

# Statistik für Digital Humanities

Vergleich mehrerer Mittelwerte: ANOVA & ANCOVA

Dr. Jochen Tiepmar

Institut für Informatik  
Computational Humanities  
**Universität Leipzig**

15. Juni 2020

[Letzte Aktualisierung: 14/06/2020, 22:13]

- 1 Motivation
- 2 ANOVA
- 3 Lokalisierung der Unterschiede
- 4 Effektstärke
- 5 ANCOVA

- Sind 2 Mittelwerte signifikant verschieden?

## 2 Arten von t-Tests

- Unabhängiger t-Test → Gruppendesign (verschiedene Probanden)  
Auch Independent-Measures oder Independent Samples t-Test
- Abhängiger t-Test → Messwiederholungsdesign (gleiche Probanden)  
Auch Matched Pairs oder Paired Samples t-Test

- Sind 2 Mittelwerte signifikant verschieden?

## 2 Arten von t-Tests

- Unabhängiger t-Test → Gruppendesign (verschiedene Probanden)  
Auch Independent-Measures oder Independent Samples t-Test
- Abhängiger t-Test → Messwiederholungsdesign (gleiche Probanden)  
Auch Matched Pairs oder Paired Samples t-Test

Warum nicht mehrere t-Tests durchführen um mehrere Mittelwerte zu vergleichen?

# Familienbezogener / Experimentbezogener Fehler

		Wirklichkeit	
		$H_0$ ist wahr	$H_1$ ist wahr
Entscheidung des Tests	für $H_0$	Spezifität True Positive Wahrscheinlichkeit: $1 - \alpha$	<b>Fehler 2. Art</b> False Negative Wahrscheinlichkeit: $\beta$
	für $H_1$	<b>Fehler 1. Art</b> False Positive Wahrscheinlichkeit: $\alpha$	Sensitivität, Trennschärfe True Negative Wahrscheinlichkeit: $1 - \beta$

		Wirklichkeit	
		H <sub>0</sub> ist wahr	H <sub>1</sub> ist wahr
Entscheidung des Tests	für H <sub>0</sub>	Spezifität True Positive Wahrscheinlichkeit: $1 - \alpha$	<b>Fehler 2. Art</b> False Negative Wahrscheinlichkeit: $\beta$
	für H <sub>1</sub>	<b>Fehler 1. Art</b> False Positive Wahrscheinlichkeit: $\alpha$	Sensitivität, Trennschärfe True Negative Wahrscheinlichkeit: $1 - \beta$

- Problem: familienbezogene / experimentbezogene Fehlerrate  $1 - (1 - \alpha)^k$  mit  $k = \text{Anzahl der Kombinationen}$   $\alpha$  ist die Typ 1 Fehlerwahrscheinlichkeit
- vergrößert sich mit jedem zusätzlichen Test
  - 3 Gruppen  $\rightarrow$  3 Tests  $\rightarrow$  Wahrscheinlichkeit für Typ 1 Fehler  $1 - 0.95^3 = 0.143$
  - 5 Gruppen:  $\rightarrow$  10 Kombinationen  $\rightarrow$  Wahrscheinlichkeit für Typ 1 Fehler  $1 - 0.95^{10} = 0.401$ , also 40%

## 1 Motivation

## 2 ANOVA

- Berechnung
- Beispiel
- Robustheit

## 3 Lokalisierung der Unterschiede

- Geplante Kontrastierung
- Post Hoc Tests

## 4 Effektstärke

## 5 ANCOVA

- Voraussetzungen
- Berechnung

- $H_0 =$  Alle Mittelwerte sind statistisch ähnlich
- Omnibus Test: Zeigt Effekt an, aber nicht wo er passiert ist
  - $\overline{X_1} = \overline{X_2} \neq \overline{X_3}$
  - $\overline{X_1} \neq \overline{X_2} = \overline{X_3}$
  - $\overline{X_1} \neq \overline{X_2} \neq \overline{X_3}$

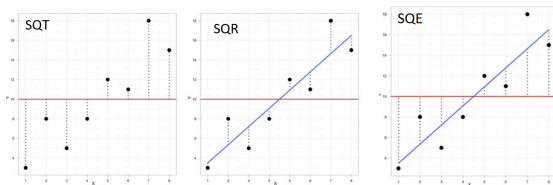


- $H_0$  = Alle Mittelwerte sind statistisch ähnlich
- Omnibus Test: Zeigt Effekt an, aber nicht wo er passiert ist
  - $\overline{X_1} = \overline{X_2} \neq \overline{X_3}$
  - $\overline{X_1} \neq \overline{X_2} = \overline{X_3}$
  - $\overline{X_1} \neq \overline{X_2} \neq \overline{X_3}$
- Erstmal nur unabhängig, nächste Woche besprechen wir u.a. die abhängige Variante (Messwiederholung)

- $H_0$  = Alle Mittelwerte sind statistisch ähnlich
- Omnibus Test: Zeigt Effekt an, aber nicht wo er passiert ist
  - $\overline{X_1} = \overline{X_2} \neq \overline{X_3}$
  - $\overline{X_1} \neq \overline{X_2} = \overline{X_3}$
  - $\overline{X_1} \neq \overline{X_2} \neq \overline{X_3}$
- Erstmal nur unabhängig, nächste Woche besprechen wir u.a. die abhängige Variante (Messwiederholung)
- Berechnung:
  - $F = \frac{MQE}{MQR} = \frac{\text{Systematische Variation}}{\text{Unsystematische Variation}}$
  - $MQE = \frac{SQE}{k-1}$
  - $MQR = \frac{SQR}{n-k}$
  - $k$  = Anzahl der Gruppen

- Regression erlaubt Abschätzen von  $Y$  für neue Werte aus  $X$
- Regressionsformel  $\hat{Y} = (b_0 + b_1 * X_1 + b_2 * X_2 + \dots)$
- Als statistisches Modell hat eine Regressionslinie eine Fitness
  - Residuenquadratsumme, Erklärte Quadratsumme,  $R^2 =$  Verhältnis beider
  - F-Test möglich um Modell zu bewerten
  - t-Test möglich um Einflußstärke des Prädiktors zu bewerten

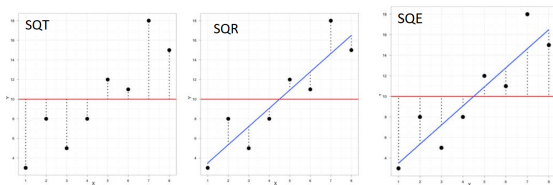
# Wiederholung Abweichungsquadrate bei Regression



Abstände von Regression zu Beobachtung sind Residuen (Residuum)

- **Quadratsumme der Abweichungen**  $SQT = \sum (y_i - \bar{y})^2$
- **Residuenquadratsumme**  $SQR = \sum (y_i - \hat{y}_i)^2$
- **Erklärte Quadratsumme**  $SQE = \sum (\bar{y} - \hat{y}_i)^2$

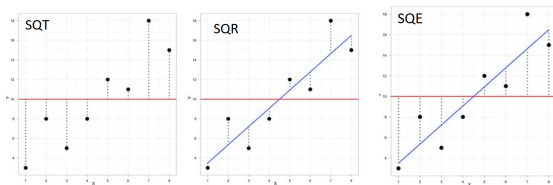
# Wiederholung Abweichungsquadrate bei Regression



Abstände von Regression zu Beobachtung sind Residuen (Residuum)

- **Quadratsumme der Abweichungen**  $SQT = \sum (y_i - \bar{y})^2$
- **Residuenquadratsumme**  $SQR = \sum (y_i - \hat{y}_i)^2$
- **Erklärte Quadratsumme**  $SQE = \sum (\bar{y} - \hat{y}_i)^2$
- $MQ_x$  = Mittelwert der Quadrate von  $x$
- $MQE = \frac{SQE}{\text{Variablenanzahl}}$
- $MQR = \frac{SQR}{\text{Beobachtungen} - \text{Regressionskoeffizienten}}$

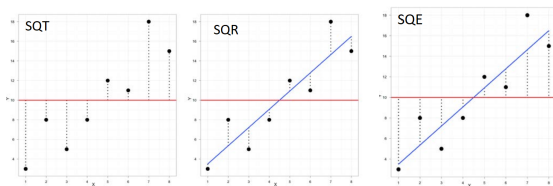
# Wiederholung Abweichungsquadrate bei Regression



Abstände von Regression zu Beobachtung sind Residuen (Residuum)

- **Quadratsumme der Abweichungen**  $SQT = \sum (y_i - \bar{y})^2$
- **Residuenquadratsumme**  $SQR = \sum (y_i - \hat{y}_i)^2$
- **Erklärte Quadratsumme**  $SQE = \sum (\bar{y} - \hat{y}_i)^2$
- $MQ_x =$  Mittelwert der Quadrate von  $x$
- $MQE = \frac{SQE}{\text{Variablenanzahl}}$
- $MQR = \frac{SQR}{\text{Beobachtungen} - \text{Regressionskoeffizienten}}$
- $F = \frac{MQE}{MQR}$

# Wiederholung Abweichungsquadrate bei Regression

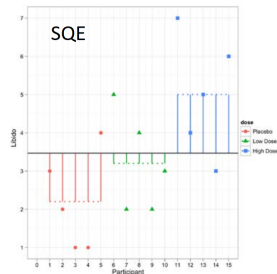
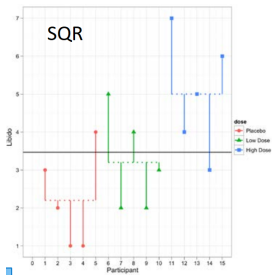
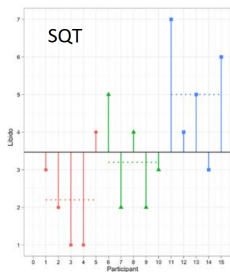


Abstände von Regression zu Beobachtung sind Residuen (Residuum)

- **Quadratsumme der Abweichungen**  $SQT = \sum (y_i - \bar{y})^2$
- **Residuenquadratsumme**  $SQR = \sum (y_i - \hat{y}_i)^2$
- **Erklärte Quadratsumme**  $SQE = \sum (\bar{y} - \hat{y}_i)^2$
- $MQ_x =$  Mittelwert der Quadrate von  $x$
- $MQE = \frac{SQE}{\text{Variablenanzahl}}$
- $MQR = \frac{SQR}{\text{Beobachtungen} - \text{Regressionskoeffizienten}}$
- $F = \frac{MQE}{MQR}$
- Was wir bei der Evaluation der Regression gemacht haben, ist eine ANOVA auf dem Linearen Modell

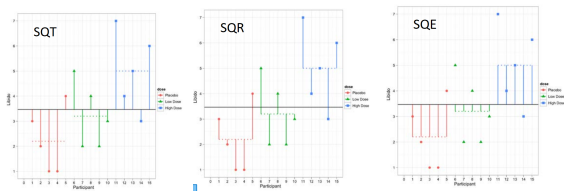
# Abweichungsquadrate bei mehreren Mittelwerten

Jetzt die eigentliche ANOVA:





# Abweichungsquadrate bei mehreren Mittelwerten



- **Quadratsumme der Abweichungen**

$$SQT = \sum (x_i - \text{Grand Mean})^2 \text{ Im Buch } SS_T$$

- **Residuenquadratsumme**

$$SQR = \sum (x_i - \overline{group})^2 = \sum s_{group}^2 * (n_{group} - 1) \text{ Im Buch } SS_R$$

- **Erklärte Quadratsumme**

$$SQE = \sum n_{group} * (\overline{group} - \text{Grand Mean})^2 \text{ Im Buch } SS_M$$

$$SQT = SQR + SQE$$

- $F = \frac{MQE}{MQR} = \frac{\text{Systematische Variation}}{\text{Unsystematische Variation}}$
- $MQE = \frac{SQE}{k-1}$
- $MQR = \frac{SQR}{n-k}$
- $k = \text{Anzahl der Gruppen}$

Interpretation:

- $F = \frac{MQE}{MQR} = \frac{\text{Systematische Variation}}{\text{Unsystematische Variation}}$
- $MQE = \frac{SQE}{k-1}$
- $MQR = \frac{SQR}{n-k}$
- $k = \text{Anzahl der Gruppen}$

Interpretation:

- Je höher  $F$ , desto besser das Modell
- $F < 1 \rightarrow \text{Unsystematische Variation ist größer als Systematische Variation}$
- $F < F_{kr} \text{ aus Tabelle} \rightarrow H_0$

- $F = \frac{MQE}{MQR} = \frac{\text{Systematische Variation}}{\text{Unsystematische Variation}}$
- $MQE = \frac{SQE}{k-1}$
- $MQR = \frac{SQR}{n-k}$
- $k = \text{Anzahl der Gruppen}$

Interpretation:

- Je höher  $F$ , desto besser das Modell
- $F < 1 \rightarrow$  Unsystematische Variation ist größer als Systematische Variation
- $F < F_{kr}$  aus Tabelle  $\rightarrow H_0$  kann nicht abgewiesen werden  $\rightarrow$

- $F = \frac{MQE}{MQR} = \frac{\text{Systematische Variation}}{\text{Unsystematische Variation}}$
- $MQE = \frac{SQE}{k-1}$
- $MQR = \frac{SQR}{n-k}$
- $k = \text{Anzahl der Gruppen}$

Interpretation:

- Je höher  $F$ , desto besser das Modell
- $F < 1 \rightarrow$  Unsystematische Variation ist größer als Systematische Variation
- $F < F_{kr}$  aus Tabelle  $\rightarrow H_0$  kann nicht abgewiesen werden  $\rightarrow$  Alle Mittelwerte sind statistisch ähnlich
- $df(\text{Numerator}) = k - 1$
- $df(\text{Denominator}) = n - k$

Zeichenlänge des Dokumententitels pro Autor

Autor 1	Autor 2	Autor 3
30	40	10
35	35	25
45	50	5

Zeichenlänge des Dokumententitels pro Autor

Autor 1	Autor 2	Autor 3
30	40	10
35	35	25
45	50	5

$\overline{Autor1} = 36.667$ ,  $\overline{Autor2} = 41.667$ ,  $\overline{Autor3} = 13.333$ ,  $G. Mean = 30.555$

Zeichenzahl des Dokumententitels pro Autor

Autor 1	Autor 2	Autor 3
30	40	10
35	35	25
45	50	5

$\overline{Autor1} = 36.667, \overline{Autor2} = 41.667, \overline{Autor3} = 13.333, G. Mean = 30.555$

$SQE = \sum n_{group} * (\overline{group} - G. Mean)^2 = 3 * (36.667 - 30.555)^2 + 3 * (41.667 - 30.555)^2 + 3 * (13.333 - 30.555)^2 = 1372.291$

$SQR = \sum (x_i - \overline{group})^2 = 116.667 + 116.667 + 216.667 = 450$



Zeichenlänge des Dokumententitels pro Autor

Autor 1	Autor 2	Autor 3
30	40	10
35	35	25
45	50	5

$\overline{Autor1} = 36.667$ ,  $\overline{Autor2} = 41.667$ ,  $\overline{Autor3} = 13.333$ ,  $G. Mean = 30.555$

$$SQE = \sum n_{group} * (\overline{group} - G. Mean)^2 = 3 * (36.667 - 30.555)^2 + 3 * (41.667 - 30.555)^2 + 3 * (13.333 - 30.555)^2 = 1372.291$$

$$SQR = \sum (x_i - \overline{group})^2 = 116.667 + 116.667 + 216.667 = 450$$

$$MQE = \frac{SQE}{k-1} = \frac{1372.291}{2} = 686.145$$

$$MQR = \frac{SQR}{n-k} = \frac{449.998}{6} = 75$$

Zeichenlänge des Dokumententitels pro Autor

Autor 1	Autor 2	Autor 3
30	40	10
35	35	25
45	50	5

$\overline{Autor1} = 36.667$ ,  $\overline{Autor2} = 41.667$ ,  $\overline{Autor3} = 13.333$ ,  $G. Mean = 30.555$

$$SQE = \sum n_{group} * (\overline{group} - G. Mean)^2 = 3 * (36.667 - 30.555)^2 + 3 * (41.667 - 30.555)^2 + 3 * (13.333 - 30.555)^2 = 1372.291$$

$$SQR = \sum (x_i - \overline{group})^2 = 116.667 + 116.667 + 216.667 = 450$$

$$MQE = \frac{SQE}{k-1} = \frac{1372.291}{2} = 686.145$$

$$MQR = \frac{SQR}{n-k} = \frac{449.998}{6} = 75$$

$$F = \frac{MQE}{MQR} = \frac{686.145}{75} = 9.149 \quad F > F_{kr95}(2, 6) = 5.79 \rightarrow H_0$$

Zeichenlänge des Dokumententitels pro Autor

Autor 1	Autor 2	Autor 3
30	40	10
35	35	25
45	50	5

$\overline{Autor1} = 36.667$ ,  $\overline{Autor2} = 41.667$ ,  $\overline{Autor3} = 13.333$ ,  $G. Mean = 30.555$

$$SQE = \sum n_{group} * (\overline{group} - G. Mean)^2 = 3 * (36.667 - 30.555)^2 + 3 * (41.667 - 30.555)^2 + 3 * (13.333 - 30.555)^2 = 1372.291$$

$$SQR = \sum (x_i - \overline{group})^2 = 116.667 + 116.667 + 216.667 = 450$$

$$MQE = \frac{SQE}{k-1} = \frac{1372.291}{2} = 686.145$$

$$MQR = \frac{SQR}{n-k} = \frac{449.998}{6} = 75$$

$$F = \frac{MQE}{MQR} = \frac{686.145}{75} = 9.149 \quad F > F_{kr95}(2, 6) = 5.79 \rightarrow H_0 \text{ kann abgewiesen werden} \rightarrow$$

Zeichenlänge des Dokumententitels pro Autor

Autor 1	Autor 2	Autor 3
30	40	10
35	35	25
45	50	5

$\overline{Autor1} = 36.667$ ,  $\overline{Autor2} = 41.667$ ,  $\overline{Autor3} = 13.333$ ,  $G. Mean = 30.555$

$$SQE = \sum n_{group} * (\overline{group} - G. Mean)^2 = 3 * (36.667 - 30.555)^2 + 3 * (41.667 - 30.555)^2 + 3 * (13.333 - 30.555)^2 = 1372.291$$

$$SQR = \sum (x_i - \overline{group})^2 = 116.667 + 116.667 + 216.667 = 450$$

$$MQE = \frac{SQE}{k-1} = \frac{1372.291}{2} = 686.145$$

$$MQR = \frac{SQR}{n-k} = \frac{449.998}{6} = 75$$

$$F = \frac{MQE}{MQR} = \frac{686.145}{75} = 9.149 \quad F > F_{kr95}(2, 6) = 5.79 \rightarrow H_0 \text{ kann}$$

abgewiesen werden  $\rightarrow$  Es gibt signifikante Unterschiede in den Mittelwerten

```
titlelength<-c(30,35,45,40,35,50,10,25,5)
group<-gl(3,3,labels=c("autor1","autor2","autor3"))
df<-data.frame(group,titlelength)
```

```
anovamodel<-aov(titlelength~group, data=df)
summary(anovamodel)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	1372	686.1	9.148	0.0151 *
Residuals	6	450	75.0		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

$Pr(> F) < 0.05 \rightarrow H_0$  kann abgewiesen werden

- ANOVA ist grundsätzlich erstmal parametrisch
- Wird oft als robust angesehen. . .
- . . . aber Robustheit in vielen Fällen nicht untersucht und viele Hinweise dagegen
- Bei gleichen Gruppengrößen eher robust.
- Bei heterogenen Varianzen kann Welchs  $F$  weiterhelfen  
Welch, B.L. (1951): *On the comparison of several mean values: An alternative approach*
- Ausführlicher im Begleitmaterial "Andy Field - Robustheit von Anova" (Moodle)

- 1 Motivation
- 2 ANOVA
  - Berechnung
  - Beispiel
  - Robustheit
- 3 Lokalisierung der Unterschiede
  - Geplante Kontrastierung
  - Post Hoc Tests
- 4 Effektstärke
- 5 ANCOVA
  - Voraussetzungen
  - Berechnung

Wie finden wir jetzt heraus, welche Gruppen sich unterscheiden ohne die familienbezogene Fehlerrate zu stark ansteigen zu lassen?



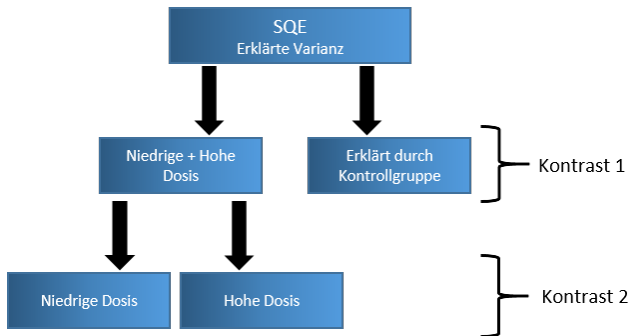
Wie finden wir jetzt heraus, welche Gruppen sich unterscheiden ohne die familienbezogene Fehlerrate zu stark ansteigen zu lassen?

- Geplante Kontrastierung: Spezifische Hypothese vorhanden
- Post Hoc Tests: Keine spezifische Hypothese vorhanden

# Geplante Kontrastierung

	Viagra	Libido		Viagra	Libido		Viagra	Libido
1	1	3	6	2	5	11	3	7
2	1	2	7	2	2	12	3	4
...			...			...		

Viagra: 1 = Kontrollgruppe, 2 = niedrige Dosierung, 3 = hohe Dosierung

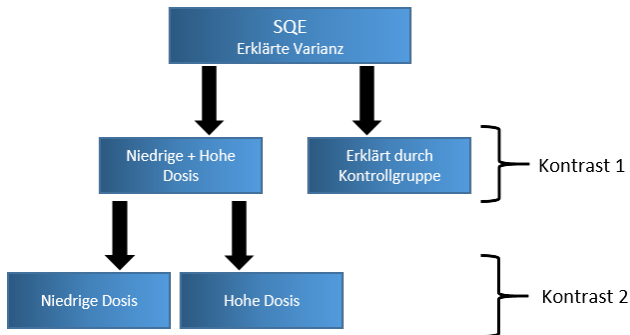


- Wiederholter Vergleich zweier Variationen
- Hierarchisches Vorgehen
- Gezieltes Annähern anhand der Hypothese , die vorher festgelegt werden muss!

- Wiederholter Vergleich zweier Variationen
- Hierarchisches Vorgehen
- Gezieltes Annähern anhand der Hypothese , die vorher festgelegt werden muss!

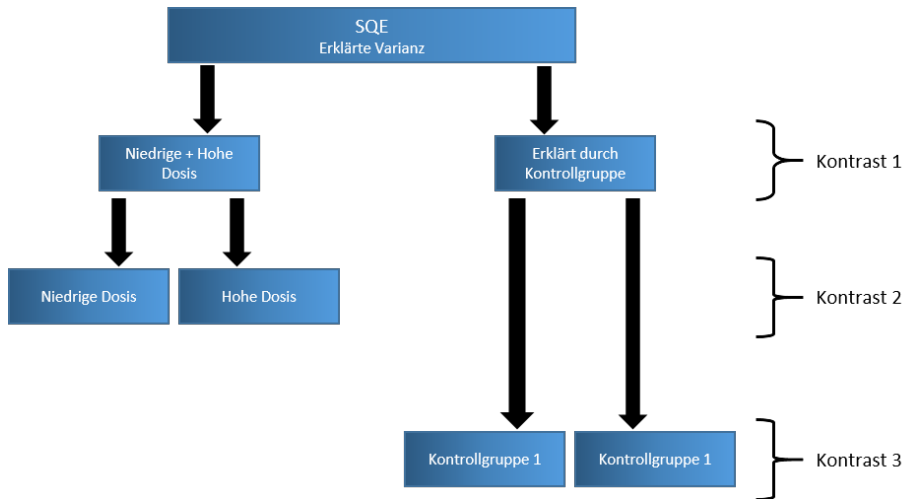
## 3 Regeln:

- Variation der Kontrollgruppen sollten gegen Variation der Experimentgruppen kontrastiert werden
- Jeder Kontrast darf nur 2 Variationen vergleichen
- Jede Gruppe darf nur Teil eines Kontrasts sein



- Auf übergeordneten Ebenen Vergleichswerte für zusammengesetzte Gruppen neu berechnen
  - bspw. Mittelwerte für Niedrige + Hohe Dosis
- Kontrastierung mit Multipler Regression mittels Dummy Variablen

# Kontraste bei 4 Gruppen (Bsp: 2 Kontrollgruppen)



# Berechnung Orthogonaler Kontraste

Schritt 1: Tabelle aufstellen

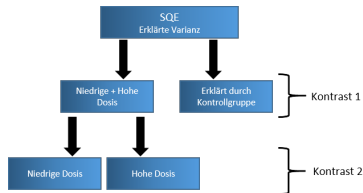
Schritt 2:

Schritt 3:

Schritt 4:

(Schritt 5:)

Gruppe	Kontrast 1	Kontrast 2
Placebo		
niedrig		
hoch		



# Berechnung Orthogonaler Kontraste

Schritt 1: Tabelle aufstellen

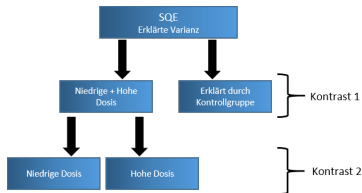
Schritt 2: Unbeteiligte Gruppen  $\rightarrow 0$

Schritt 3: Linke Gruppe +, Rechte Gruppe -

Schritt 4:

(Schritt 5:)

Gruppe	Kontrast 1	Kontrast 2
Placebo	-	0
niedrig	+	+
hoch	+	-





# Berechnung Orthogonaler Kontraste

Schritt 1: Tabelle aufstellen

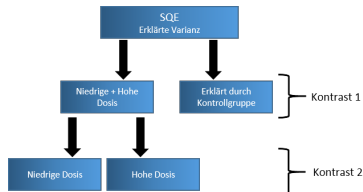
Schritt 2: Unbeteiligte Gruppen  $\rightarrow 0$

Schritt 3: Linke Gruppe  $-$ , Rechte Gruppe  $+$

Schritt 4: Gegenüberliegende Gruppenzahl eintragen

(Schritt 5:)

Gruppe	Kontrast 1	Kontrast 2
Placebo	-2	0
niedrig	+1	+1
hoch	+1	-1



# Berechnung Orthogonaler Kontraste

Schritt 1: Tabelle aufstellen

Schritt 2: Unbeteiligte Gruppen  $\rightarrow 0$

Schritt 3: Linke Gruppe  $-$ , Rechte Gruppe  $+$

Schritt 4: Gegenüberliegende Gruppenzahl eintragen

(Schritt 5: Orthogonalität prüfen)

Gruppe	Kontrast 1	Kontrast 2	Kontrast 1 * Kontrast 1
Placebo	-2	0	0
niedrig	+1	+1	+1
hoch	+1	-1	-1
SUM	0	0	0

# Berechnung Orthogonaler Kontraste

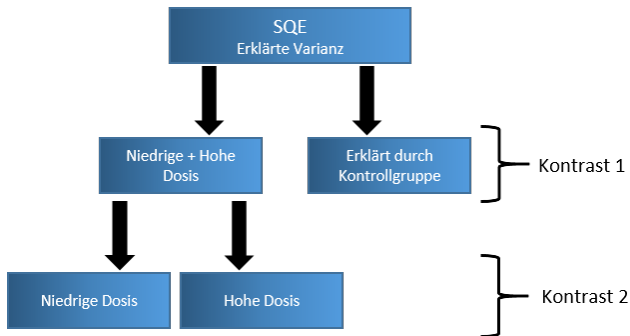
	Viagra	Libido
1	1	3
2	1	2
...		
6	2	5
7	2	2
...		
11	3	7
12	3	4
...		



	Viagra	Libido	Kontr1	Kontr2
1	1	3	-2	0
2	1	2	-2	0
...				
6	2	5	1	1
7	2	2	1	1
...				
11	3	7	1	-1
12	3	4	1	-1
...				

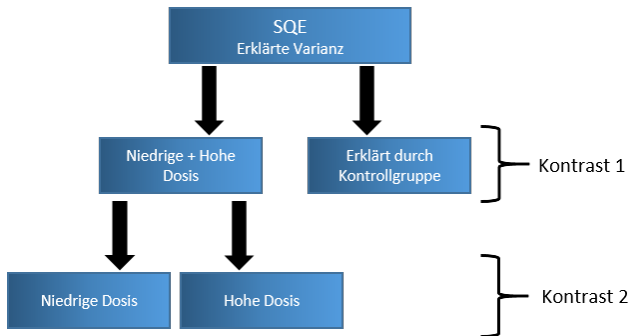
# Kontrastierung über Multiple Regression

- Regressionsformel  $\hat{Y} = (b_0 + b_1 * \text{Kontrast1} + b_2 * \text{Kontrast2} + \dots)$



# Kontrastierung über Multiple Regression

- Regressionsformel  $\hat{Y} = (b_0 + b_1 * \text{Kontrast1} + b_2 * \text{Kontrast2} + \dots)$



- $b_0 = \text{Grand Mean}$
- $b_i = \frac{\overline{\text{Links}} - \overline{\text{Rechts}}}{\text{Gruppenzahl}_{\text{Kontrast } i}}$
- $b_1 = \frac{\overline{\text{Niedrig} + \text{Hoch}} - \overline{\text{Kontrollgruppe}}}{3}$ ,  $b_2 = \frac{\overline{\text{Niedrig}} - \overline{\text{Hoch}}}{2}$

# Kontrastierung über Multiple Regression

```
libido<-c(3,2,1,1,4,5,2,4,2,3,7,4,5,3,6)
dose<-gl(3,5, labels = c("Placebo", "Low Dose", "High Dose"))
viagraDataContrast<-data.frame(dose, libido)
viagraData<-data.frame(dose, libido)
contrast1<-c(-2,1,1)
contrast2<-c(0,1,-1)
contrasts(viagraDataContrast$dose)<-cbind(contrast1, contrast2)
viagraPlannedContrast<-aov(libido~dose, data = viagraDataContrast)
viagraPlanned<-aov(libido~dose, data = viagraData)
```

```
summary.lm(viagraPlannedContrast)
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	3.4667	0.3621	9.574	5.72e-07	***
dosecontrast1	0.6333	0.2560	2.474	0.0293	* //t-Wert f"ur Kontrast1
dosecontrast2	0.9000	0.4435	2.029	0.0652	. //t-Wert f"ur Kontrast2

```
summary.lm(viagraPlanned)
```

```
//Das hier ist falsch
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	2.2000	0.6272	3.508	0.00432	**
doseLow Dose	1.0000	0.8869	1.127	0.28158	
doseHigh Dose	2.8000	0.8869	3.157	0.00827	**

Grundidee:

- Mache paarweise Tests
- Angepasste Signifikanz verhindert, dass der kummulierte Typ 1 Fehler den Toleranzbereich überschreitet

## Grundidee:

- Mache paarweise Tests
- Angepasste Signifikanz verhindert, dass der kummulierte Typ 1 Fehler den Toleranzbereich überschreitet
- (konservativ) Bonferroni Korrektur:  $p_{kr} = \frac{\alpha}{\text{Anzahl der Vergleiche}}$   
Bonferroni, C.E. (1936): *Teoria Statistica Delle Classi e Calcolo Delle Probabilità*



## Grundidee:

- Mache paarweise Tests
- Angepasste Signifikanz verhindert, dass der kummulierte Typ 1 Fehler den Toleranzbereich überschreitet
- (konservativ) Bonferroni Korrektur:  $p_{kr} = \frac{\alpha}{\text{Anzahl der Vergleiche}}$   
Bonferroni, C.E. (1936): *Teoria Statistica Delle Classi e Calcolo Delle Probabilità*
- Pro Test steigt die Wahrscheinlichkeit von Typ 2 Fehlern (Vorhandener Effekt nicht festgestellt)

Verbesserung durch Abstufung:

Hochberg, Y. (1988): *A sharper Bonferroni procedure for multiple tests of significance*

Holm, S. (1979): *A simple rejective multiple test procedure*

Hommel, G. (1988): *A stagewise rejective multiple test procedure based on a modified Bonferroni test*

Grundidee Holm:1979 : Bereits abgearbeitete Vergleiche werden nicht mehr beachtet

- Berechne  $p$  für jeden Vergleich
- Sortiere  $p$  absteigend und vergebe Rang  $j$
- $p_{kr} = \frac{\alpha}{j}$
- From  $j=6$  TO  $j=1$ :  
if  $p < p_{kr} \rightarrow$  Unterschied ist signifikant
- Stoppe sobald  $p > p_{kr}$
- Alle weiteren Unterschiede sind nicht signifikant

Verbesserung durch Aufstufung:

Benjamini, Y. & Hochberg, Y. (1995): *Controlling the false discovery rate – a practical and powerful approach to multiple testing*

Grundidee : Lieber Typ 2 Fehler als Typ 1 Fehler optimieren

- Berechne  $p$  für jeden Vergleich
- Sortiere  $p$  aufsteigend und vergebe Rang  $j$
- $p_{kr} = \frac{j}{k} * \alpha$  mit  $k$  = Anzahl der Vergleiche
- From  $j=6$  TO  $j=1$ :  
if  $p > p_{kr} \rightarrow$  Unterschied ist nicht signifikant
- Stoppe sobald  $p < p_{kr}$
- Alle weiteren Unterschiede sind signifikant

Daten Siehe Superhero.dat im Moodle

\* = signifikant

	p
NT,Super	.000
Super,Hulk	.0014
Spider,Super	.0127
NT,Spider	.0252
NT,Hulk	.1704
Spider,Hulk	.3431

Daten sind erfunden aber Studien existieren dazu bspw:

Davis, P. & SurrIDGE, J.& Hole, S. & Munro-Davies, L. (2007 ): *Superhero-related injuries in paediatrics: a case series*

" Commercial role models for girls are less likely to show risk-taking behaviour; for instance, none of the authors have ever experienced any 'My Little Pony'-related injuries"

Daten Siehe Superhero.dat im Moodle

\* = signifikant

	p	Bonferroni $\frac{\alpha}{k}$
NT,Super	.000	.0083*
Super,Hulk	.0014	.0083*
Spider,Super	.0127	.0083
NT,Spider	.0252	.0083
NT,Hulk	.1704	.0083
Spider,Hulk	.3431	.0083

Daten Siehe Superhero.dat im Moodle

\* = signifikant

	p	Bonferroni $\frac{\alpha}{k}$	$j_1$	Holm $\frac{\alpha}{j_1}$
NT,Super	.000	.0083*	6	.0083*
Super,Hulk	.0014	.0083*	5	.0100*
Spider,Super	.0127	.0083	4	.0125
NT,Spider	.0252	.0083	3	.0167
NT,Hulk	.1704	.0083	2	.0250
Spider,Hulk	.3431	.0083	1	.0500

Daten Siehe Superhero.dat im Moodle

\* = signifikant

	p	Bonferroni $\frac{\alpha}{k}$	$j_1$	Holm $\frac{\alpha}{j_1}$	$j_2$	B. & H. $\frac{j}{k} * \alpha$
NT,Super	.000	.0083*	6	.0083*	1	.0083*
Super,Hulk	.0014	.0083*	5	.0100*	2	.0167*
Spider,Super	.0127	.0083	4	.0125	3	.0250*
NT,Spider	.0252	.0083	3	.0167	4	.0333*
NT,Hulk	.1704	.0083	2	.0250	5	.0417
Spider,Hulk	.3431	.0083	1	.0500	6	.0500

- 1 Motivation
- 2 ANOVA
  - Berechnung
  - Beispiel
  - Robustheit
- 3 Lokalisierung der Unterschiede
  - Geplante Kontrastierung
  - Post Hoc Tests
- 4 Effektstärke
- 5 ANCOVA
  - Voraussetzungen
  - Berechnung



- Determinationskoeffizient  $R^2 = \frac{SQE}{SQT}$
- Weniger Bias:  $\Omega^2 \omega^2 = \frac{SQE - (k * MQR)}{SQT + MQR}$  mit  $k =$  Anzahl der Gruppen

- Determinationskoeffizient  $R^2 = \frac{SQE}{SQT}$
- Weniger Bias:  $\Omega^2 \omega^2 = \frac{SQE - (k * MQR)}{SQT + MQR}$  mit  $k =$  Anzahl der Gruppen
- Guidelines für  $\omega^2$   
Kirk, R.E. (1996): *Practical Significance: A concept whose time has come*
  - .01 → gering
  - .06 → moderat
  - .14 → stark
  - sehr kontextabhängig

- 1 Motivation
- 2 ANOVA
  - Berechnung
  - Beispiel
  - Robustheit
- 3 Lokalisierung der Unterschiede
  - Geplante Kontrastierung
  - Post Hoc Tests
- 4 Effektstärke
- 5 ANCOVA
  - Voraussetzungen
  - Berechnung

# Analysis of Covariance: ANCOVA

- ANOVA + Kovariate
- Kovariate: stetige Variablen, die nicht Teil des Experiments sind, aber die abhängige Variable beeinflussen

# Analysis of Covariance: ANCOVA

- ANOVA + Kovariate
- Kovariate: stetige Variablen, die nicht Teil des Experiments sind, aber die abhängige Variable beeinflussen
- ANOVA ist eigentlich "verkleidete" Regression und ANCOVA "verkleidete" multiple Regression

# Analysis of Covariance: ANCOVA

- ANOVA + Kovariate
- Kovariate: stetige Variablen, die nicht Teil des Experiments sind, aber die abhängige Variable beeinflussen
- ANOVA ist eigentlich "verkleidete" Regression und ANCOVA "verkleidete" multiple Regression

Anwendungsszenarien:

- Fehlervarianz innerhalb der Gruppen reduzieren
- Eliminierung von Bias durch Störvariablen

# Analysis of Covariance: ANCOVA

- ANOVA + Kovariate
- Kovariate: stetige Variablen, die nicht Teil des Experiments sind, aber die abhängige Variable beeinflussen
- ANOVA ist eigentlich "verkleidete" Regression und ANCOVA "verkleidete" multiple Regression

Anwendungsszenarien:

- Fehlervarianz innerhalb der Gruppen reduzieren
- Eliminierung von Bias durch Störvariablen

Wir schauen uns das Thema nur ganz grob an, genauer nachlesen bspw. hier:

Stevens, J.P. (2002): *Applied multivariate statistics for the social sciences*

Wildt, A.R. & Athola, O. (1978): *Analysis of covariance*

# Beispiel

Dosis	Viagra	Libido
1	1	3
1	1	2
...		

Dosis	Viagra	Libido
2	2	7
2	2	5
...		

Dosis	Viagra	Libido
3	3	9
3	3	5
...		

ANOVA Beispiel von vorher (mit geänderten Werten)

Dosis: 1 = Kontrollgruppe, 2 = niedrige Dosierung, 3 = hohe Dosierung



# Beispiel

Dosis	Viagra	Libido	Libido Partner_In
1	1	3	4
1	1	2	1
...			

Dosis	Viagra	Libido	Libido Partner_In
2	2	7	5
2	2	5	3
...			

Dosis	Viagra	Libido	Libido Partner_In
3	3	9	1
3	3	5	3
...			

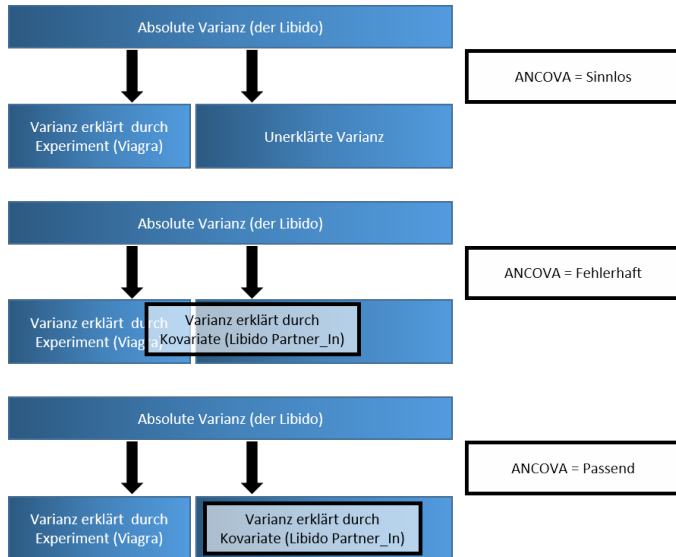
ANCOVA Beispiel mit Kovariate *Libido Partner\_In*

Dosis: 1 = Kontrollgruppe, 2 = niedrige Dosierung, 3 = hohe Dosierung

Zusätzlich zu Voraussetzungen für ANOVA muss gelten:

- Unabhängigkeit der Kovariaten von Prädiktoren
- Homogenität der Regressionslinien Prädiktoren ( $X$ )  $\rightarrow$  Kovariate ( $Y$ )

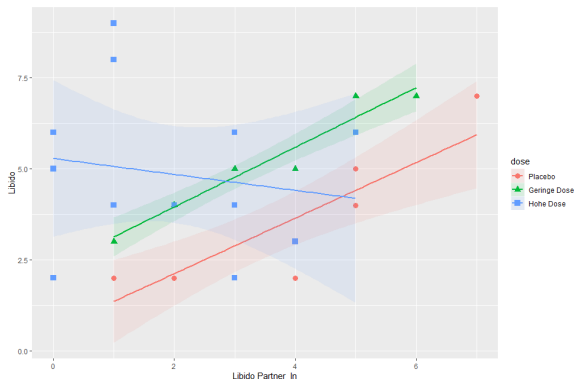
# Unabhängigkeit der Kovariaten



- Sicherstellen durch Randomisierung
- Test: ANOVA mit Kovariate als Outcome und unabhängige Variablen als Prädiktoren
- Wenn Test signifikante Unterschiede anzeigt → Analyse abbrechen

- Sicherstellen durch Randomisierung
- Test: ANOVA mit Kovariate als Outcome und unabhängige Variablen als Prädiktoren
- Wenn Test signifikante Unterschiede anzeigt → Analyse abbrechen
- Negativbeispiel: Analyse des Effekts von Ängstlichkeit auf Verhalten, Depression als Kovariate identifiziert → Depression korreliert stark mit Ängstlichkeit

# Homogenität der Regressionslinien



Diagrammscript im Moodle (`regressionslinienPlot.R`)

- Pro Gruppe einfache Regression mit Kovariater als Outcome und Unabhängiger Variable als Prädiktor
- Regressionslinien im Beispiel sind heterogen (schlecht für ANCOVA)

- Regressionskoeffizienten verschiedener Gruppen sollten etwa gleich sein
- Heterogene Regressionslinien nicht zwangsläufig schlecht, können selbst zu interessanten Hypothesen führen
- → Multilevel Linear Models

- Quadratsummen auf 4 Arten berechenbar (Typ I, Typ II, Typ III, Typ IV)
- Typ I berechnet Einfluß eines Prädiktors abhängig von vorher hinzugefügten Prädiktoren
  - Einfügereihenfolge verändert Ergebnis, wird in R standardmäßig verwendet 😞
- Typ III berechnet SQ nach Einfügen aller anderen Prädiktoren
  - Per default verwenden bei ANOVA mit mehreren Prädiktoren



```
library(car)
libido<-c(3,2,5,2,2,2,7,2,4,7,5,3,4,4,7,5,4,9,2,6,3,4,4,4,6,4,6,2,8,5)
partnerLibido<-c(4,1,5,1,2,2,7,4,5,5,3,1,2,2,6,4,2,1,3,5,4,3,3,2,0,1,3,0,1,0)
dose<-c(rep(1,9),rep(2,8), rep(3,13))
dose<-factor(dose, levels = c(1:3), labels = c("Placebo",
        "Geringe Dose", "Hohe Dose"))
viagraData<-data.frame(dose, libido, partnerLibido)

contrasts(viagraData$dose)<-cbind(c(-2,1,1), c(0,-1,1)) #Siehe ANOVA Vorlesung
viagraModel<-aov(libido~ partnerLibido + dose, data = viagraData)
Anova(viagraModel, type="III") #Typ III Summenquadrate verwenden

Anova Table (Type III tests)

Response: libido
          Sum Sq Df F value    Pr(>F)
(Intercept)  76.069  1 25.0205 3.342e-05 ***
partnerLibido 15.076  1  4.9587  0.03483 * <---Kovariate signifikant wenn p<0.05
dose          25.185  2  4.1419  0.02745 *
Residuals    79.047 26
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# Test auf Homogenität der Regressionslinien

```
library(car)
libido<-c(3,2,5,2,2,2,7,2,4,7,5,3,4,4,7,5,4,9,2,6,3,4,4,4,6,4,6,2,8,5)
partnerLibido<-c(4,1,5,1,2,2,7,4,5,5,3,1,2,2,6,4,2,1,3,5,4,3,3,2,0,1,3,0,1,0)
dose<-c(rep(1,9),rep(2,8), rep(3,13))
dose<-factor(dose, levels = c(1:3), labels = c("Placebo",
        "Geringe Dose", "Hohe Dose"))
viagraData<-data.frame(dose, libido, partnerLibido)
contrasts(viagraData$dose)<-cbind(c(-2,1,1), c(0,-1,1)) #Kontrastierung

hoRS<-update(viagraModel, .~. + partnerLibido:dose) # Interaction Term
Anova(hoRS, type="III") #Typ III Summenquadrate verwenden
```

Anova Table (Type III tests)

Response: libido

	Sum Sq	Df	F value	Pr(>F)	
(Intercept)	53.542	1	21.9207	9.323e-05	***
partnerLibido	17.182	1	7.0346	0.01395	*
dose	36.558	2	7.4836	0.00298	**
partnerLibido:dose	20.427	2	4.1815	0.02767	*
Residuals	58.621	24			

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

//Pr<0.05: Regressionslinien  
//heterogen (schlecht)

- ANOVA vergleicht mehrere Mittelwerte als Omnibus Test
- F-Verteilung
- Robustheit
- Familienbezogene Fehler bei paarweisen Vergleichen
- Lokalisierung von Unterschieden
  - Kontrastierung
  - Post Hoc Tests
- Determinationskoeffizient  $R^2$  und quadriertes Omega  $\omega^2$  als Effektstärke
- Übersprungen: Weitere Kontrastierungen (Nicht-Orthogonal, Polynomial)

- Analysis of Covariance (ANCOVA) = ANOVA + Kovariate
- Kovariate sollen Varianz außerhalb des Experiments teilweise erklären
- Voraussetzungen:
  - Siehe ANOVA
  - Unabhängigkeit der Kovariaten von Prädiktoren
  - Homogenität der Regressionslinien Prädiktoren ( $X$ )  $\rightarrow$  Kovariate ( $Y$ )
- ANCOVA ist ein extrem komplexes Thema mit vielen Fallstricken, also sehr vorsichtig verwenden und Voraussetzungen absichern
- Übersprungen: Kontrastierung (Siehe ANOVA), Post-Hoc Tests (Tukey und Dunnets PH-Tests), Robustheit (Siehe Wilcox, 2005)