

# Beispiel $\chi^2$ -Verfahren

André Meichtry

## Contents

Daten aus AMT S.93	1
Sind Antwort und Freundlichkeit unabhängig?	2
Als $\chi^2$ -Test auf Unabhängigkeit (approximativ)	2
Als Fisher-Exakt-Test (exact)	2
Als GLM	2
Als LogLineares Modell*	5

## Daten aus AMT S.93

```
data<-matrix(c(13,15,3,22,5,19),ncol=3,
             dimnames=list(Antwort=c("ja","nein"),Freundlichkeit=c("freundlich","forsch","sachlich")))
datatable<-as.table(data)
datatable
```

```
      Freundlichkeit
Antwort freundlich forsch sachlich
ja             13      3      5
nein           15     22     19
```

```
addmargins(datatable)
```

```
      Freundlichkeit
Antwort freundlich forsch sachlich Sum
ja             13      3      5    21
nein           15     22     19    56
Sum            28     25     24    77
```

```
margin.table(datatable,1)
```

```
Antwort
ja nein
21  56
```

```
margin.table(datatable,2)
```

```
Freundlichkeit
freundlich forsch sachlich
28         25         24
```

```
prop.table(datatable)
```

```
      Freundlichkeit
```

```
Antwort freundlich  forsch sachlich
ja      0.168831 0.038961 0.064935
nein    0.194805 0.285714 0.246753
```

```
prop.table(datatable,1)
```

```
      Freundlichkeit
Antwort freundlich  forsch sachlich
ja      0.61905 0.14286  0.23810
nein    0.26786 0.39286  0.33929
```

```
prop.table(datatable,2)
```

```
      Freundlichkeit
Antwort freundlich  forsch sachlich
ja      0.46429 0.12000  0.20833
nein    0.53571 0.88000  0.79167
```

```
summary(datatable)
```

```
Number of cases in table: 77
Number of factors: 2
Test for independence of all factors:
  Chisq = 8.6, df = 2, p-value = 0.013
```

## Sind Antwort und Freundlichkeit unabhängig?

### Als $\chi^2$ -Test auf Unabhängigkeit (approximativ)

```
chisq.test(datatable)
```

```
Pearson's Chi-squared test
```

```
data:  datatable
X-squared = 8.62, df = 2, p-value = 0.013
```

Unabhängigkeit von Freundlichkeit und Antwortverhalten kann verworfen werden.

### Als Fisher-Exakt-Test (exact)

```
fisher.test(datatable)
```

```
Fisher's Exact Test for Count Data
```

```
data:  datatable
p-value = 0.013
alternative hypothesis: two.sided
```

### Als GLM

Zählraten kann man mit GLM (Generalisiertes Lineares Modell) anpassen, mit Annahme einer Poisson-Verteilung und log-Link (siehe Theorie Generalisierte Lineare Modelle). Dazu brauchen wir die Daten als data.frame.

```
datadf<-as.data.frame(datatable)
datadf
```

```
  Antwort Freundlichkeit Freq
1      ja      freundlich  13
2     nein      freundlich  15
3      ja         falsch   3
4     nein         falsch  22
5      ja      sachlich   5
6     nein      sachlich  19
```

Das reduzierte Modell ohne Antwort:Freundlichkeit Interaktionseffekt (entspricht der Nullhypothese: Effekt von Antwortverhalten auf die Anzahl ist unabhängig von Freundlichkeit) kann verworfen werden.

```
modglm<-glm(Freq~Antwort*Freundlichkeit,data=datadf,family="poisson")
modglm0<-glm(Freq~Antwort+Freundlichkeit,data=datadf,family="poisson")
anova(modglm,modglm0,test="LRT")
```

Analysis of Deviance Table

```
Model 1: Freq ~ Antwort * Freundlichkeit
Model 2: Freq ~ Antwort + Freundlichkeit
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1         0         0.00
2         2         8.65 -2    -8.65    0.013
```

Beim Argument test hätten wir statt LRT auch Chisq schreiben können, das ist äquivalent. Wir können die Nullhypothese, dass es keinen Interaktionseffekt, dass also Antwortverhalten und Freundlichkeit unabhängig sind, verwerfen:

```
summary(modglm)
```

Call:

```
glm(formula = Freq ~ Antwort * Freundlichkeit, family = "poisson",
    data = datadf)
```

Deviance Residuals:

```
[1] 0 0 0 0 0 0 0
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	2.565	0.277	9.25	<2e-16
Antwortnein	0.143	0.379	0.38	0.706
Freundlichkeitfalsch	-1.466	0.641	-2.29	0.022
Freundlichkeitsachlich	-0.956	0.526	-1.82	0.069
Antwortnein:Freundlichkeitfalsch	1.849	0.723	2.56	0.011
Antwortnein:Freundlichkeitsachlich	1.192	0.629	1.89	0.058

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 2.5496e+01 on 5 degrees of freedom
Residual deviance: 1.3323e-15 on 0 degrees of freedom
AIC: 37.17
```

Number of Fisher Scoring iterations: 3

Interpretation: Wir sehen, dass die Residuendevianz dieses Modells Null ist! Das kommt daher, dass dieses

Modell das saturierte, maximal mögliche Modell ist mit gleich vielen Parametern (6) wie Daten (6). Der Koeffizient für Antwortnein, 0.1431008 ist **der Unterschied in log-Anzahlen zwischen Nein-Sagern und Ja-Sagern** (für Referenzkategorie auf Freundlichkeit (freundlich)), also ist 1.1538462 das Verhältnis der Anzahlen von Nein-Sagern relativ zu Ja-Sagern (für Referenzkategorie auf Freundlichkeit (freundlich)).

R hat als Default-Kodierung sogenannte Treatment-Kontraste, die wir schon oft angetroffen haben. Eine alternative Parametrisierung wäre die **Means-Parameterisierung**.

```
summary(modglmMeans<-glm(Freq~Antwort:Freundlichkeit-1,data=datadf,family="poisson"))
```

Call:

```
glm(formula = Freq ~ Antwort:Freundlichkeit - 1, family = "poisson",
    data = datadf)
```

Deviance Residuals:

```
[1] 0 0 0 0 0 0
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
Antwortja:Freundlichkeitfreundlich	2.565	0.277	9.25	< 2e-16
Antwortnein:Freundlichkeitfreundlich	2.708	0.258	10.49	< 2e-16
Antwortja:Freundlichkeitforsch	1.099	0.577	1.90	0.05706
Antwortnein:Freundlichkeitforsch	3.091	0.213	14.50	< 2e-16
Antwortja:Freundlichkeitsachlich	1.609	0.447	3.60	0.00032
Antwortnein:Freundlichkeitsachlich	2.944	0.229	12.83	< 2e-16

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 2.7651e+02 on 6 degrees of freedom
Residual deviance: 3.3307e-15 on 0 degrees of freedom
AIC: 37.17
```

Number of Fisher Scoring iterations: 3

Wenn wir auf die geschätzten Parameter die Exponentialfunktion anwenden, dann erhalten wir gerade die Anzahlen.

```
as.data.frame(exp(summary(modglmMeans)$coef[,1]))
```

	exp(summary(modglmMeans)\$coef[, 1])
Antwortja:Freundlichkeitfreundlich	13
Antwortnein:Freundlichkeitfreundlich	15
Antwortja:Freundlichkeitforsch	3
Antwortnein:Freundlichkeitforsch	22
Antwortja:Freundlichkeitsachlich	5
Antwortnein:Freundlichkeitsachlich	19

Möchten wir z.B. Konfidenzintervalle für Anzahlen, ist das zu haben mit

```
confint(modglmMeans)
```

	2.5 %	97.5 %
Antwortja:Freundlichkeitfreundlich	1.96724	3.0634
Antwortnein:Freundlichkeitfreundlich	2.15540	3.1747
Antwortja:Freundlichkeitforsch	-0.29305	2.0511
Antwortnein:Freundlichkeitforsch	2.64189	3.4817
Antwortja:Freundlichkeitsachlich	0.58373	2.3746
Antwortnein:Freundlichkeitsachlich	2.45837	3.3627

## Als LogLineares Modell\*

Eine dritte Analysemehtode ist das **log-lineare Modell**. Wir gehen hier nicht näher darauf ein.

```
library(MASS)
loglm(~Antwort+Freundlichkeit,data=datatable)
```

Call:

```
loglm(formula = ~Antwort + Freundlichkeit, data = datatable)
```

Statistics:

	X <sup>2</sup>	df	P(> X <sup>2</sup> )
Likelihood Ratio	8.6537	2	0.013209
Pearson	8.6218	2	0.013421

Unabhängigkeitsmodell kann verworfen werden.