **zwischen** Gruppenmittelwerten



schwache Evidenz für Unterschiede zwischen den Populationsmittelwerten → H_0 wird beibehalten



starke Evidenz für Unterschiede zwischen den Populationsmittelwerten → H_0 wird verworfen, H_1 angenommen

in der Varianzanalyse wird die *Varianz innerhalb* mit der *Varianz zwischen* verglichen - daher der Name,

$$F = \frac{Varianz\ zwischen}{Varianz\ innerhalb}$$

↙

Teststatistik der Varianzanalyse!

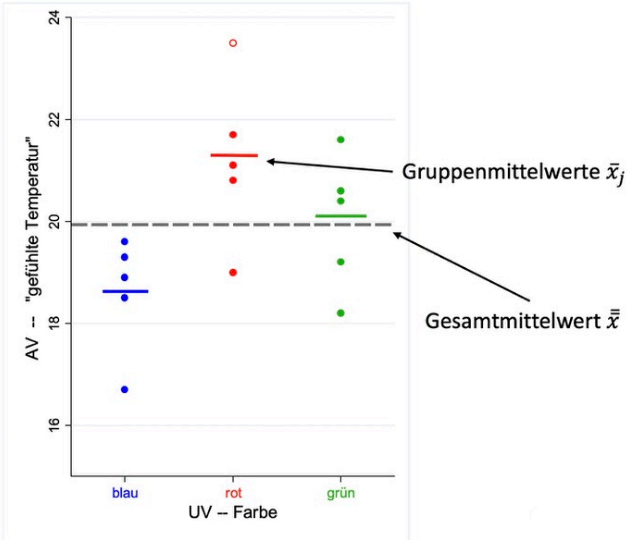
Teststatistik =	$\frac{\text{erklärte Varianz}}{\text{nicht erkl. Varianz}}$	=	$\frac{\text{Signal}}{\text{Rauschen}}$	=	$\frac{\text{Effekt}}{\text{Fehler}}$

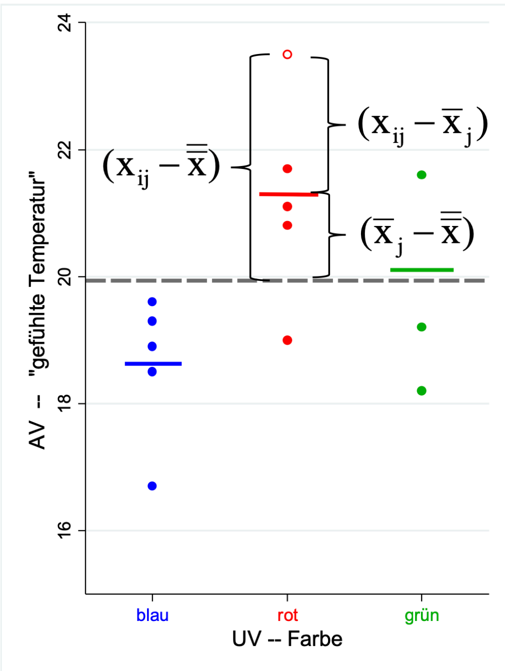
→ dieses *Prinzip* findet sich in Abwandlungen in vielen Signifikanztests wieder...

Varianzzerlegung
- benötigt werden Maße für die Varianz zwischen und innerhalb!
- unsere Daten:

blau	rot	grün
16,7	19,0	18,2
18,5	20,8	19,2
18,3	21,1	20,4
19,3	21,7	20,6
19,6	23,5	21,6
$\bar{x}_{blau}$ = 18,60	$\bar{x}_{rot}$ = 21,22	$\bar{x}_{grün}$ = 20,00
$\hat{\sigma}^2_{blau}$ = 1,30	$\hat{\sigma}^2_{rot}$ = 2,64	$\hat{\sigma}^2_{grün}$ = 1,74

$\bar{\bar{x}}$ = 19,94





die Abweichung eines Messwerts vom Gesamtmittelwert lässt sich zerlegen in
- die Abweichung des Messwerts von seinem Gruppenmittelwert und
- die Abweichung seines Gruppenmittelwerts vom Gesamtmittelwert

$$(x_{ij} - \bar{\bar{x}}) = (x_{ij} - \bar{x}_j) + (\bar{x}_j - \bar{\bar{x}})$$

23,5 – 19,94 = (23,5 – 21,22) + (21,22 – 19,94)

→ die Abweichung der Gruppenmittelwerte vom Gesamtmittelwert (und somit von der Varianz der Gruppenmittelwerte) kann als Wirkung der uV interpretiert werden
→ **systematische Varianz**,

→ die Abweichung der Messwerte von ihrem Gruppenmittelwert (und somit die Varianz innerhalb der Gruppen) kann nicht durch die uV erklärt werden → **Fehlervarianz**

Quadratsummen in der einfaktoriellen Varianzanalyse

$$QS_{Ges} = \sum_{j=1}^{k}\sum_{i=1}^{n_i}(x_{ij} - \bar{\bar{x}})^2 \qquad QS_{inn} = \sum_{j=1}^{k}\sum_{i=1}^{n_i}(x_{ij} - \bar{x}_j)^2 \qquad QS_{zw} = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2$$

$$\sum_j\sum_i(x_{ij} - \bar{\bar{x}})^2 = \sum_j\sum_i(x_{ij} - \bar{x}_j)^2 + \sum_j n_j(\bar{x}_j - \bar{\bar{x}})^2$$

$$QS_{gesamt} = QS_{inn} + QS_{zw}$$

Gesamtvariation — Fehlervariation — systemische Variation

blau	rot	grün
16,7	19,0	18,2
18,5	20,8	19,2
18,3	21,1	20,4
19,3	21,7	20,6
19,6	23,5	21,6

$\bar{x}_{blau} = 18{,}60$ $\bar{x}_{rot} = 21{,}22$ $\bar{x}_{gtün} = 20{,}00$

$\bar{\bar{x}} = 19{,}94$

$$QS_{Ges} = \sum_{j=1}^{k}\sum_{i=1}^{n_i}(x_{ij} - \bar{\bar{x}})^2$$

$$QS_{Ges} = (16{,}7 - 19{,}94)^2 + (18{,}5 - 19{,}94)^2 + ... + (19 - 19{,}94)^2 + (20{,}8 - 19{,}94)^2 + ... + (18{,}2 - 19{,}94)^2 + ... + (21{,}6 - 19{,}94)^2 = 39{,}90$$

blau	rot	grün
16,7	19,0	18,2
18,5	20,8	19,2
18,3	21,1	20,4
19,3	21,7	20,6
19,6	23,5	21,6

$\bar{x}_{blau} = 18{,}60$ $\bar{x}_{rot} = 21{,}22$ $\bar{x}_{gtün} = 20{,}00$

$\bar{\bar{x}} = 19{,}94$

$$QS_{zw} = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2 \qquad QS_{inn} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_j)^2$$

$$QS_{zw} = 5 \cdot (18{,}6 - 19{,}94)^2 + 5 \cdot (21{,}22 - 19{,}94)^2 + 5 \cdot (20 - 19{,}94)^2 = 17{,}19$$

$$QS_{inn} = (16{,}7 - 18{,}6)^2 + (18{,}5 - 18{,}6)^2 + ... + (19 - 21{,}22)^2 + (20{,}8 - 21{,}22)^2 + ... + (18{,}2 - 20)^2 + ... + (21{,}6 - 20)^2 = 22{,}71$$

→ aus diesen Quadratsummen werden Varianzschätzungen berechnet,
→ dazu werden Freiheitsgrade benötigt,

- Freiheitsgrade innerhalb: $df_{inn} = N - k$
- Freiheitsgrade zwischen: $df_{zw} = k - 1$

→ im Beispiel bei $N = 15$ Teilnehmern und $k = 3$ Gruppen:
- $df_{zw} = 3 - 1 = 2$

Zwischen: $\hat{\sigma}^2_{zw} = MS_{zw} = \frac{QS_{zw}}{df_{zw}}$ Im Beispiel: $\hat{\sigma}^2_{zw} = \frac{17{,}19}{2} = 8{,}59$

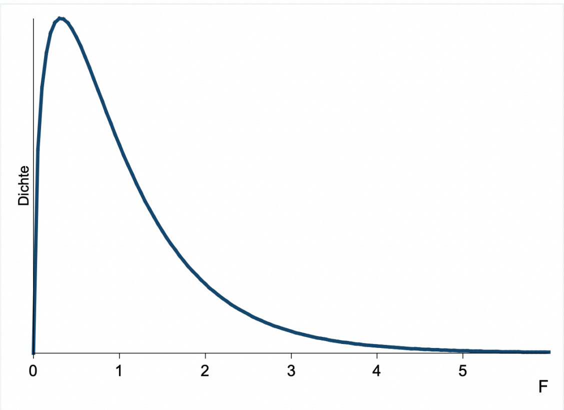
Innerhalb: $\hat{\sigma}^2_{inn} = MS_{inn} = \frac{QS_{inn}}{df_{inn}}$ Im Beispiel: $\hat{\sigma}^2_{inn} = \frac{22{,}71}{12} = 1{,}89$

F-Wert, *F*-Verteilung und Signifikanzprüfung

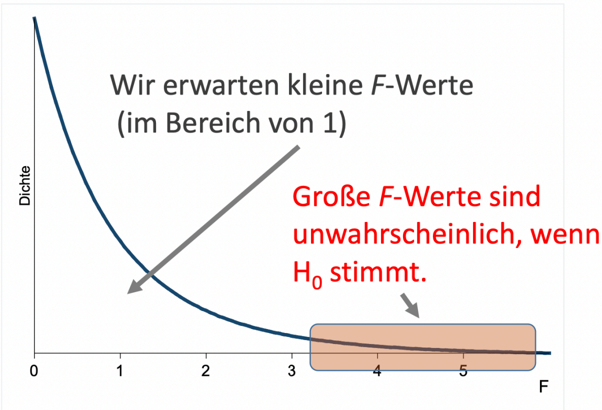
$$F = \frac{Varianz\ zwischen}{Varianz\ innerhalb} = \frac{\hat{\sigma}^2_{zw}}{\hat{\sigma}^2_{inn}}$$ Im Beispiel: $$F = \frac{8{,}59}{1{,}89} = 4{,}54$$

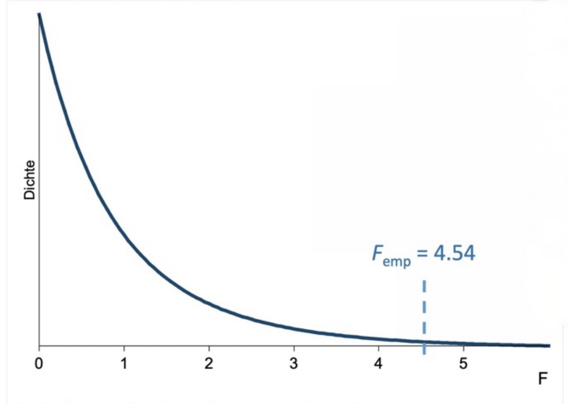
? in welchem Bereich sollte der beobachtete *F*-Wert liegen, **wenn die Nullhypothese stimmt?**
→ *F*-Verteilung
→ standardisierte Stichprobenkennwerteverteilung,
? wenn die $\mathbf{H_0}$ **korrekt** ist und wir unsere Studie unendlich oft wiederholen und jeweils einen F-Wert bestimmen, wie verteilen sich dann die resultierenden unendlich vielen F-Werte?
? wie sieht die *F*-Verteilung aus?
→ abhängig von den Freiheitsgraden

mit $df_{zw} = 3$ und $df_{inn} = 40$



unser Fall:
mit $df_{zw} = 2$ und $df_{inn} = 12$



	*p*-Wert: mit welcher Wahrscheinlichkeit tritt dieser oder ein noch größerer *F*- Wert auf, wenn die H_0 stimmt? entspricht dem Flächenanteil rechts von $F_{emp} = 4.54$ hier: $p = 0.034$ Signifikanzniveau: $\alpha = 0.05$ **! signifikant, da p < α !**

Schlussfolgerung

→ wir verwerfen die H_0 und nehmen die H_1 an,

→ wir schließen also, dass nicht alle untersuchten Populationsmittelwerte gleich sind:

  → die mittlere Temperaturschätzung in rot, blau und grün beleuchteten Räumen ist in der Population nicht gleich,

→ wir haben aber nicht geprüft, **welche** der Populationsmittelwerte sich unterscheiden,

→ bei einem nicht-signifikanten Ergebnis hätten wir ohne vorherige Poweranalyse nichts schließen können

→ weiterhin wichtig:

  → Effektstärken in der Varianzanalyse
  → Power
  → post-hoc Test
  → Voraussetzungen und Voraussetzungsprüfung
  → Kontraste und Bezug zum allgemeinen linearen Modell

Effektstärken in der Varianzanalyse

→ als Effektstärke wird in der Varianzanalyse η^2 (griechisch: *eta*) verwendet,

→ η^2 gibt den Anteil der durch die uV erklärten Variation an der gesamten Variation an,

$$\eta^2 = \frac{QS_{zw}}{QS_{Ges}}$$

→ falls die Quadratsummen nicht verfügbar sind:

$$\eta^2 = \frac{F \cdot df_{zw}}{F \cdot df_{zw} + df_{inn}}$$

→ im Beispiel:

$$\eta^2 = \frac{17{,}19}{39{,}90} = 0{,}43 \qquad \text{oder} \qquad \eta^2 = \frac{4{,}54 \cdot 2}{4{,}54 \cdot 2 + 12} = 0{,}43$$

! 43% der gesamten Varianz in den Messwerten gehen also auf die Raumfarbe zurück!

→ η^2 beschreibt den Effekt in der Stichprobe, liefert aber **keine** unverzerrte erwartungstreue Schätzung des Effekts in der Population,
→ eine entsprechende Schätzung erbringt ω^2 (griechisch: *omega*),
→ auch *omega* beschreibt den Anteil der erklärten Variation an der gesamten Variation,
→ ω^2 wird etwas kleiner ausfallen als η^2

$$\omega^2 = \frac{QS_{zw} - df_{zw} \cdot \hat{\sigma}^2_{inn}}{QS_{ges} + \hat{\sigma}^2_{inn}}$$

- im Beispiel:

$$\omega^2 = \frac{17{,}19 - 2 \cdot 1{,}89}{39{,}90 + 1{,}89} = 0{,}32$$

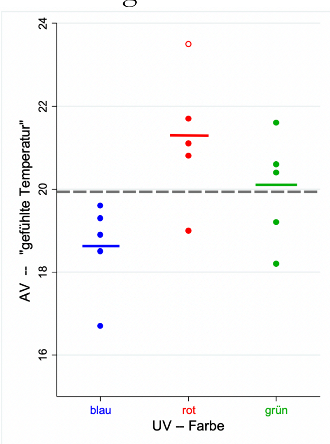
→ der relativ deutliche Unterschied zwischen ω^2 und η^2 ist hier durch die kleine Stichprobengröße verursacht!

→ Konventionen zur Beurteilung der Effektgrößen:
  - kleiner Effekt: $\eta^2 = 0{,}01$
  - mittlerer Effekt: $\eta^2 = 0{,}06$
  - großer Effekt: $\eta^2 = 0{,}14$

→ auch auf ω^2 anwendbar,
→ Effektstärken sollten nach Möglichkeit immer im Kontext anderer Forschungsergebnisse interpretiert werden!
  → Konventionen sind eine *Krücke*, die hilft, wenn keinerlei andere Informationen vorliegen!

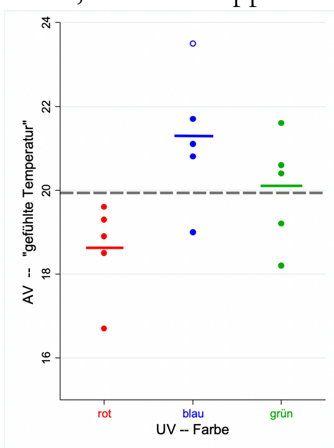
<u>Probleme mit η^2 und ω^2</u>

Originaldaten:



$QS_{zw} = 17{,}19$, $QS_{inn} = 22{,}71$, $\eta^2 = 0{,}43$

gleiche Daten, andere Gruppenzuordnung:



$QS_{zw} = 17{,}19$, $QS_{inn} = 22{,}71$, $\eta^2 = 0{,}43$

→ tatsächlich geht in beiden Untersuchungen 43% der Gesamtvariation auf die Raumfarbe zurück,

→ passt zur Omnibus-Hypothese der Varianzanalyse

→ η^2 und ω^2 liefern insoweit eine brauchbare Aussage über die Stärke des Effekts, η^2 und ω^2 können daher auch als Effektmaß in Poweranalysen verwendet werden,

→ **aber:** offensichtlich wurde **nicht** in beiden Untersuchungen der gleiche Effekt gefunden,

→ η^2 und ω^2 erlauben also keine inhaltliche Aussage über die Richtung des Effekts!

Power

Teilnehmer pro Gruppe (n)	$\alpha = 0{,}01$			$\alpha = 0{,}05$			$\alpha = 0{,}10$		
	$\eta^2 = 0{,}01$ (klein)	$\eta^2 = 0{,}06$ (mittel)	$\eta^2 = 0{,}14$ (groß)	$\eta^2 = 0{,}01$ (klein)	$\eta^2 = 0{,}06$ (mittel)	$\eta^2 = 0{,}14$ (groß)	$\eta^2 = 0{,}01$ (klein)	$\eta^2 = 0{,}06$ (mittel)	$\eta^2 = 0{,}14$ (groß)
Drei Gruppen ($df_{zw} = 2$)									
5	0,01	0,03	0,07	0,06	0,11	0,21	0,11	0,19	0,33
10	0,02	0,06	0,21	0,07	0,20	0,44	0,13	0,30	0,58
20	0,03	0,17	0,54	0,10	0,37	0,78	0,17	0,50	0,86
30	0,04	0,30	0,79	0,12	0,54	0,93	0,20	0,66	0,96
40	0,05	0,43	0,92	0,15	0,68	0,98	0,24	0,78	0,99
50	0,06	0,56	0,97	0,18	0,78	0,99	0,27	0,86	*
Vier Gruppen ($df_{zw} = 3$)									
5	0,01	0,03	0,08	0,06	0,12	0,24	0,12	0,20	0,36
10	0,02	0,07	0,25	0,07	0,21	0,50	0,13	0,32	0,63
20	0,03	0,20	0,65	0,10	0,42	0,85	0,17	0,55	0,91
30	0,04	0,36	0,88	0,13	0,61	0,96	0,21	0,72	0,98
40	0,05	0,52	0,97	0,16	0,75	0,99	0,25	0,84	*
50	0,06	0,66	0,99	0,19	0,85	*	0,29	0,91	*

Benötigte Anzahl von Teilnehmern pro Gruppe, damit in einer einfaktoriellen Varianzanalyse mit dem Signifikanzkriterium $\alpha = 0{,}05$ eine Power von 80% erreicht wird

	$\eta^2 = 0{,}01$ (klein)	$\eta^2 = 0{,}06$ (mittel)	$\eta^2 = 0{,}14$ (groß)
Drei Gruppen ($df_{zw} = 2$)	322	52	21
Vier Gruppen ($df_{zw} = 3$)	274	45	18
Fünf Gruppen ($df_{zw} = 4$)	240	39	16

Omnibus-Hypothese und post-hoc Tests

→ Problem: durch den gleichzeitigen Vergleich aller Gruppen wird eine Omnibus-Hypothese getestet,
  - → bei einem signifikanten Ergebnis bleibt unklar, zwischen welchen Gruppen ein Unterschied besteht,

→ post-hoc Tests versuchen **nach** einem signifikanten Ergebnis in der ANOVA zu ermitteln, wo dieser Unterschied auftrat, indem sie **alle möglichen Paarvergleiche** durchführen,

→ sie testen damit Hypothesen, die **vor** der Untersuchung nicht bestanden → eigentlich eher hypothesengenerierend als hypothesentestend,

→ post-hoc Tests stehen vor demselben Problem, wie mehrere *t*-Tests → entweder steigt der kumulierte α-Fehler oder die Power sinkt,

→ verschiedene post-hoc Tests bieten hier unterschiedliche Kompromisse, sind also unterschiedlich teststark,

post-hoc Tests

→ SPSS bietet derzeit 18 unterschiedliche Verfahren an,

→ eine grobe Übersicht:

t-Tests/LSD	↑	große Power/*liberale* Kontrolle des α-Fehlers
Studentized Newman-Keuls Test		
REGWQ		
Tukey-Test		
Scheffé-Test		
Dunn		
Bonferroni-Korrektur	↓	geringe Power/*strenge* Kontrolle des α-Fehlers

Anwendungsvoraussetzungen der einfaktoriellen Varianzanalyse

1. die aV muss mindestens Intervallskalenniveau besitzen,
2. Stichproben/Gruppen müssen unabhängig sein,
   - → andernfalls ist die Varianzanalyse für unabhängige Gruppen das falsche Verfahren!
3. alle Stichproben/Gruppen müssen aus normalverteilten Populationen stammen,
4. Varianzhomogenität: alle Stichproben/Gruppen müssen aus Populationen mit gleicher Varianz stammen → **Varianzhomogenität**
   - → sind die Voraussetzungen 3 oder 4 nicht erfüllt, so entspricht die Stichprobenkennwerteverteilung nicht exakt der *F*-Verteilung,
   - → folglich ist der errechnete *p*-Wert nicht exakt korrekt
     - → Konsequenz?
   - → infolge von Voraussetzungsverletzungen…

     …kann der tatsächliche *p*-Wert unterschätzt werden
     - wir erhalten **zu viele** signifikante Ergebnisse, wenn die H_0 stimmt,
     - das tatsächliche α-Niveau ist **größer** als angenommen,
     - der Test entscheidet **progressiv**,

     …der tatsächliche *p*-Wert überschätzt werden
     - wir erhalten **zu wenige** signifikante Ergebnisse, wenn die H_0 stimmt,
     - das tatsächliche α-Niveau ist **kleiner** als angenommen,
     - der Test entscheidet **konservativ**,

     …der ermittelte *p*-Wert immer noch näherungsweise mit dem tatsächlichen

übereinstimmen,
- das α-Niveau liegt in etwa beim angenommenen Wert,
- das Verfahren ist **robust**

→ ist die Varianzanalyse robust?
- sie ist in jedem Fall robuster, wenn die Stichproben in den verschiedenen Gruppen **gleich groß** und **groß** sind → gilt für alle Verfahren des ALM im Zweifel sollten Sie nicht davon ausgehen, dass mögliche Voraussetzungsverletzungen ignoriert werden können,

→ was tun bei Voraussetzungsverletzungen?
- **Welch's *F*** und **Brown-Forsythe *F*** korrigieren den Einfluss heterogener Varianzen

→ diese Verfahren liefern ein verändertes *F* und veränderte Freiheitsgrade - i.d.R. mit einem Powerverlust verbunden,

→ der **Kruskal-Wallis-Test** ist eine non-parametrische Alternative, z.B. wenn auch die Normalverteilungsannahme problematisch ist - ebenfalls mit Powerverlust verbunden,

→ *robuste* Varianten der Varianzanalyse: Bootstrapping...

Prüfung der Voraussetzungen

→ die Voraussetzungen beziehen sich auf die zugrundeliegenden Populationen,

→ es lässt sich also nie mit Sicherheit sagen, dass die Voraussetzungen erfüllt sind,

→ es existieren Tests, zur Prüfung der Voraussetzungen:
- der Levene-Test prüft die Nullypothese, dass die Varianz in allen Populationen
- gleich ist,
- der Kolmogorov-Smirnov-Test prüft die Nullhypothese, dass eine Variable normalverteilt ist,
    - diese Tests sind mit Vorsicht zu genießen!
        - kleine Power bei kleiner Stichprobe - gerade dann sind Voraussetzungsverletzungen aber relevant!
        - große Power bei großer Stichprobe - dann können Voraussetzungsverletzungen aber eher ignoriert werden

→ sie sollten sich Ihre Daten vorab angesehen haben!

→ verschiedene graphische Verfahren...

Bezug zum ALM und Kontraste

? wie könnte man die gesamte ursprüngliche Varianzanalyse auch durch eine Regression ersetzen?
- in der Regression müssen Prädiktoren (x_1, x_2...) und Kriterium (y) eigentlich intervallskaliert sein,
    - unsere uV ist offensichtlich nominalskaliert,
- wir umgehen dieses Problem, indem wir die uV in mehrere zwei-stufige Variablen zerlegen,
- die resultierenden Variablen heißen Dummy-Variablen
- zwei-stufige Variablen können behandelt werden, als wären sie intervallskaliert

Dummy-Codierung

→ bei *k* Bedingungen werden *k* – 1 Dummy-Variablen benötigt:

Bedingung	Dummy 1 (rot?)	Dummy 2 (grün?)
blau	0	0
rot	1	0
grün	0	1

0 nein, 1 ja

Regressionsmodell: $\hat{y} = b_0 + b_1 * x_1 + b_2 * x_2$

Dummy 1 Dummy 2

→ die Regression wird für jeden Teilnehmer den bestmöglichen Wert vorhersagen,

→ der bestmögliche Wert entspricht hier dem Mittelwert in der jeweiligen Gruppe,

- also, für blau - unsere Referenzbedingung:
  $\hat{y} = b_0 + b_1 * 0 + b_2 * 0$
  $\hat{y} = b_0$
  → b_0 entspricht also dem Mittelwert in der Bedingung blau - allgemein dem Mittelwert in der mit Nullen codierten Referenzbedingung,
- also, für rot - unsere Referenzbedingung:
  $\hat{y} = b_0 + b_1 * 1 + b_2 * 0$
  $\hat{y} = b_0 + b_1$
  → da $\hat{y}$ der Mittelwert in der Bedingung rot sein muss und b_0 der Mittelwert in der Bedingung blau ist, muss b_1 die Mittelwertsdifferenz zw. den Bedingungen rot und blau sein,
- also, für grün - unsere Referenzbedingung:
  $\hat{y} = b_0 + b_1 * 0 + b_2 * 1$
  $\hat{y} = b_0 + b_2$
  → da $\hat{y}$ der Mittelwert in der Bedingung grün sein muss und b_0 der Mittelwert in der Bedingung blau ist, muss b_2 die Mittelwertsdifferenz zw. den Bedingungen grün und blau sein,

→ die entsprechende Regression im SPSS

	Tn_ID	UV	AV	Dummy1	Dummy2
1	1	blau	16,7	0	0
2	2	blau	18,5	0	0
3	3	blau	18,9	0	0
4	4	blau	19,3	0	0
5	5	blau	19,6	0	0
6	6	rot	19,0	1	0
7	7	rot	20,8	1	0
8	8	rot	21,1	1	0
9	9	rot	21,7	1	0
10	10	rot	23,5	1	0
11	11	grün	18,2	0	1
12	12	grün	19,2	0	1
13	13	grün	20,4	0	1
14	14	grün	20,6	0	1
15	15	grün	21,6	0	1

Ergebnis:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,656[a]	,431	,336	1,3756

a. Predictors: (Constant), Dummy2, Dummy1

η^2 und ω^2 aus der Varianzanalyse!

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	17,188	2	8,594	4,541	,034[b]
	Residual	22,708	12	1,892		
	Total	39,896	14			

a. Dependent Variable: AV

b. Predictors: (Constant), Dummy2, Dummy1

Der F-Test und alle entsprechenden Zwischenergebnisse!

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	18,600	,615		30,234	,000
	Dummy1	2,620	,870	,757	3,011	,011
	Dummy2	1,400	,870	,405	1,609	,134

a. Dependent Variable: AV

t-Test für den Vergleich zwischen blau und rot!

t-Test für den Vergleich zwischen blau und grün!

Diese t-Tests unterliegen einer α-Inflation!

Kontraste

- ! wir können nun durch andere Codierungen andere Vergleiche zwischen Gruppen
- ! spezifizieren
  diese Kontraste können so gestaltet sein, dass Sie keiner α-Inflation unterliegen - orthogonale Kontraste!
- ! in jedem Fall sollten für die entsprechenden Vergleiche **vorab** Hypothesen bestehen - *planned comparisons*

orthogonale Kontraste

- ! Aufteilung der systematischen Variation QS_{zw} in unterschiedliche **Paarvergleiche**,
- ! diese Paarvergleiche müssen **unabhängig** sein,
  - Information aus einem Vergleich, darf in keinem anderen Vergleich erneut verwendet werden
  - konkret: eine einzelne Bedingung kann nur in einem Vergleich vorkommen!
- ! bei *k* Bedingungen sind *k* -1 orthogonale Vergleiche/Kontraste möglich,
- ! die gewünschten Vergleiche werden durch andere Codierungen der Dummy-Variablen spezifiziert - den Bedingungen werden **Gewichte** zugeordnet,
- ! welche Vergleiche gewünscht sind, ergibt sich aus vorab definierten Hypothesen
- ! Hypothesen für die Beispielstudie:
  - das Temperaturempfinden ist bei den kalten Farben **blau** und **grün** geringer als bei der Farbe **rot**
  - Vergleich des (Gesamt-)Mittelwerts in den Bedingungen **blau** und **grün** mit der Bedingung **rot** - Kontrast 1
  - das Temperaturempfinden ist bei **blau** niedriger als bei **grün** - Kontrast 2

→ entsprechende Kontrastkodierung:

Bedingung	Kontrast 1	Kontrast 2
blau	-1	-1
rot	2	0
grün	-1	1

→ orthogonale Kontraste:

Bedingung	Kontrast 1	Kontrast 2	Produkt
blau	-1	-1	1
rot	2	0	0
grün	-1	1	-1
Summe	0	0	0

→ Regeln für Kontrastgewichte:
  - Gruppen, die in einem Kontrast zusammengefasst werden, erhalten das gleiche Gewicht,
  - die Summe der Gewichte ist 0
  - Vorzeichen codieren die Richtung des erwarteten Effekts
  - aus einem Kontrast ausgeschlossene Bedingungen erhalten eine 0
  - **zwei Kontraste sind unabhängig, wenn die Summe der Produkte der Gewichte 0 ist!**

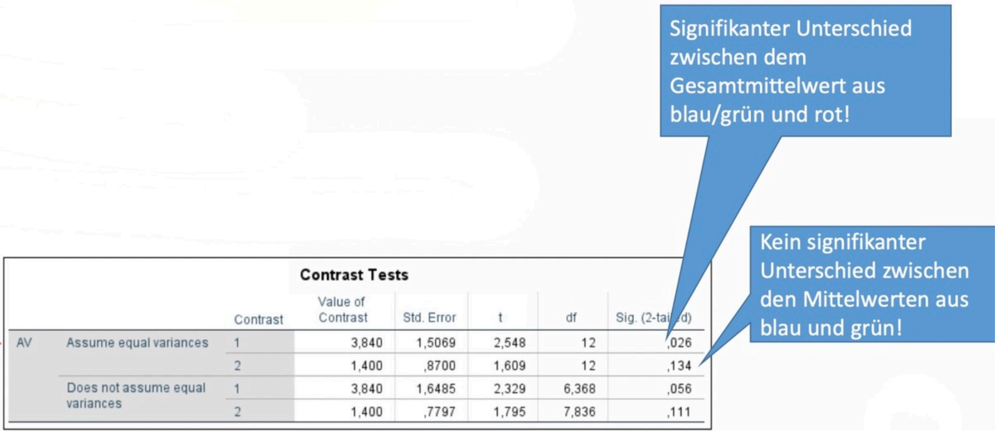
Regressionsmodell: $\hat{y} = b_0 + b_1 * x_1 + b_2 * x_2$

↙ ↘

Kontrast 1 - Vergleich von blau, grün mit rot

Kontrast 2 – Vergleich blau mit grün

Ergebnis:



Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
AV	Assume equal variances	1	3,840	1,5069	2,548	12	,026
		2	1,400	,8700	1,609	12	,134
	Does not assume equal variances	1	3,840	1,6485	2,329	6,368	,056
		2	1,400	,7797	1,795	7,836	,111

→ nicht-orthogonale Kontraste

- Hypothesen:
  - das Temperaturempfinden ist bei **rot** höher als bei **grün**
  - das Temperaturempfinden ist bei **grün** höher als bei **blau**

Bedingung	Kontrast 1	Kontrast 2	Produkt
blau	0	-1	0
rot	1	0	0
grün	-1	1	-1
Summe	0	0	-1

→ gegen nicht-orthogonale Kontraste spricht nichts, wenn man entsprechende Hypothesen hat,

→ allerdings sollte eine Korrektur von α zumindest erwogen werden!

Beispiel:

die Datei Puppies.sav (im Ordner *SPSS-Dateien aus Field*) enthält Daten aus einer fiktiven Studie, deren Hintergrund in der tiergestützten Therapie besteht - die Teilnehmer der Kontrollgruppe haben keinen Kontakt mit einem Welpen, die Teilnehmer der beiden Experimentalgruppen verbringen 15min oder 30min mit einem Welpen - die aV ergibt sich aus *Happiness-Ratings* (0 bis 10, je höher desto mehr *happiness*),

wir wollen untersuchen, ob der Welpenkontakt wirkt, ob also Unterschiede im Wohlbefinden zwischen den verschiedenen Bedingungen bestehen - darüber hinaus haben wir aber auch zwei spezifischere Hypothesen:

a) unabhängig von der Dauer des Kontakts verbessert jeder Umgang mit Welpen das Wohlbefinden

b) eine längere Kontaktdauer führt zu einem größeren Wohlbefinden als eine kürzere,

- führen Sie eine Varianzanalyse durch!
- prüfen Sie dabei die Voraussetzung der Varianzhomogenität und interpretieren das Ergebnis dieser Prüfung!
- interpretieren Sie das Ergebnis der Varianzanalyse!
- bestimmen Sie eine geeignete Effektgröße!
- wählen Sie mindestens zwei post-hoc Tests aus und interpretieren Sie die Ergebnisse!
- bestimmen Sie Kontraste, um die spezifischere Hypothese zu prüfen und interpretieren Sie die Ergebnisse!

F-Verteilung

Nenner-*df*	Fläche	Zähler-*df* 1	2	3	4	5	6	7	8	9	10
1	0,75	5,83	7,50	8,20	8,58	8,82	8,98	9,10	9,19	9,26	9,32
	0,90	39,90	49,50	53,60	55,80	57,20	58,20	58,90	59,40	59,90	60,20
	0,95	161,00	200,00	216,00	225,00	230,00	234,00	237,00	239,00	241,00	242,00
2	0,75	2,57	3,00	3,15	3,23	3,28	3,31	3,34	3,35	3,37	3,38
	0,90	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39
	0,95	18,50	19,00	19,20	19,20	19,30	19,30	19,40	19,40	19,40	19,40
	0,99	98,50	99,00	99,20	99,20	99,30	99,30	99,40	99,40	99,40	99,40
3	0,75	2,02	2,28	2,36	2,39	2,41	2,42	2,43	2,44	2,44	2,44
	0,90	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23
	0,95	10,10	9,55	9,28	9,12	9,10	8,94	8,89	8,85	8,81	8,79
	0,99	34,10	30,80	29,50	28,70	28,20	27,90	27,70	27,50	27,30	27,20
4	0,75	1,81	2,00	2,05	2,06	2,07	2,08	2,08	2,08	2,08	2,08
	0,90	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92
	0,95	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
	0,99	21,20	18,00	16,70	16,00	15,50	15,20	15,00	14,80	14,70	14,50
5	0,75	1,69	1,85	1,88	1,89	1,89	1,89	1,89	1,89	1,89	1,89
	0,90	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30
	0,95	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
	0,99	16,30	13,30	12,10	11,40	11,00	10,70	10,50	10,30	10,20	10,10
6	0,75	1,62	1,76	1,78	1,79	1,79	1,78	1,78	1,77	1,77	1,77
	0,90	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94
	0,95	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
	0,99	13,70	10,90	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	0,75	1,57	1,70	1,72	1,72	1,71	1,71	1,70	1,70	1,69	1,69
	0,90	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70
	0,95	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
	0,99	12,20	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	0,75	1,54	1,66	1,67	1,66	1,66	1,65	1,64	1,64	1,64	1,63
	0,90	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54
	0,95	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
	0,99	11,30	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	0,75	1,51	1,62	1,63	1,63	1,62	1,61	1,60	1,60	1,59	1,59
	0,90	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42
	0,95	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
	0,99	10,60	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	0,75	1,49	1,60	1,60	1,59	1,59	1,58	1,57	1,56	1,56	1,55
	0,90	3,28	2,92	2,73	2,61	2,52	2,46	2,41	2,39	2,35	2,32
	0,95	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
	0,99	10,00	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85

F-Verteilung (Fortsetzung)

Nenner-*df*	Fläche	Zähler-*df* 1	2	3	4	5	6	7	8	9	10
11	0,75	1,47	1,58	1,58	1,57	1,56	1,55	1,54	1,53	1,53	1,52
	0,90	3,23	2,86	2,66	1,54	2,45	2,39	2,34	2,30	2,27	2,25
	0,95	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
	0,99	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	0,75	1,46	1,56	1,56	1,55	1,54	1,53	1,52	1,51	1,51	1,50
	0,90	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19
	0,95	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
	0,99	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
13	0,75	1,45	1,54	1,54	1,53	1,52	1,51	1,50	1,49	1,49	1,48
	0,90	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14
	0,95	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
	0,99	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	0,75	1,44	1,53	1,53	1,52	1,51	1,50	1,48	1,48	1,47	1,46
	0,90	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10
	0,95	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
	0,99	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
15	0,75	1,43	1,52	1,52	1,51	1,49	1,48	1,47	1,46	1,46	1,45
	0,90	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06
	0,95	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
	0,99	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	0,75	1,42	1,51	1,51	1,50	1,48	1,48	1,47	1,46	1,45	1,45
	0,90	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03
	0,95	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
	0,99	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	0,75	1,42	1,51	1,50	1,49	1,47	1,46	1,45	1,44	1,43	1,43
	0,90	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00
	0,95	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
	0,99	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	0,75	1,41	1,50	1,49	1,48	1,46	1,45	1,44	1,43	1,42	1,42
	0,90	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98
	0,95	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
	0,99	8,29	6,01	5,09	4,58	5,24	4,01	3,84	3,71	3,60	3,51
19	0,75	1,41	1,49	1,49	1,47	1,46	1,44	1,43	1,42	1,41	1,41
	0,90	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96
	0,95	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
	0,99	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	0,75	1,40	1,49	1,48	1,46	1,45	1,44	1,42	1,42	1,41	1,40
	0,90	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94
	0,95	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
	0,99	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37

mehrfaktorielle Varianzanalyse

→ Ziel: Signifikanztest zur Überprüfung von Mittelwertsunterschieden in Untersuchungen mit mindestens zwei uVs/Faktoren

→ dazu wird die Varianz der aVs analysiert,

→ die Idee der mehrfaktoriellen Varianzanalyse ist dieselbe wie bei der einfaktoriellen Varianzanalyse

  → Anwendung in Untersuchungen mit mehrfaktoriellem Design (Versuchsplan)!

→ Beispiel:

		Faktor A: Schultyp		
		Realschule (Spalte $j = 1$)	Gymnasium (Spalte $j = 2$)	Zeilenmittelwert
Faktor B: Lernsoftware	Programm A (Zeile $l = 1$)	8 10 $\bar{x}_{11} = 10$ 12	19 18 $\bar{x}_{21} = 19$ 20	$\bar{x}_{\bullet 2} = 14{,}5$
	Programm B (Zeile $l = 2$)	13 13 $\bar{x}_{12} = 14$ 16	14 15 $\bar{x}_{22} = 16$ 19	$\bar{x}_{\bullet 1} = 15$
	Spaltenmittelwert	$\bar{x}_{1\bullet} = 12$	$\bar{x}_{2\bullet} = 17{,}5$	$\bar{\bar{x}} = 14{,}75$

Anzahl der Zahlen = Anzahl der Faktoren

2 x 2 Design

Zahlen: Stufen pro Faktor!

		Faktor A: Schultyp		
		Realschule (Spalte $j = 1$)	Gymnasium (Spalte $j = 2$)	Zeilenmittelwert
Faktor B: Lernsoftware	Programm A (Zeile $l = 1$)	8 10 $\bar{x}_{11} = 10$ 12	19 18 $\bar{x}_{21} = 19$ 20	$\bar{x}_{\bullet 2} = 14{,}5$
	Programm B (Zeile $l = 2$)	13 13 $\bar{x}_{12} = 14$ 16	14 15 $\bar{x}_{22} = 16$ 19	$\bar{x}_{\bullet 1} = 15$
	Spaltenmittelwert	$\bar{x}_{1\bullet} = 12$	$\bar{x}_{2\bullet} = 17{,}5$	$\bar{\bar{x}} = 14{,}75$

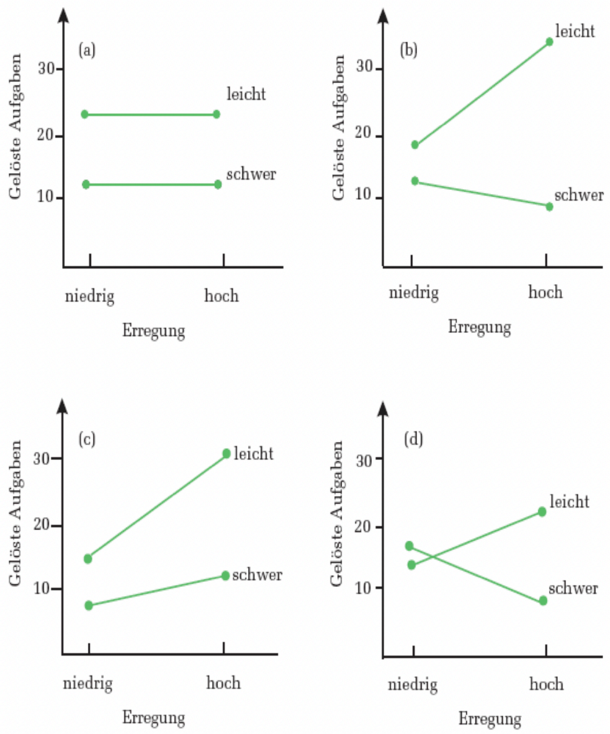
Haupteffekt von Faktor B (Lernsoftware)

Haupteffekt von Faktor A (Schultyp)

Zudem: Interaktion!

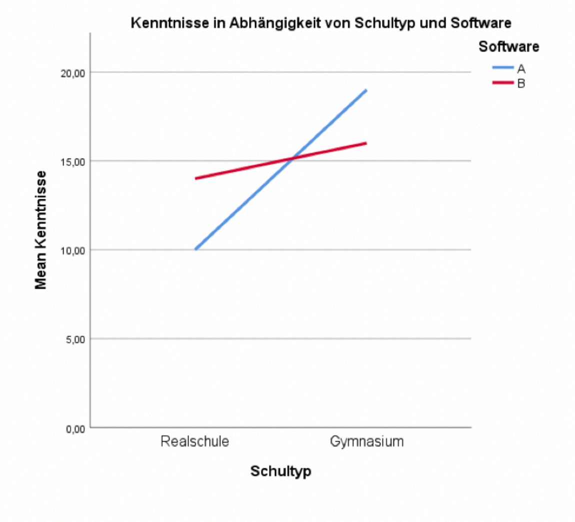
→ was ist eine Interaktion?

- bedingter Mittelwertsunterschied
- die Wirkung einer uV ist abhängig von der Ausprägung mindestens einer anderen uV
- z.B. *die Lernsoftware A führt bei Gymnasiasten zu einer Verbesserung der Geometriekenntnisse, bei Realschülern hingegen zu einer Verschlechterung*

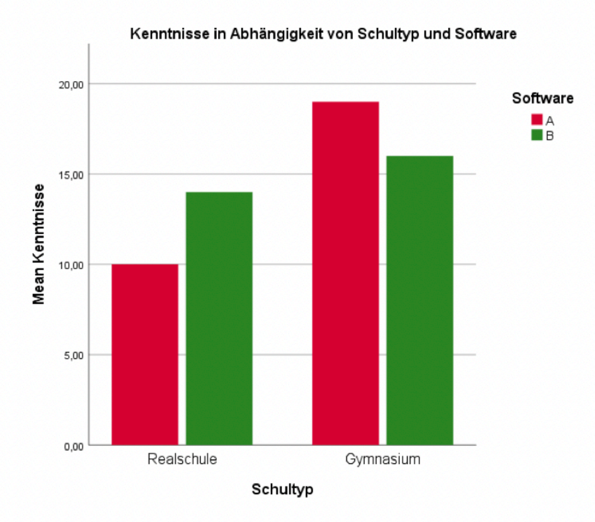


Haupteffekte und Interaktionen?

Interaktionen im Liniendiagramm



Interaktionen im Balkendiagramm

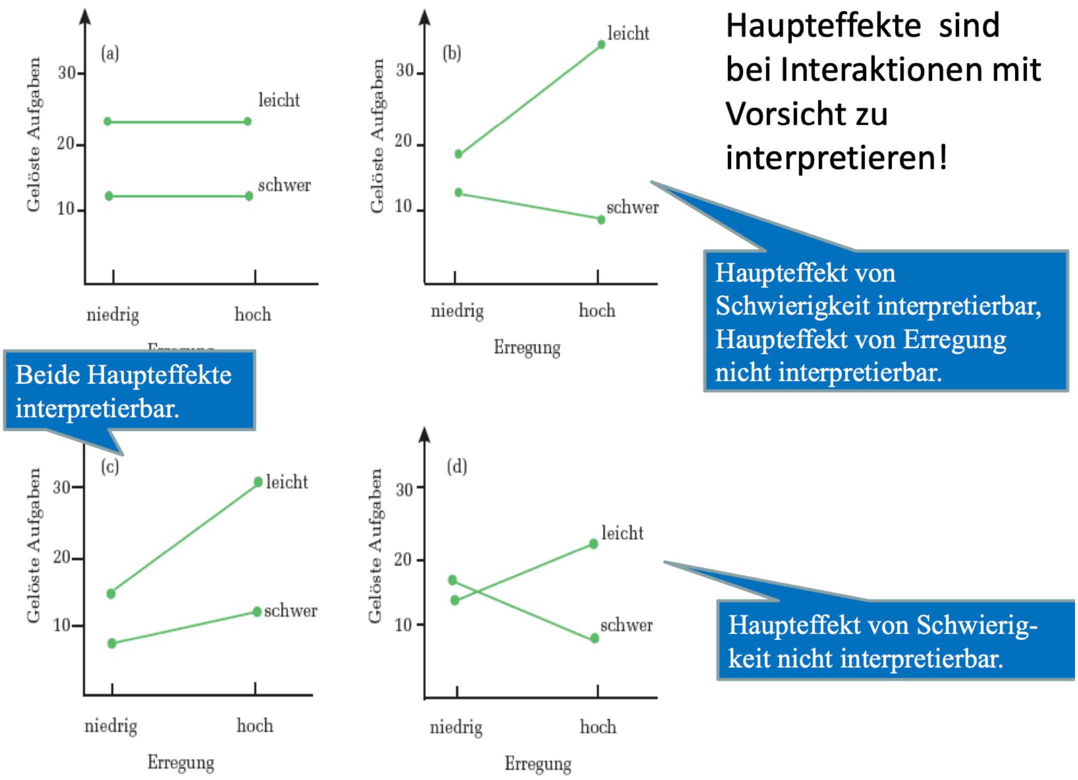


Interaktionen in der Kreuztabelle

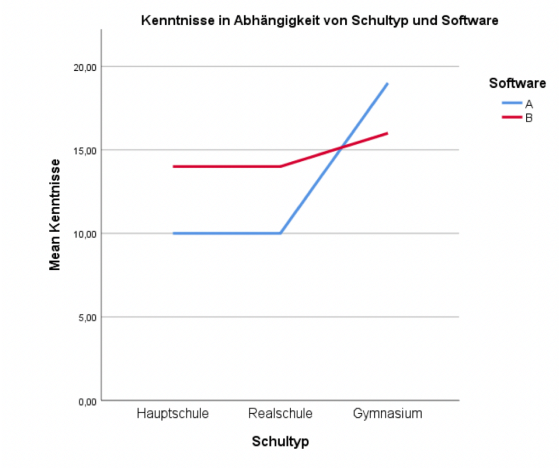
		Schultyp	
		Realschule	Gymnasium
Lernsoftware	Programm A	10	19
	Programm B	14	16

19 – 10 = 9
16 – 14 = 2

→ ungleiche Differenzen zeigen Interaktion an, die dann natürlich noch geprüft werden muss!

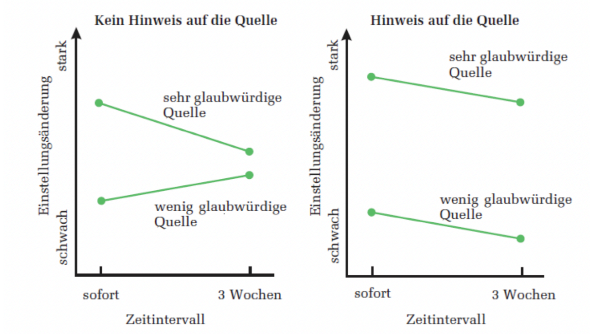


Effekte im 3 x 2 Design



immer noch maximal
2Haupteffekte und eine Interaktion

Effekte im 2 x 2 x 2 Design



Interaktion zweiter Ordnung
- die Interaktion von *Glaubwürdigkeit* und *Zeitintervall* ist davon abhängig, ob ein Quellenhinweis gegeben wurde,

bei 3Faktoren gibt es 7mögliche Effekte:
- Haupteffekt Faktor 1
- Haupteffekt Faktor 2
- Haupteffekt Faktor 3
- Interaktion Faktor 1 und 2
- Interaktion Faktor 1 und 3
- Interaktion Faktor 2 und 3
- Interaktion Faktor 1, 2 und 3 → Interaktion 2. Ordnung)

α-Kumulation in der mehrfaktoriellen Varianzanalyse

→ die Tests der verschiedenen Effekte verhindern eine Kumulation des α-Fehlers **nicht**!

→ bei 2Faktoren und 3Effekten beläuft sich das kumulierte α näherungsweise auf: $1 - 0{,}95^3 = 14{,}3\%$

→ bei 3Faktoren und 7Effekten beläuft sich das kumulierte α näherungsweise auf: $1 - 0{,}95_7 = 30{,}2\%$

→ also:
- ! nicht alle Effekte testen und vorbehaltlos interpretieren!
- ! Interpretation auf diejenigen Effekte beschränken, für die es Hypothesen gab!
- ! alternativ: α korrigieren!

! mit der mehrfaktoriellen Varianzanalyse wird geprüft, ob in der Population Haupteffekte und Interaktionseffekte bestehen!

! mit einer zweifaktoriellen Varianzanalyse werden also 3Nullhypothesen geprüft
- die uV A hat keinen Effekt: H_0: $\mu_{.1} = \mu_{.2} = ... = \mu_{.k}$
- die uV B hat keinen Effekt: H_0: $\mu_{1.} = \mu_{2.} = ... = \mu_{m.}$
- zwischen uV A und uV B besteht keine Interaktion: H_0: $\mu_{11} - \mu_{12} = \mu_{21} - \mu_{22}$

! die Idee ist bei allen Varianten der Varianzanalyse dieselbe:
- ! die Variabilität der aV wird zerlegt in systematische Variabilität, die durch die Wirkung der uVs erklärt werden kann und unsystematische Variabilität, die nicht durch die uVs erklärt werden kann,
- ! als Maß für die jeweilige Variabilität dient stets eine Quadratsumme,
- ! aufgrund der unterschiedlichen Quadratsummen werden verschiedene Schätzungen der Populationsvarianz ermittelt,

! ist die H_0 korrekt, so ist zu erwarten, dass sich die Varianzschätzungen aufgrund systematischer und unsystematischer Variabilität nur zufällig unterscheiden → F = 1

! in einer zweifaktoriellen Varianzanalyse kann die systematische Variabilität auf drei Quellen zurückgeführt werden: uV A, uV B, Interaktion der uVs,

! es gilt:

QS_{ges}	=	QS_A	+	QS_B	+	QS_{A*B}	+	QS_{inn}
Gesamtvariabilität	=	Einfluss uV A	+	Einfluss uV B	+	Einfluss Interaktion	+	Fehler

↘ ↓ ↙

systematische Variabilität

Interaktionsterm

$$\text{Interaktionseffekt} = \text{systematische Abweichung} - \text{Effekt Faktor A} - \text{Effekt Faktor B}$$

$$\text{Interaktionseffekt} = (\bar{x}_{jl} - \bar{\bar{x}}) - (\bar{x}_{j\bullet} - \bar{\bar{x}}) - (\bar{x}_{\bullet l} - \bar{\bar{x}})$$

→ welche Zellenmittelwerte wären ohne Interaktion zu erwarten?

- additiver Effekt!

		Faktor A: Schultyp		
		Realschule (Spalte $j = 1$)	Gymnasium (Spalte $j = 2$)	Zeilen-mittelwert
Faktor B: Lernsoftware	Programm A (Zeile $l = 1$)	8 10 $\bar{x}_{11} = 10$ 12	19 18 $\bar{x}_{21} = 19$ 20	$\bar{x}_{\bullet 2} = 14{,}5$
	Programm B (Zeile $l = 2$)	13 13 $\bar{x}_{12} = 14$ 16	14 15 $\bar{x}_{22} = 16$ 19	$\bar{x}_{\bullet 1} = 15$
	Spaltenmittelwert	$\bar{x}_{1\bullet} = 12$	$\bar{x}_{2\bullet} = 17{,}5$	$\bar{\bar{x}} = 14{,}75$

14,5 – 14,75 = - 0,25
→ Effekt von Programm A!

12 – 14,75 = - 2,75 → Effekt von Realschule!

- erwarteter Mittelwert in Realschule/Programm A: 14,75 – 2,75 – 0,25 = **11,75**
- die Differenz zwischen erwartetem und beobachtetem Mittelwert könnte auf die Interaktion zurückgehen!

	Gesamt-mittelwert $(\bar{\bar{x}})$	+	Effekt Faktor A $(\bar{x}_{j\bullet} - \bar{\bar{x}})$	+	Effekt Faktor B $(\bar{x}_{\bullet l} - \bar{\bar{x}})$	=	Gruppen-mittelwert $(\bar{x}_{jl})$
Realsch. / Progr. A	14,75	+	(−2,75)	+	(−0,25)	=	11,75
Realsch. / Progr. B	14,75	+	(−2,75)	+	0,25	=	12,25
Gymn. / Progr. A	14,75	+	2,75	+	(−0,25)	=	17,25
Gymn. / Progr. B	14,75	+	2,75	+	0,25	=	17,75

additive Effekte für alle Zellen

Quadratsummen in der zweifaktoriellen Varianzanalyse:

$$QS_{gesamt} = \sum_{j}\sum_{l}\sum_{i}(x_{ijl} - \bar{\bar{x}})^2 \qquad QS_{inn} = \sum_{j}\sum_{l}\sum_{i}(x_{ijl} - \bar{x}_{jl})^2$$

$$QS_A = \sum_{j}^{k} n \cdot m(\bar{x}_{j\bullet} - \bar{\bar{x}})^2 \qquad QS_B = \sum_{l}^{m} n \cdot k(\bar{x}_{\bullet l} - \bar{\bar{x}})^2$$

$$QS_{A \times B} = QS_{gesamt} - QS_A - QS_B - QS_{inn}$$

j	Index der Stufen von uV A
l	Index der Stufen von uV B
i	Index der Messwerte in Zelle jl
n	Anzahl der Messwerte in Zelle jl
	Mittelwert in Zelle jl
	Mittelwert der Messwerte in der Bedingung j → Spaltenmittelwert
	Mittelwert aller Messwerte in der Bedingung l → Zeilenmittelwert

Varianzschätzungen in der zweifaktoriellen Varianzanalyse:

$$\hat{\sigma}^2_{inn} = \frac{QS_{inn}}{df_{inn}}$$ df_{inn} = k + m (n-1)

$$\hat{\sigma}^2_A = \frac{QS_A}{df_A}$$ df_A = k - 1

$$\hat{\sigma}^2_B = \frac{QS_B}{df_B}$$ df_B = m - 1

$$\hat{\sigma}^2_{A \times B} = \frac{QS_{A \times B}}{df_{A \times B}}$$ df_{A*B} = (k-1) (m-1)

Prüfgrößen in der zweifaktoriellen Varianzanalyse:

für den Effekt der uV A:

$$F_A = \frac{\hat{\sigma}^2_A}{\hat{\sigma}^2_{inn}}$$

für den Effekt der uV B:

$$F_B = \frac{\hat{\sigma}^2_B}{\hat{\sigma}^2_{inn}}$$

für den Effekt der Interaktion:

$$F_{A \times B} = \frac{\hat{\sigma}^2_{A \times B}}{\hat{\sigma}^2_{inn}}$$

ANOVA-Tabelle für die Ergebnisse aus dem Rechenbeispiel 15.1

Varianzquelle	*QS*	*df*	Varianzschätzung	*F*	*p*
Faktor Schultyp	90,75	1	90,75	24,2	0,001
Faktor Lernsoftware	0,75	1	0,75	0,2	0,667
Interaktion	36,75	1	36,75	9,8	0,014
Fehler	30,00	8	3,75		
gesamt	158,25	11			

Effektstärken in der zweifaktoriellen Varianzanalyse:

→ η^2: Anteil der durch einen bestimmten Effekt erklärten Varianz an der gesamten Varianz

$$\eta_A^2 = \frac{QS_A}{QS_{Ges}} \qquad \eta_B^2 = \frac{QS_B}{QS_{Ges}} \qquad \eta_{A\times B}^2 = \frac{QS_{A\times B}}{QS_{Ges}}$$

unser Beispiel:

$$\eta_A^2 = \frac{90{,}75}{158{,}25} = 0{,}57 \qquad \eta_B^2 = \frac{0{,}75}{158{,}25} = 0{,}005 \qquad \eta_{A\times B}^2 = \frac{36{,}75}{158{,}25} = 0{,}23$$

→ 57% der Varianz in den Leistungen können durch den Faktor *Schultyp* erklärt werden!

→ häufig verwendete Alternative → partielles η_2

- Unterschied zu η_2: beim partiellen η_2 steht nicht die gesamte Variablilität der Messwerte im Nenner, sondern nur diejenige Variabiliät, die noch nicht durch andere Effekte erklärt werden konnte,
- ein partielles η_2 gibt also den Anteil der durch einen bestimmten Effekt erklärten Varianz an der nicht anderweitig erklärten Varianz an,

$$\eta_A^2 = \frac{QS_A}{QS_A + QS_{inn}} \qquad \eta_B^2 = \frac{QS_B}{QS_B + QS_{inn}} \qquad \eta_{A\times B}^2 = \frac{QS_{A\times B}}{QS_{A\times B} + QS_{inn}}$$

- falls die Quadratsummen nicht verfügbar sind:

$$\eta_A^2 = \frac{F_A \cdot df_A}{F_A \cdot df_A + df_{inn}}$$

- unser Beispiel:

$$\eta_A^2 = \frac{90{,}75}{90{,}75+30} = 0{,}75 \qquad \eta_B^2 = \frac{0{,}75}{0{,}75+30} = 0{,}02 \qquad \eta_{A\times B}^2 = \frac{36{,}75}{36{,}75+30} = 0{,}55$$

→ erwartungstreue Effektstärke: ω2
- Varianzaufklärung in der Population
- wir benötigen korrigierte Varianzschätzungen für jeden Effekt:

$$\hat{\sigma}_{A'}^2 = \frac{df_A \cdot (\hat{\sigma}_A^2 - \hat{\sigma}_{inn}^2)}{N} \qquad \hat{\sigma}_{AxB'}^2 = \frac{df_{AXB} \cdot (\hat{\sigma}_{A\,X\,B}^2 - \hat{\sigma}_{inn}^2)}{N}$$

- korrigierte Schätzung für die gesamte Varianz:

$$\hat{\sigma}_{total'}^2 = \hat{\sigma}_{A'}^2 + \hat{\sigma}_{B'}^2 + \hat{\sigma}_{AxB'}^2 + \hat{\sigma}_{inn}^2$$

- Effektberechnung:

$$\omega^2 = \frac{\hat{\sigma}_{effect}^2}{\hat{\sigma}_{total}^2}$$

- unser Beispiel:

$$\hat{\sigma}_{A'}^2 = \frac{1 \cdot (90{,}75 - 3{,}75)}{12} = 7{,}25 \qquad \hat{\sigma}_{B'}^2 = \frac{1 \cdot (0{,}75 - 3{,}75)}{12} = -0{,}25$$

$$\hat{\sigma}_{AxB'}^2 = \frac{1 \cdot (36{,}75 - 3{,}75)}{12} = 2{,}75 \qquad \hat{\sigma}_{total'}^2 = 7{,}25 - 0{,}25 + 2{,}75 + 3{,}75 = 13{,}5$$

Also: $\omega_A^2 = \frac{7{,}25}{13{,}5} = 0.54$

negative Varianz? → unsinnig!
die korrigierte Effektstärke für uV B ist 0!

Power in der mehrfaktoriellen Varianzanalyse

→ der in der Varianzanalyse getestete Effekt einer uV ist grundsätzlich **ein partieller Effekt!**
- systematische Variation anderer uVs bleibt unberücksichtigt,

→ mehrfaktorielle Designs bieten daher die Möglichkeit die Power für den Effekt einer uV zu erhöhen,
- die zusätzlichen uVs können Variation aufklären und damit die Fehlervariation reduzieren,
  - ! **dies ist einer der Gründe, mehrfaktorielle Designs anzuwenden!**

		Faktor A: Schultyp		
		Realschule (Spalte $j = 1$)	Gymnasium (Spalte $j = 2$)	Zeilen-mittelwert
Faktor B: Lernsoftware	Programm A (Zeile $l = 1$)	8 10 $\quad \bar{x}_{11} = 10$ 12	19 18 $\quad \bar{x}_{21} = 19$ 20	$\bar{x}_{\bullet 2} = 14,5$
	Programm B (Zeile $l = 2$)	13 13 $\quad \bar{x}_{12} = 14$ 16	14 15 $\quad \bar{x}_{22} = 16$ 19	$\bar{x}_{\bullet 1} = 15$
	Spaltenmittelwert	$\bar{x}_{1\bullet} = 12$	$\bar{x}_{2\bullet} = 17,5$	$\bar{\bar{x}} = 14,75$

mehrfaktorielle Auswertung für Schultyp:
$QS_A = 90,75$
$QS_{inn} = 30$
$\eta^2 = 0,75 \rightarrow$ **F = 24,2**

einfaktorielle Auswertung für Schultyp:
$QS_A = 90,75$
$QS_{inn} = 30 + 0,75 + 36,75 = 67,5$
$\eta^2 = 0,57 \rightarrow$ **F = 13,44**

Tabelle 15.7

Benötigte Anzahl von Teilnehmern pro Gruppe, damit in einer mehrfaktoriellen Varianzanalyse zu 2 × 2- und 3 × 2-Designs mit dem Signifikanzkriterium $\alpha = 0,05$ eine Power von 80% erreicht wird.

	$\eta_p^2 = ,01$ (klein)	$\eta_p^2 = ,06$ (mittel)	$\eta_p^2 = ,14$ (groß)
2 × 2-Design:			
Alle Effekte	197	33	14
3 × 2-Design:			
Haupteffekt des zweistufigen Faktors	132	22	9
Haupteffekt des dreistufigen Faktors und Interaktionseffekt	162	27	11

→ größte relevante Gruppengröße auswählen!

Anwendungsvoraussetzungen der mehrfaktoriellen Varianzanalyse

- → identisch mit den Voraussetzungen der einfaktoriellen Varianzanalyse
- → die aV muss mindestens Intervallskalenniveau besitzen
- → Stichproben **in allen Zellen** müssen unabhängig sein
- → **alle Stichproben** müssen aus normalverteilten Populationen stammen
- → Varianzhomogenität: **alle Stichproben** müssen aus Populationen mit gleicher Varianz stammen
    - → bei fehlender Varianzhomogenität gibt es auch hier eine Korrektur nach Welch, die allerdings in SPSS nicht verfügbar ist
    - → ansonsten: wiederum Bootstrapping...

<u>post-hoc Tests</u>

- → für die Haupteffekte stehen dieselben Tests zur Verfügung wie in der einfaktoriellen Varianzanalyse
- → diese sind natürlich nur notwendig und sinnvoll, wenn ein Faktor mehr als 2Stufen hat
    - → also überflüssig im 2 x 2 Design...

<u>Kontraste</u>

- → für die Haupteffekte identisch mit der Situation in der einfaktoriellen Varianzanalyse
- → allerdings stehen in SPSS nur vordefinierte Kontraste zur Verfügung, z.B. Helmert-Kontrast - Vergleich jeder Bedingung mit der Kombination aus allen folgenden Bedingungen - also ein orthogonaler Kontrast
    - → eigene Kontraste kann man nur über die Syntax spezifizieren
- → i.d.R. wird man jedoch vor allem an der weiteren Untersuchung einer Interaktion interessiert sein - gängiges Mittel: **simple (main) effects**

<u>simple effects</u>

- → hat eine uV auf jeder Stufe der anderen uV einen Effekt?
    - unser Beispiel: hat die uV *Lernprogramm* einen Effekt in Hauptschule, Realschule, Gymnasium?
- → diese Vergleiche können so spezifiziert werden, dass sie unabhängig sind
    - keine α-Kumulation
- → auch dafür wird in SPSS Syntax benötigt

↙

in unserem Beispiel:

```
GLM Kenntnisse by Schultyp Software
/EMMEANS = TABLES(Schultyp*Software) COMPARE(Software).
```

Univariate Tests

Dependent Variable: Kenntnisse

Schultyp		Sum of Squares	df	Mean Square	F	Sig.
Hauptschule	Contrast	24,000	1	24,000	6,545	,025
	Error	44,000	12	3,667		
Realschule	Contrast	24,000	1	24,000	6,545	,025
	Error	44,000	12	3,667		
Gymnasium	Contrast	13,500	1	13,500	3,682	,079
	Error	44,000	12	3,667		

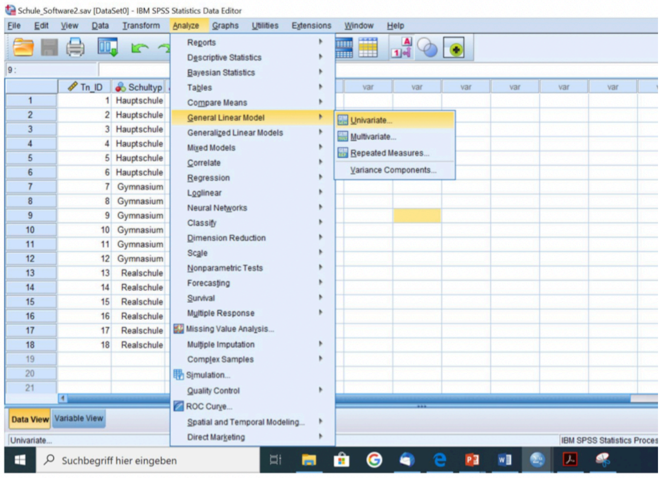
Each F tests the simple effects of Software within each level combination of the other effects shown. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

signifikanter Effekt der Lernsoftware in der Hauptschule!

kein signifikanter Effekt der Lernsoftware am Gymnasium!

wo in SPSS?

Tn_ID	Schultyp	Software	Kenntnisse
1	Hauptschule	A	8,00
2	Hauptschule	A	10,00
3	Hauptschule	A	12,00
4	Hauptschule	B	13,00
5	Hauptschule	B	13,00
6	Hauptschule	B	16,00
7	Gymnasium	A	19,00
8	Gymnasium	A	18,00
9	Gymnasium	A	20,00
10	Gymnasium	B	14,00
11	Gymnasium	B	15,00
12	Gymnasium	B	19,00
13	Realschule	A	8,00
14	Realschule	A	10,00
15	Realschule	A	12,00
16	Realschule	B	13,00
17	Realschule	B	13,00
18	Realschule	B	16,00



Beispiel:
die Datei Fugazi.sav im Ordner *SPSS-Dateien aus Field* enthält Daten aus einer fiktiven Studie, in der untersucht wird, ob sich der Musikgeschmack mit dem Alter verändert - jüngere (bis 40 Jahre) und ältere (ü40) TN haben dazu angegeben, wie gut ihnen die Musik von einer von 3unterschiedlichen Bands/Musikern gefällt - auf einer Skala von -100 bis 100,

a) führen Sie eine entsprechende 2-faktorielle Varianzanalyse durch und interpretieren Sie die Ergebnisse zu allen drei möglichen Effekten!
b) bestimmen und interpretieren Sie entsprechende Effektstärken!
c) erstellen Sie ein Interaktionsdiagramm!
d) untersuchen Sie die Interaktion weiter mittels simple effects!
e) was ergab Ihre Voraussetzungsprüfung?
f) wenden Sie einen geeigneten post-hoc Test auf den Faktor *Musik* an!
g) führen Sie zu Illustrationszwecken eine Analyse mit einem vordefinierten Kontrast für diesen Faktor durch!

Varianzanalyse mit abhängigen Stichproben

→ Grundidee aller inferenzstatistischen Verfahren für abhängige Stichpoben:
- die Fehlervariation geht zum Teil auf systematische Unterschiede zwischen Personen zurück → natürliche Variation,
- diese Variation zwischen Personen soll von der Fehlervariation getrennt und rausgerechnet werden → **größere Power!**

→ Experiment: den Teilnehmern werden einzeln und in randomisierter Reihenfolge Wörter präsentiert - diese werden in unterschiedlicher Häufigkeit wiederholt - die Tabelle zeigt die mittleren Häufigkeitsschätzungen pro Person für Wörter, die 2x, 4x und 6x präsentiert wurden,

Tabelle 15.8

Ergebnisse des Beispielexperiments zu Häufigkeitsschätzungen (fiktive Daten)

Teilnehmer	Häufigkeit			
	2	4	6	Zeilen-mittelwert
1	2,50	3,50	3,75	$\bar{x}_{\bullet 1} = 3{,}25$
2	2,00	2,75	4,25	$\bar{x}_{\bullet 2} = 3{,}00$
3	3,75	4,25	5,50	$\bar{x}_{\bullet 3} = 4{,}50$
4	3,25	4,50	5,75	$\bar{x}_{\bullet 4} = 4{,}50$
5	3,50	4,50	4,75	$\bar{x}_{\bullet 5} = 4{,}25$
Spalten-mittelwert	$\bar{x}_{1\bullet} = 3{,}00$	$\bar{x}_{2\bullet} = 3{,}90$	$\bar{x}_{3\bullet} = 4{,}80$	$\bar{\bar{x}} = 3{,}90$

„Personen-gebundene“ Variation

→ personengebundene Variation im vorher-nachher-Design
→ der einfachere Fall: within-Vergleich von 2Bedingungen → *t*-Test für abhängige Stichproben
- hier: Anzahl erinnerter Wörter in einer Liste mit nicht-assoziierten vs. assoziierten Wörtern,

Hypothetische Ergebnisse eines Gedächtnisexperiments im within-subjects Design

Teilnehmer	Kontroll-bedingung	Experimental-bedingung	Differenz
Teilnehmer 1	13	15	2
Teilnehmer 2	4	6	2
Teilnehmer 3	16	18	2
Teilnehmer 4	9	11	2
Teilnehmer 5	19	20	1
Mittelwert	*12,2*	*14*	

Niveau- Unterschiede: Teilnehmer 1 erinnert mehr Wörter als Teilnehmer

die Differenzen enthalten diese Niveau- Unterschiede **nicht** mehr!

→ Folge: weniger Variation in den Differenzen, mehr Variation in den Messwerten!

abhängige Daten

→ die Nicht-Berücksichtigung von Abhängigkeiten bei der Analyse führt immer zu einer oft drastischen α-Inflation!

→ das ist ein großer Fehler!

→ jeder Statistiker wird immer nach Abhängigkeiten in den Daten suchen

→ dringend bei der Studienplanung berücksichtigen!

→ within-Designs ermöglichen es, personengebundene Variation aus der Fehlervariation rauszurechnen

→ **sie erhöhen damit die Power!**

→ das ist ein wesentlicher Grund für die Verwendung solcher Designs!

personengebundene Variation in komplexeren Designs

Teilnehmer	Häufigkeit			
	2	4	6	Zeilen-mittelwert
1	2,50	3,50	3,75	$\bar{x}_{\bullet 1} = 3{,}25$
2	2,00	2,75	4,25	$\bar{x}_{\bullet 2} = 3{,}00$
3	3,75	4,25	5,50	$\bar{x}_{\bullet 3} = 4{,}50$
4	3,25	4,50	5,75	$\bar{x}_{\bullet 4} = 4{,}50$
5	3,50	4,50	4,75	$\bar{x}_{\bullet 5} = 4{,}25$
Spalten-mittelwert	$\bar{x}_{1\bullet} = 3{,}00$	$\bar{x}_{2\bullet} = 3{,}90$	$\bar{x}_{3\bullet} = 4{,}80$	$\bar{\bar{x}} = 3{,}90$

→ Personenmittelwerte/Zeilenmittelwerte abziehen!

Geringere Fehlervariation!

Teilnehmer	Häufigkeit			
	2	4	6	Zeilen-mittelwert
1	−0,75	0,25	0,50	$\bar{x}_{\bullet 1} = 0$
2	−1,00	−0,25	1,25	$\bar{x}_{\bullet 2} = 0$
3	−0,75	−0,25	1,00	$\bar{x}_{\bullet 3} = 0$
4	−1,25	0,00	1,25	$\bar{x}_{\bullet 4} = 0$
5	−0,75	0,25	0,50	$\bar{x}_{\bullet 5} = 0$
Spalten-mittelwert	$\bar{x}_{1\bullet} = -0{,}90$	$\bar{x}_{2\bullet} = 0$	$\bar{x}_{3\bullet} = 0{,}90$	$\bar{\bar{x}} = 0$

Keine Personengebundene Variation!

Unveränderte systematische Variation!

→ gewöhnliche, einfaktorielle VA mit $df_{inn} = (k-1) * (n-1)$ bringt korrektes Ergebnis!

Varianzanalyse mit abhängigen Stichproben

→ rechnerisches Vorgehen wie in zweifaktorieller Varianzanalyse!

→ Quadratsummen:

$$QS_{gesamt} = \sum_{j}^{k} \sum_{i}^{n} (x_{ji} - \bar{\bar{x}})^2 \qquad QS_{UV} = \sum_{j}^{k} n \cdot (\bar{x}_j - \bar{\bar{x}})^2$$

$$QS_{Pers} = \sum_{i}^{n} k \cdot (\bar{x}_i - \bar{\bar{x}})^2 \qquad QS_{UV \times Pers} = QS_{gesamt} - QS_{UV} - QS_{Pers}$$

j	Index der Stufen von uV
i	Index der TN
n	Anzahl der TN
	Mittelwert in Zelle jl
	Mittelwert der Messwerte in der Bedingung j → Spaltenmittelwert
	Mittelwert aller Messwerte in der Bedingung i → Zeilenmittelwert

Varianzschätzungen in der Varianzanalyse mit abhängigen Stichproben:

$$\hat{\sigma}^2_{UV} = \frac{QS_{UV}}{df_{UV}} \qquad df_{uV} = k - 1 \qquad \hat{\sigma}^2_{Fehler} = \frac{QS_{UV \times Pers}}{df_{UV \times Pers}} \qquad df_{uV\&Pers} = (k - 1)*(m - 1)$$

Prüfgröße:

$$F = \frac{\hat{\sigma}^2_{UV}}{\hat{\sigma}^2_{Fehler}}$$

ANOVA-Tabelle für die Ergebnisse aus dem Rechenbeispiel 15.2

Varianzquelle	*QS*	*df*	Varianzschätzung	*F*	*p*
Personen	6,225	4	1,556		
Häufigkeit	8,100	2	4,050	31,24	0,0002
Fehler	1,025	8	0,128		
gesamt	15,350	14			

→ wir könnten auch einen *F*-Wert für *Personen* bestimmen:

- prüft, ob sich Personen im Mittel unterscheiden - also, ob Niveauunterschiede existieren,
- meist zu offensichtlich, um es tatsächlich zu testen!

Effektstärken in der Varianzanalyse mit Messwiederholung

η²:
$$\eta^2 = \frac{QS_{UV}}{QS_{gesamt}}$$

η_p^2:
$$\eta_P^2 = \frac{QS_{UV}}{QS_{UV} + QS_{UV \times Pers}} \qquad \eta_P^2 = \frac{F \cdot df_{UV}}{F \cdot df_{UV} + df_{UV \times Pers}}$$

→ η_p ist mit dem η^2 aus einem between-subjects Design **nicht** vergleichbar!
  - → nach Möglichkeit beide Effektgrößen berichten!
  - → das Problem begründet ein eigenes Forschungsfeld, z.B. Olejnik & Algina, 2003 – *Generalized Eta Square*

→ auch hier könnte man eine korrigierte Schätzung des Effekts in der Population berechnen - ω^2,
  - → ziemlich mühselig
  - → sparen wir uns!

Power in der Varianzanalyse mit abhängigen Stichproben
Power-Tabelle für gewöhnliche einfaktorielle VA:

Power in der einfaktoriellen Varianzanalyse*

Teilnehmer pro Gruppe (n)	$\alpha = 0,01$			$\alpha = 0,05$			$\alpha = 0,10$		
	$\eta^2 = 0,01$ (klein)	$\eta^2 = 0,06$ (mittel)	$\eta^2 = 0,14$ (groß)	$\eta^2 = 0,01$ (klein)	$\eta^2 = 0,06$ (mittel)	$\eta^2 = 0,14$ (groß)	$\eta^2 = 0,01$ (klein)	$\eta^2 = 0,06$ (mittel)	$\eta^2 = 0,14$ (groß)
Drei Gruppen ($df_{zw} = 2$)									
5	0,01	0,03	0,07	0,06	0,11	0,21	0,11	0,19	0,33
10	0,02	0,06	0,21	0,07	0,20	0,44	0,13	0,30	0,58
20	0,03	0,17	0,54	0,10	0,37	0,78	0,17	0,50	0,86
30	0,04	0,30	0,79	0,12	0,54	0,93	0,20	0,66	0,96
40	0,05	0,43	0,92	0,15	0,68	0,98	0,24	0,78	0,99
50	0,06	0,56	0,97	0,18	0,78	0,99	0,27	0,86	*

→ **aber:**

- η_p anstelle von η^2 - da das partielle η_p größer ist, ist auch die Power größer!
- Teilnehmer in der Studie anstatt Teilnehmer pro Gruppe!

→ eine a priori Power-Analyse in der Varianzanalyse für abhängige Stichproben erfordert, dass man neben dem between-subjects-Effekt auch die personengebundene Varianz abschätzen kann,

→ sinnvoller Indikator der personengebundenen Varianz: Korrelation der Messwertreihen

→ bei perfekter Korrelation ist $\eta_p = 1$

→ bei einer Korrelation von 0 ist $\eta_p = \eta^2$ - die Daten sind unabhängig!

Anwendungsvoraussetzungen der Varianzanalyse für abhängige Stichproben

→ die aV muss mindestens Intervallskalenniveau besitzen,

→ **alle Stichproben** müssen aus normalverteilten Populationen stammen

→ und → Sphärizität!

↙

→ Varianzhomogenität: die Daten in **allen Bedingungen** stammen aus Populationen mit gleicher Varianz,

→ homogene Kovarianzen: die Kovarianzen zwischen **allen Bedingungen** sind in der Population gleich - Kovarianz = unstandardisierte Korrelation,

→ inhaltlich: die Abhängigkeit zwischen allen Bedingungen soll gleich groß sein,

→ pragmatisch ausreichend:

→ die Differenzen der Messwerte zwischen zwei Bedingungen haben stets die gleiche Varianz,

→ unser Beispiel:

Teilnehmer	Häufigkeit			
	2	4	6	Zeilen-mittelwert
1	2,50	3,50	3,75	$\bar{x}_{\bullet 1} = 3{,}25$
2	2,00	2,75	4,25	$\bar{x}_{\bullet 2} = 3{,}00$
3	3,75	4,25	5,50	$\bar{x}_{\bullet 3} = 4{,}50$
4	3,25	4,50	5,75	$\bar{x}_{\bullet 4} = 4{,}50$
5	3,50	4,50	4,75	$\bar{x}_{\bullet 5} = 4{,}25$
Spalten-mittelwert	$\bar{x}_{1\bullet} = 3{,}00$	$\bar{x}_{2\bullet} = 3{,}90$	$\bar{x}_{3\bullet} = 4{,}80$	$\bar{\bar{x}} = 3{,}90$

4 - 2	**6 - 2**	**6 - 4**
1	1,25	0,25
0,75	2,25	1,50
0,5	1,75	1,25
1,25	2,5	1,25
1	1,25	0,25
$\hat{\sigma}^2 = 0{,}08$	$\hat{\sigma}^2 = 0{,}33$	$\hat{\sigma}^2 = 0{,}36$

! **je stärker diese Varianzen sich unterscheiden, desto stärker ist Sphärizität verletzt!**

wie wird darauf geprüft?

→ generell: Verletzungen der Sphärizitätsannahme sind schwerwiegend und führen zu progressiven Entscheidungen,

- man sollte also dringend ermitteln, ob die Sphärizitätsannahme verletzt sein könnte,

→ Mauchly's Test
- Nullhypothese: die Varianzen der Differenzen sind alle gleich,
- signifikantes Ergebnis impliziert also Verletzung der Sphärizitätsannahme,

→ aber:
- übliches Powerproblem!
- within-Designs arbeiten oft mit kleinen Stichproben
- wir laufen also Gefahr, Voraussetzungsverletzungen zu übersehen
- kurz: sie wollen diesen Test nicht verwenden!

Sphärizitätsschätzungen

→ es gibt mehrere Methoden, um zu schätzen, wie stark die Sphärizitätsannahme verletzt ist:
- Greenhouse-Geiser ε
- Huynh-Feldt ε

→ maximum für beide Schätzungen: 1
- dieser Wert zeigt an, dass keine Abweichung von Sphärizität erkennbar ist,
- je niedriger der Wert, desto größer die Abweichung,
- Minimum für Greenhouse-Geiser: $1 / df_{uV}$ - SPSS: *lower bound*

was tun, wenn keine Sphärizität besteht?

→ die Sphärizitätsschätzungen ε können für eine Korrektur der Freiheitsgrade des *F*-Werts verwendet werden,

→ die Freiheitsgrade werden mit ε multipliziert
  → je stärker die Sphärizitätsannahme verletzt ist, desto stärker werden die Freiheitsgrade reduziert,
  → kleinere Freiheitsgrade bedeuten, dass der *p*-Wert wächst
    - die falsch-positiv Rate ist also wieder kontrolliert!

→ sie sollten immer eine solche Korrektur verwenden!

→ aber welche?
  → schwierige Frage...
  → typische Empfehlung:
  wenn Greenhouse-Geisser ε > 0,75, dann Huynh-Feldt
  → sonst: Greenhouse-Geisser

→ post-hoc Tests funktionieren bei Verletzungen der Sphärizitätsannahme meist nicht,

→ am ehesten: Bonferroni-Korrektur! - in SPSS nur über Tricks verfügbar...

Kontraste

→ wie gehabt…

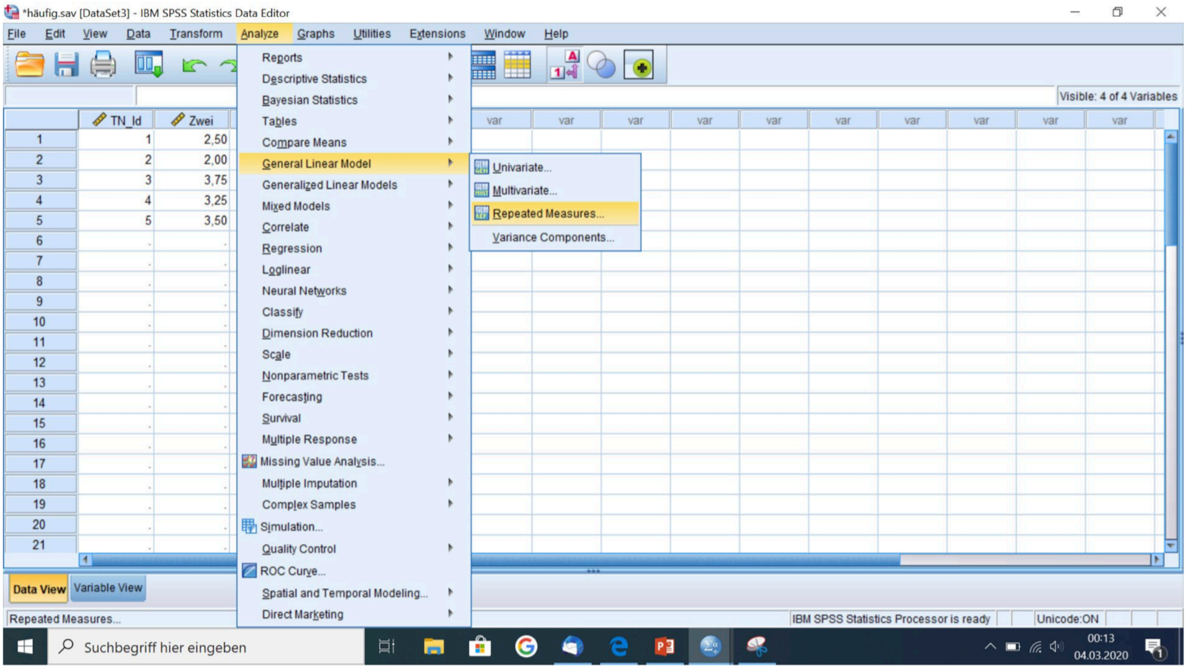
→ in SPSS stehen nur vordefinierte Kontraste zur Verfügung, z.B. Helmert-Kontrast: Vergleich jeder Bedingung mit der Kombination aus allen folgenden Bedingungen - also ein orthogonaler Kontrast,

→ eigene Kontraste kann man nur über die Syntax spezifizieren, wo in SPSS?

TN_Id	Zwei	Vier	Sechs
1	2,50	3,50	3,75
2	2,00	2,75	4,25
3	3,75	4,25	5,50
4	3,25	4,50	5,75
5	3,50	4,50	4,75

verändertes Datenformat beachten!

die uV ist **keine** gesonderte Variable!



ein Beispiel:
die Datei RovingEye.sav im Ordner *SPSS-Dateien aus Field* enthält Daten aus einer fiktiven Eye-Tracking-Studie, in der untersucht wurde, wie oft die 20TN unter Alkoholeinfluss in einem Nachtclub anderen Besuchern hinterherschauen - eine typische Field-Studie,
den TNn wurden an 4Abenden 1, 2, 3, und 4Bier *verabreicht*

a) prüfen Sie, ob die Menge der getrunkenen Biere einen Einfluss auf die Neigung, anderen Nachtclubbesuchern hinterherzuschauen, hat!
b) was ergab Ihre Sphärizitätsprüfung? wie interpretieren Sie die Ergebnisse?
c) bestimmen und interpretieren Sie entsprechende Effektstärken!
d) führen Sie zu Illustrationszwecken einen geeigneten post-hoc Test durch und wenden Sie einen Helmert-Kontrast auf die Daten an!
e) interpretieren Sie die Ergebnisse!
f) wie erklären sich die Unterschiede?

Tutorium - 12.Januar

1. zu welcher Analyse gehören diese Effektstärken, welche wird bei gleicher Datengrundlage größer sein und welche würde man in einer eigenen Analyse berichten?

$$\eta^2 = \frac{QS_{zw}}{QS_{Ges}} \qquad \omega^2 = \frac{QS_{zw} - df_{zw} \cdot \hat{\sigma}^2_{inn}}{QS_{ges} + \hat{\sigma}^2_{inn}}$$

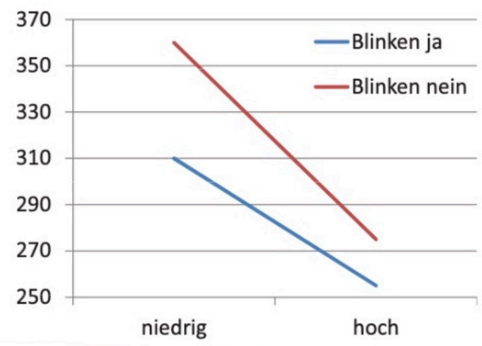
→ Effektstärken der Varianzanalyse - eta-Quadrat und Omega-Quadrat
→ eta-Quadrat ist größer,
→ Omega-Quadrat hat weniger Verzerrung als das (partielle) Eta-Quadrat, da es die Anzahl der Gruppen bei der Berechnung der Varianzaufklärung miteinbezieht - es kann daher auch für eine ANOVA berechnet werden, deren Zellengrößen unterschiedlich sind,

Effektstärke (auch Effektgröße) bezeichnet die Größe eines statistischen Effekts - sie kann zur Verdeutlichung der praktischen Relevanz von statistisch signifikanten Ergebnissen herangezogen werden - zur Messung werden verschiedene Effektmaße verwendet!

2. Aufstellen von Hypothesen anhand dieser Ergebnisse und Bestimmung der Kontrastgewichte!

In einer Studie zu Reaktionszeiten beim Autofahren wurden die Intensität und die Stetigkeit eines Warnsignals untersucht. Sie fanden folgende Reaktionszeiten:

		Intensität	
		niedrig	hoch
Blinken	ja	310	255
	nein	360	275



→ explorative Studie,
→ Kontraste waren hypothesengeleitet,
→ Umwandlung in datengeleitete Kontraste,
→ 2uVs - potentiell 2Haupteffekte,
→ Generierung von 3Hypothesen:
- Intensität hat einen Effekt auf Reaktionszeit - hoch vs. niedrig
- hohe Intensität verkürzt Reaktionszeit,
- wenn das Warnsignal blinkt, dann verkürzt sich die Reaktionszeit,
- bei niedriger Intensität wirkt sich das Blinken etwas stärker aus,

→ Haupteffekt von Intensität → hohe Intensität von Warnsignalen mit Blinken senkt die Reaktionszeit beim Autofahren,

→ Haupteffekt von Warnsignal → wenn das Warnsignal blinkt, dann verkürzt sich die Reaktionszeit,

→ Interaktion   Abstand der *BlinkenGruppen*   das Blinken verkürzt die Reaktionszeit mehr wenn die Intensität des Warnsignals niedrig ist gegenüber wenn sie hoch ist,

→ parallele Linien in Grafik → **keine Interaktion**!

→ Lambda-Gewichte festlegen:

Ermittlung der λ anhand der Werte aus Studie 1:

MW = 300

		Intensität niedrig	Intensität hoch
Blinken	ja	310	255
	nein	360	275

Differenz zu gefundenen Werten:

		Intensität niedrig	Intensität hoch
Blinken	ja	10	-45
	nein	60	-25

Aufhübschen der λ:

		Intensität niedrig	Intensität hoch
Blinken	ja	2	-9
	nein	12	-5

→ Lambda-Gewichte bekommt jede einzelne Gruppe - sie drücken aus, welche Gruppe einen höheren und welche Gruppe einen niedrigeren Wert bekommt - Umsetzung der Hypothesen,

→ **Kontrastanalysen** für unabhängige Stichproben
- wenn man drei oder mehr Gruppen miteinander hinsichtlich eines bestimmten Merkmals vergleichen möchte, und man hat auch eine Hypothese darüber, wie sie sich unterscheiden, dann verwendet man eine Kontrastanalyse,
- statistisches Verfahren zur Untersuchung gerichteter Hypothesen,
- Alternative zur normalen Varianzanalyse,
- Vorteile:
  - sie macht post-hoc-Tests überflüssig,
  - weist eine höhere Teststärke auf,
  - hat besser interpretierbare Effektgrößen,
  - ermöglicht den Vergleich mehrerer verschiedener Hypothesen,
- Berechnung:
  - ähnlich der Varianzanalyse - auch hier werden Varianzen miteinander verglichen - die Fehlervarianz ist dabei identisch mit der Fehlervarianz der ANOVA, und die Varianz zwischen den Gruppen wird ersetzt durch die Varianz des Kontrastes,

→ die einzelnen Lambdas **müssen** 0 ergeben! wie?

→ wie?

- durch Mittelwertzentrierung,

		Intensität	
		niedrig	hoch
Blinken	ja	310	255
	nein	360	275

↙

310 + 360 + 255 + 275 = 1200 : 4 = **300**

↙

310 - 300/ 360 - 300/ 255 - 300/ 275 - 300

↙

		Intensität	
		niedrig	hoch
Blinken	ja	10	-45
	nein	60	-25
		70	-70

= **0**

: 5

↓

		Intensität	
		niedrig	hoch
Blinken	ja	2	-9
	nein	12	-5

↙

- diese Kontraste drücken die Hypothesen aus - auf eine exaktere Weise

3. Berechnung der Varianz und statistischer Test

$$QS_{Kontrast} = \frac{\left(\sum_{j}^{4} \lambda_j \cdot \bar{x}_j\right)^2}{\sum_{j}^{4} \frac{\lambda_j^2}{n_j}}$$

- ist der Kontrast signifikant?

- Mittelwert * Lambda
- alles aufaddieren dann quadrieren

- F-Test - man teilt eine Art von Zwischengruppen (ergänzt durch die Lambda-gewichte) durch die Varianz in den Gruppen

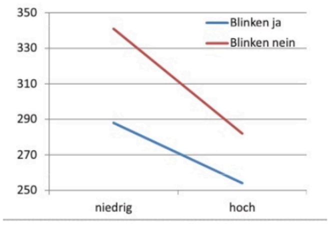
$$F_{Kontrast}(1,12) = \frac{\hat{\sigma}^2_{Kontrast}}{\hat{\sigma}^2_{inn}}$$

- Varianz$_{zw}$ analog zu Varianz$_{Kontraste}$

- je höher der F-Wert desto signifikanter

→ Beispielberechnung:

		Intensität niedrig	Intensität hoch
Blinken	ja	288	254
	nein	341	282



$$QS_{Kontrast} = \frac{\left(\sum_j^4 \lambda_j \cdot \bar{x}_j\right)^2}{\sum_j^4 \frac{\lambda_j^2}{n_j}} = \frac{(2\cdot 288 - 9\cdot 254 + 12\cdot 341 - 5\cdot 282)^2}{\frac{2^2}{4} + \frac{(-9)^2}{4} + \frac{12^2}{4} + \frac{(-5)^2}{4}} = \frac{944784}{63.5} \approx 14878{,}488$$

$$F_{Kontrast}(1,12) = \frac{\hat{\sigma}^2_{Kontrast}}{\hat{\sigma}^2_{inn}} = \frac{14878{,}488}{915} = 16{,}26 \qquad F_{krit}(1,12) = 9{,}33$$

→ F-Verteilung endet bei 0 und ist nach oben unendlich weit offen,

4. was sind mögliche Gründe dafür, dass WissenschaftlerInnen auf die Idee kommen, p-Hacking zu betreiben?
   - ! **Daten werden so lange erhoben oder ausgewertet bis ehemals nicht signifikante Ergebnisse signifikant werden - der p-Wert wird gehackt, künstlich unter die 5%-Grenze gedrückt - m.H. von Metaanalysen ist es möglich, p-Hacking aufzudecken,**
   - das liegt z.T. an der Anreizstruktur in der Wissenschaft, dass häufiger signifikante Ergebnisse veröffentlicht werden,
   - Journal-Struktur,
   - Publikationsdruck,
   - Unwissenheit,
   - Drittmittelprojekte,

5. was sind mögliche Wege, um Probleme wie HARKing und p-Hacking einzudämmen?
   - Präregistrierung,
   - registered report,
   - Probleme identifizieren - RPP, SSRP, many labs, …
   - strukturierte Interessensvertretung - open science Bewegung

6.

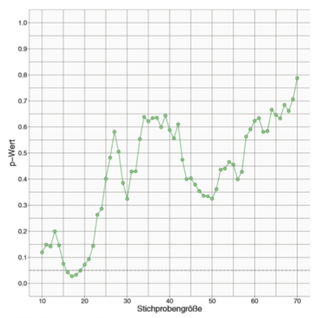


Abbildung 20.2: Simulation von *p*-Werten aus einer Studie mit falscher Alternativhypothese. In der Studie wird zunächst mit zehn Teilnehmern ein *t*-Test durchgeführt. Anschließend wird jeweils ein Teilnehmer pro Gruppe hinzugefügt und der Test wiederholt. Die gepunktete Linie bezeichnet die Signifikanzgrenze von $\alpha = 5\%$ (modifiziert nach Simmons et al., 2011, Fig. 2, S. 4).

dance of the p-value und *optional stopping*

www.shinyapps.org/apps/p-hacker

7. wie kann man p-hacking beschreiben?
   → Anpassung einiger Kennwerte wie **α**, damit p-Wert signifikant wird,
   → Manipulation der Daten, um Signifikanz herzustellen, die eigentlich nicht vorhanden ist,
   → selektive Ergebnisse berichten,
     ! **Daten werden so lange erhoben oder ausgewertet bis ehemals nicht signifikante Ergebnisse signifikant werden - der p-Wert wird gehackt, künstlich unter die 5%-Grenze gedrückt - m.H. von Metaanalysen ist es möglich, p-Hacking aufzudecken,**

8. Möglichkeiten, um p-hacking zu betreiben:
   → weitere Erhöhung der Power: die Wahrscheinlichkeit ein signifikantes Ergebnis zu finden steigt auch dann, wenn es tatsächlich keinen Effekt gibt → *optional stopping:* Vergrößerung der Stichprobe, wenn nicht signifikante Ergebnisse beobachtet wurden - Ende der Studie, sobald Signifikanz auftritt
   → Durchführung mit 3Bedingungen, wovon eine Bedingung im Bericht unerwähnt bleibt,
   → neben uV und aV auch Erhebung nicht theoriegeleiteter Daten, wie z.B. Alter, Gecshlecht, etc

9. kurze Zusammenfassung des HARKing
   → ***H**ypothesizing **A**fter the **R**esults are **K**nown* → das Aufstellen von Hypothesen nach dem Bekanntwerden von Ergebnissen
   → Beispiel: ein Sportschütze hat es sich zur Aufgabe gemacht, eine weit entfernte Zielscheibe genau in der Mitte zu treffen - einige Beobachter haben sich um ihn geschart - der Sportschütze tut sein Vorhaben aber nicht genau kund (ich stelle die Hypothese auf, dass es mir gelingt, aus dieser weiten Distanz diese Zielscheibe zu treffen), sondern signalisiert den Zuschauern nur, dass er schießen wolle - sein erster Schuss geht daneben und er trifft den äußeren Ring der Zielscheibe - danach wendet sich der Schütze an die Schaulustigen und behauptet: genau diesen Punkt am Rand wollte ich treffen - damit ist belegt, wie gut ich als Bogenschütze bin…
   → Sammlung großer Datenmengen und Durchführung zahlreicher Signifikanztests - für eine beobachtetes signifikantes Resultat wird im Nachhinein nach einer passenden Hypothese gesucht und das gesamte Vorgehen als Test dieser Hypothese ausgegeben,

10. eine nach außen gerichtete für 12min andauernde Aktivität (TV, lesen) auszuüben wird als angenehmer bewertet als sich 12min einer *Denkperiode* hinzugeben, in der man sich einfach mit den eigenen Gedanken unterhält,
    würden sich Vpn mehr amüsieren, wenn sie etwas zu tun hätten?
    30TN,
    mehr Genuss bei externen Aktivitäten, Konzentration war einfacher möglich, Gedanken wanderten weniger
    - hat die Replikation einen statistisch signifikanten Effekt in die gleiche Richtung gefunden wie die ursprüngliche Studie?
      → JA - gleiches Ergebnis und ungefähr gleiche Effektstärke,
      → Menschen würden sogar lieber Elektroschocks über sich ergehen lassen, als sich 12min mit ihren Gedanken zu beschäftigen,

11. https://80000hours.org/psychology-replication-quiz-mobile/

a. es scheint, als ob nur etwas mehr als die Hälfte der psychologischen Ergebnisse in Top-Journalen korrekt sind,

b. herauszufinden, was in den Sozialwissenschaften wahr ist, kann daher schwierig sein - aus diesem Grund führen wir eingehende Überprüfungen von Themen durch, bei denen es viele zwielichtige Studien gibt und Menschen Fehlinformationen verbreiten, sodass Sie dies nicht tun müssen - z.B. haben wir über 60Studien darüber gesichtet, was einen Traumjob ausmacht, alles gelesen, was es darüber zu lesen gab, ob Geld glücklich macht, und festgestellt, dass die herkömmliche Meinung über Stress größtenteils falsch ist,

c. aus irgendeinem Grund haben die Leute viel bessere Intuitionen über diese Experimente als darüber, welche staatlichen Sozialprogramme funktionieren und welche nicht - um zu sehen, ob Sie das besser können als der Zufall, versuchen Sie unser viel schwierigeres Quiz: die guten, die schlechten und die unwirksamen: Sozialprogramme in Amerika,

d. Vorhersagemärkte und Expertenumfragen waren ziemlich gut darin, herauszufinden, welche Ergebnisse legitim waren und welche nicht - wir haben über die Bedeutung und das enorme Potenzial zur Verbesserung der institutionellen Entscheidungsfindung unter Verwendung dieser strukturierten Tools geschrieben - dies kann u.a. verhindern, dass ungerechte Kriege, riesige Geldbeträge für vorhersehbar schlechte Zuschüsse verschwendet werden und Menschen durch schädliche medizinische Behandlungen sterben - wir haben einen der Pioniere dieser Art von Psychologie in unserem Podcast interviewt,

e. es zeigt die Ernsthaftigkeit der sozialwissenschaftlichen Replikationskrise - sie denken vielleicht, dass Sie eine genaue statistische Analyse benötigen, um herauszufinden, welche Artikel falsch sind, aber für die meisten Menschen stellt sich heraus, dass *wenn die Behauptung des Artikels keinen intuitiven Sinn ergibt, der Artikel falsch ist* eine gute Abkürzung dafür ist, dass es zu 70% funktioniert - wenn Sie eher Ihrem Bauchgefühl folgen als den Artikeln in den besten Zeitschriften zu glauben, ist dies ein sehr enttäuschender Zustand und deutet darauf hin, dass wir viel Geld für unzuverlässige Forschung verschwenden - wir haben über den hohen potenziellen Wert von Karrieren geschrieben, die mit wirkungsvoller und ehrlicher Forschung verbracht werden, insbesondere in der Wissenschaft, aber das System sieht ein bisschen kaputt aus,

f. sogar bei erfolgreichen Replikationen waren die geschätzten Effektstärken kleiner als in der ursprünglichen Studie - bei den 13 replizierten Studien … betrugen die Effektstärken nur etwa 75 % derjenigen in den ursprünglichen Studien - dies liefert eine Schätzung der Überschätzung der Wirkung Größen von True Positives - unter den erfolglosen Wiederholungen gab es im Wesentlichen keine Beweise für den ursprünglichen Befund: die durchschnittliche relative Effektgröße lag sehr nahe bei Null