Statistik für Digital Humanities

Mehrfaktorielle ANOVA

Dr. Jochen Tiepmar

Institut für Informatik
Abteilung Computational Humanities
Universität Leipzig

20. Januar 2020

[Letzte Aktualisierung: 02/02/2020, 11:34]

Mehrfaktorielle ANOVA

Bekannt:

- (einfaktorielle) ANOVA: Experiment durch Veränderung einer unabhängigen Variable (Gruppenzuordnung)
- ANCOVA: Experiment durch Veränderung einer unabhängigen Variable (Gruppenzuordnung) unter Eliminierung einer Variable außerhalb des Experiments

Jetzt:

- Mehrfaktorielle ANOVA: Experiment durch Veränderung <u>mehrerer</u> unabhängigen Variablen (Gruppenzuordnungen)

Arten Mehrfaktorieller ANOVA

- Unabhängig mehrfaktoriell: Mehrere Prädiktoren, jeder gemessen an verschiedenen Probanden $\rightarrow$ GLM 3/5, Heute
- Abhängig mehrfaktoriell: Mehrere Prädiktoren, jeder gemessen an gleichen Probanden $\rightarrow$ GLM 4/5
- Gemischt mehrfaktoriell: Mehrere Prädiktoren, einige mit gleichen Probanden gemessen, einige mit verschiedenen $\rightarrow$ GLM 5/5

Arten von ANOVA

- Einfaktorielle Unabhängige / One Way Independent ANOVA: 1 Prädiktor, verschiedene Probanden
- Two Way Repeated Measures ANOVA: 2 Prädiktoren, gleiche Probanden
- Zweifaktorielle Gemischte ANOVA: 2 Prädiktoren, 1 gemessen mit denselben Probanden und 1 gemessen mit verschiedenen Probanden
- Three Way Independent ANOVA: 3 Prädiktoren, jeweils verschiedene Probanden
- ...

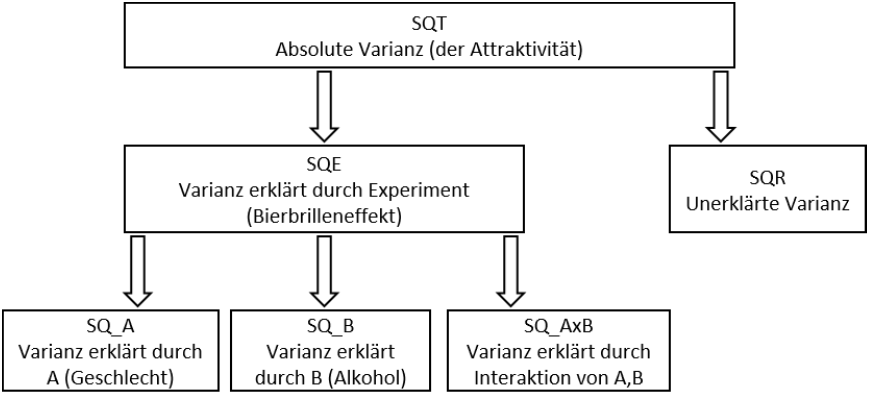
Zweifaktorielle Unabhängige ANOVA (GLM 3/5)

Beispiel Bierbrilleneffekt nach Geschlecht:

Alkohol	Kein		2 Gläser		4 Gläser	
Geschlecht	Frau	Mann	Frau	Mann	Frau	Mann
Attraktivität	65	50	70	45	55	30
des Dates	70	55	65	60	65	30
	60	80	60	85	70	30
	60	65	70	65	55	55
	60	70	65	70	55	35
	55	75	60	70	60	20
	60	75	60	80	50	45
	55	65	50	60	50	40
Sum	485	535	500	535	460	285
Mean	60.625	66.875	62.50	66.875	57.50	35.625
Var	24.55	106.70	42.86	156.70	50	117.41

Datensatz im Moodle (goggles.csv)

Zweifaktorielle Unabhängige ANOVA



Abweichungsquadrate bei mehreren Mittelwerten

Genau wie vorher bei ANOVA:

- $SQT = \sum(x_i - Grand\ Mean)^2$
- $SQR = \sum(x_i - \overline{group})^2 = \sum s^2_{group} * (n_{group} - 1)$, $MQR = \frac{SQR}{n-k}$
- $SQE = \sum n_{group} * (\overline{group} - Grand\ Mean)^2$, $MQE = \frac{SQE}{k-1}$

Jetzt neu:

- **Effekt des Prädiktors A**
  $SQ_A = \sum n_{groupA} * (\overline{groupA} - Grand\ Mean)^2$
  $MQ_A = \frac{SQ_A}{k_A - 1}$ mit k_A = Gruppenanzahl in A
- **Interaktionseffekt**
  $SQ_{AxB} = SQE - SQ_A - SQ_B$
  $MQ_{AxB} = \frac{SQ_{AxB}}{(k_A - 1) * (k_B - 1)}$ mit k = Gruppenanzahl

F-Ratios

- $SQR = \sum(x_i - \overline{group})^2 = \sum s^2_{group} * (n_{group} - 1)$, $MQR = \frac{SQR}{n-k}$

- **F-Ratio für Effekt A**
  $SQ_A = \sum n_{groupA} * (\overline{groupA} - Grand\ Mean)^2$
  $MQ_A = \frac{SQ_A}{k_A - 1}$ mit k = Gruppenanzahl
  $F_A = \frac{MQ_A}{MQR}$
- **F-Ratio Interaktionseffekt**
  $SQ_{AxB} = SQE - SQ_A - SQ_B$
  $MQ_{AxB} = \frac{SQ_{AxB}}{(k_A - 1) * (k_B - 1)}$ mit k = Gruppenanzahl
  $F_{AxB} = \frac{MQ_{AxB}}{MQR}$

Beispiel $F_{Geschlecht}$

Frau:

Kein	2 Gläser	4 Gläser
65	70	55
70	65	65
...	...	...

Mann:

Kein	2 Gläser	4 Gläser
50	45	30
55	60	30
...	...	...

$\overline{Frau} = 60.21$, $\overline{Mann} = 56.46$, $GrandMean = 58.33$

$MQR = 83.036$

$SQ_{Geschlecht} = \sum n_{Frau} * (\overline{Frau} - Grand\ Mean)^2 + n_{Mann} * (\overline{Mann} - Grand\ Mean)^2 =$
$24 * (60.21 - 58.33)^2 + 24 * (56.46 - 58.33)^2 = 168.75$

$MQ_{Geschlecht} = \frac{SQ_{Geschlecht}}{k_{Geschlecht} - 1} = \frac{168.75}{1} = 168.75$

$F_{Geschlecht} = \frac{MQ_{Geschlecht}}{MQR} = \frac{168.75}{83.036} = 2.032$

Beispiel $F_{Alkohol}$

Kein:

Frau	Mann
65	50
70	55
...	...

2 Gl.:

Frau	Mann
70	45
65	60
...	...

4 Gl.:

Frau	Mann
55	30
65	30
...	...

$\overline{Kein} = 63.75$, $\overline{2\ Gl} = 64.6875$, $\overline{4\ Gl} = 46.5625$, $GrandMean = 58.33$

$MQR = 83.036$

$SQ_{Alkohol} = \sum n_{Kein} * (\overline{Kein} - Grand\ Mean)^2 + n_{2\ Gl} * (\overline{2\ Gl} - Grand\ Mean)^2 + n_{4\ Gl} * (\overline{4\ Gl} - Grand\ Mean)^2 =$

$16 * (63.75 - 58.33)^2 + 16 * (64.6875 - 58.33)^2 + 16 * (46.5625 - 58.33)^2 = 3332.292$

$MQ_{Alkohol} = \frac{SQ_{Alkohol}}{k_{Alkohol} - 1} = \frac{3332.292}{2} = 1666.146$

$F_{Alkohol} = \frac{MQ_{Alkohol}}{MQR} = \frac{1666.146}{83.036} = 20.065$

Beispiel $F_{Alkohol\ x\ Geschlecht}$

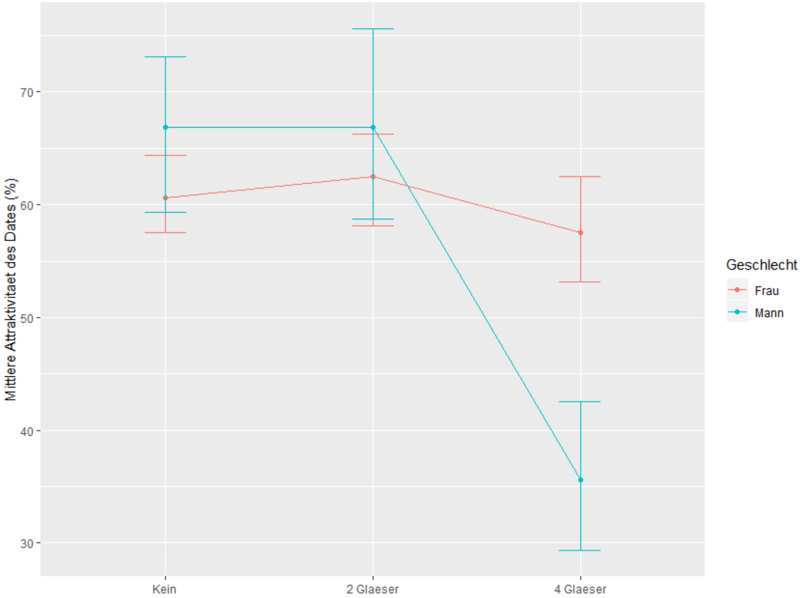
$SQ_{Alkohol} = 3332.292$
$SQ_{Geschlecht} = 168.75$
$SQE = 5479.167$
$MQR = 83.036$
$SQ_{Alkohol\ x\ Geschlecht} = SQE - SQ_{Alkohol} - SQ_{Geschlecht} = 5479.167 - 18.75 - 332.292 = 1978.125$
$MQ_{Alkohol\ x\ Geschlecht} = \frac{SQ_{Alkohol\ x\ Geschlecht}}{(k_{Alkohol}-1)*(k_{Geschlecht}-1)} = \frac{1978.125}{2} = 989.062$
$F_{Alkohol\ x\ Geschlecht} = \frac{MQ_{Alkohol\ x\ Geschlecht}}{MQR} = \frac{989.062}{83.036} = 11.911$

Beispiel

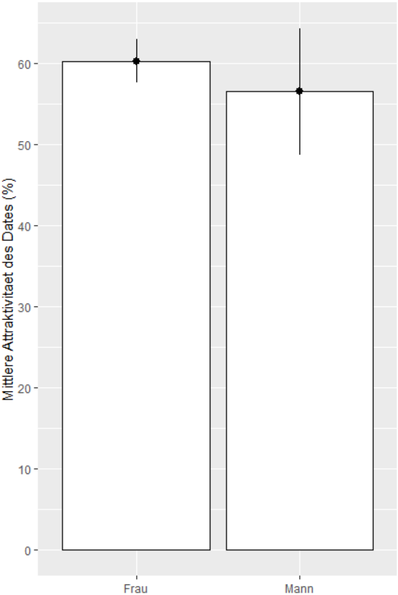
$F_{Geschlecht} = 2.032$
$F_{Alkohol} = 20.065$
$F_{Alkohol\ x\ Geschlecht} = 11.911$

- Interpretation analog zu ANOVA
  - $\rightarrow$ eigentlich machen wir nichts anderes als je 1 ANOVA für jeden Effekt und die Interaktion
- Kontrastierung und Post-Hoc Tests anwendbar
- Einfache Effektanalyse möglich indem man Unterschiede bei Variation eines Faktors analysiert (Siehe Unterschiede im Interaktionsgraph)

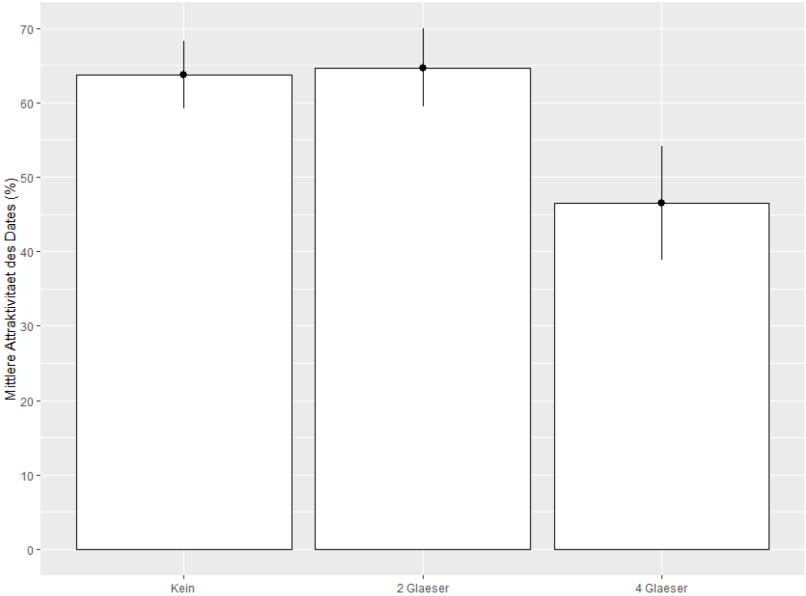
Interaktionsgraph



Visualisierung



Visualisierung



Effektstärke

Bestimmbar mittels ω^2, aber sehr umständlich. Deshalb R Skript:

```
omega_factorial<-function(n,a,b, SQa, SQb, SQab, SQr)
{
        MQa<-SQa/(a-1)
        MQb<-SQb/(b-1)
        MQab<-SQab/((a-1)*(b-1))
        MQr<-SQr/(a*b*(n-1))
        varA<-((a-1)*(MQa-MQr))/(n*a*b)
        varB<-((b-1)*(MQb-MQr))/(n*a*b)
        varAB<-((a-1)*(b-1)*(MQab-MQr))/(n*a*b)
        varTotal<-varA + varB + varAB + MQr

        print(paste("Omega-Squared A: ", varA/varTotal))
        print(paste("Omega-Squared B: ", varB/varTotal))
        print(paste("Omega-Squared AB: ", varAB/varTotal))
        }

omega_factorial(8,2,3,169,3332,1978,3488)
```

Zusammenfassung

- Mehrfaktorielle ANOVA: Variation mehrerer Variablen, Mehrfache Gruppierung
- 2-Faktorielle unabhängige ANOVA: 2 Gruppierungen (Alkohol und Geschlecht) mit verschiedenen Probanden
- Interaktionsgrad als Messwert
- Berechnung analog zu einfaktorieller ANOVA aber für jede Gruppierung einzeln sowie für Interaktion
  - $\rightarrow$ F-Ratio interpretieren
  - $\rightarrow$ Kontrastierung, Post Hoc Tests
  - $\rightarrow$ Effektstärke
- Einfache Effektanalysen bereits mit Visualisierung möglich
- Robust: Wilcox (2005)
- Im Moodle finden sich ein paar teils sehr komplexe Beispielanalysen als Anschauungsbeispiele