

Lösung zu Kapitel 10: Beispiel 5

Die Datei `us-election80.csv` enthält die Variablen Wahlverhalten (`vote`), Hautfarbe (`race`), politische Einstellung (`polview`) und eine Gewichtungvariable (`weight`). Es soll ein geeignetes Modell für das Wahlverhalten mit den Prädiktoren ‚Hautfarbe‘ und ‚politische Einstellung‘ gefunden werden.

- Wir lesen `us-election80.csv` ein und speichern die Daten in `gss` ab. Da `gss$vote` in `gss` mit 1 und 2 codiert ist, wir aber für die logistische Regression 0 und 1 benötigen erstellen wir ein neues Objekt `vote`, das um 1 verminderte Werte enthält. `gss$race` machen wir mit `factor` zu einem Faktor, dem wir entsprechende Labels zu weisen und in `race` abspeichern. `gss$polview` und `gss$weight` speichern wir unverändert in den Objekten `polv` und `weig` ab.

R

```
> gss <- read.csv2("us-election80.csv", header = TRUE)
> vote <- gss$vote - 1
> race <- factor(gss$race, labels = c("weiß", "nicht weiß"))
> polv <- gss$polview
> weig <- gss$weight
```

- Zunächst schätzen wir ein ‚volles‘ Modell (d.h. alle Haupteffekte und Interaktionen), das wir in `modell.1` abspeichern. Da wir eine logistische Regression rechnen wollen müssen wir in `glm` die Einstellung `family=binomial` treffen und zusätzlich müssen wir noch unsere Gewichtungvariable mit `weights=weig` berücksichtigen. Wenn wir ein `summary` auf `modell.1` machen sehen wir, dass der Interaktionseffekt nicht signifikant ($p = .676$) ist.

R

```
> modell.1 <- glm(vote ~ race * polv, family = binomial, weights = weig)
> summary(modell.1)
```

Call:

```
glm(formula = vote ~ race * polv, family = binomial, weights = weig)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-14.9869	-2.8491	0.2292	1.8785	13.9109

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.04464	0.26754	7.642	2.13e-14 ***
racenicht weiß	2.35760	1.42859	1.650	0.0989 .
polv	-0.49570	0.06052	-8.190	2.61e-16 ***
racenicht weiß:polv	0.12995	0.31084	0.418	0.6759

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	1269.7	on 23	degrees of freedom
Residual deviance:	1096.8	on 20	degrees of freedom

AIC: 1104.8

Number of Fisher Scoring iterations: 4

- Wir können daher weitergehen und den Interaktionseffekt entfernen. Hierfür verwenden wir `update`, geben `modell.1` an und entfernen den hierarchisch höchsten Term mit `-race:polv` und speichern alles in `modell.2` ab.

Ein Summary zeigt, dass nun alle Parameter signifikant sind. Wenn wir einen Likelihood-Ratio-Test der Modelle mit `anova` machen und einen χ^2 -Test mit `test="Chisq"` anfordern sehen wir, dass das reduzierte Modell nicht signifikant schlechter ist als das volle ($p = .677$).

In `modell.3` und `modell.4` nehmen wir zusätzlich noch einmal `race` und einmal `polview` heraus, was in den Likelihood-Ratio-Tests jedoch zu signifikanten Ergebnissen (beide $p < .001$) führt, wodurch wir uns letztlich für `modell.2` entscheiden.

R

```
> modell.2 <- update(modell.1, . ~ . - race:polv)
> summary(modell.2)
```

Call:

```
glm(formula = vote ~ race + polv, family = binomial, weights = weig)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-14.9804	-2.8330	0.1848	1.8448	13.9171

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.02395	0.26238	7.714	1.22e-14 ***
racenicht weiß	2.93691	0.47221	6.219	4.99e-10 ***
polv	-0.49084	0.05927	-8.282	< 2e-16 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1269.7 on 23 degrees of freedom
Residual deviance: 1097.0 on 21 degrees of freedom
AIC: 1103

Number of Fisher Scoring iterations: 4

R

```
> anova(modell.2, modell.1, test = "Chisq")
```

Analysis of Deviance Table

Model 1: vote ~ race + polv

Model 2: vote ~ race * polv

	Resid. Df	Resid. Dev	Df	Deviance	P(> Chi)
1	21	1097.0			
2	20	1096.8	1	0.17374	0.6768

R

```
> modell.3 <- update(modell.2, . ~ . - race)
> modell.4 <- update(modell.2, . ~ . - polv)
> anova(modell.3, modell.2, test = "Chisq")
```

Analysis of Deviance Table

Model 1: vote ~ polv

Model 2: vote ~ race + polv

	Resid. Df	Resid. Dev	Df	Deviance	P(> Chi)
1	22	1177.4			
2	21	1097.0	1	80.432	< 2.2e-16 ***

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

R

```
> anova(modell.4, modell.2, test = "Chisq")
```

Analysis of Deviance Table

Model 1: vote ~ race

Model 2: vote ~ race + polv

	Resid. Df	Resid. Dev	Df	Deviance	P(> Chi)
1	22	1174.2			
2	21	1097.0	1	77.192	< 2.2e-16 ***

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