

# Statistik für Digital Humanities

## Analysis of Covariance: ANCOVA (GLM 2/5)

Dr. Jochen Tiepmar

Institut für Informatik  
Abteilung Computational Humanities  
**Universität Leipzig**

20. Januar 2020

[Letzte Aktualisierung: 02/02/2020, 11:33]

# Analysis of Covariance: ANCOVA

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

Anwendungsszenarien:

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

Wir schauen uns das Thema nur ganz grob an.

# 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 Kovariante *Libido Partner\_In*

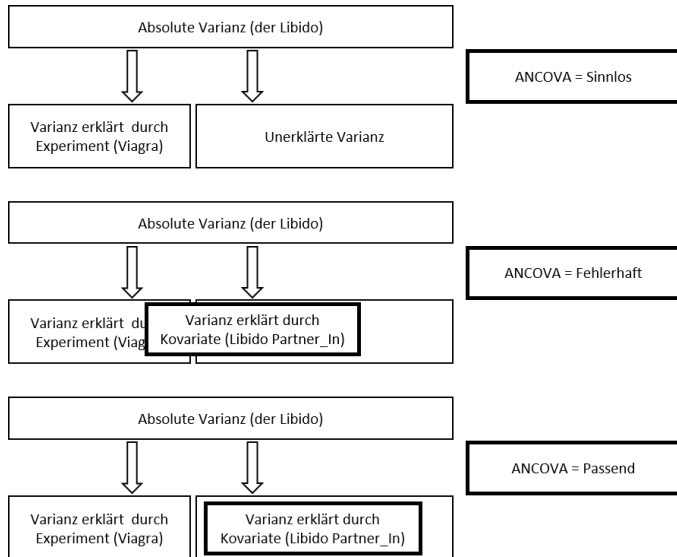
Dosis: 1 = Kontrollgruppe, 2 = niedrige Dosierung, 3 = hohe Dosierung

# Vorraussetzungen für ANCOVA

Zusätzlich zu Vorraussetzungen für ANOVA muss gelten:

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

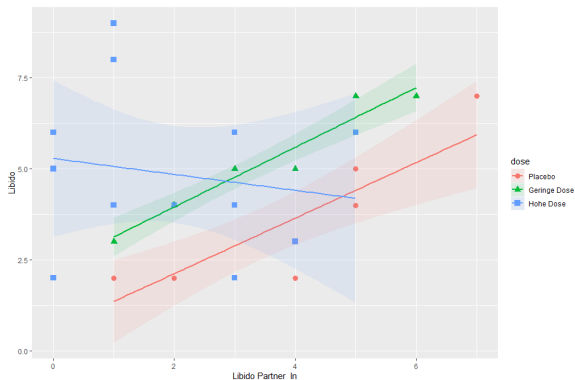
# Unabhängigkeit der Kovariaten



# Unabhängigkeit der Kovariaten

- 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)

# Homogenität der Regressionslinien

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



# Typen von Quadratsummen

- 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

# Berechnung

```
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 *
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)) //Siehe ANOVA Vorlesung

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	*
					//Pr<0.05: Regressionslinien //heterogen (schlecht)
Residuals	58.621	24			

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

# Zusammenfassung

- 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 ( $X$ ) → Kovariate ( $Y$ )
- ANCOVA ist ein extrem komplexes Thema mit vielen Fallstricken, also sehr vorsichtig verwenden und Voraussetzungen absichern
- Übersprungen: Kontrastierung, Post-Hoc Tests, Robustheit (Wilcox)