

1. Aufgabe

Bei einer Gruppe von Personen soll eine binäre Zielgrösse ($y = 0$ oder $y = 1$) durch eine erklärende Variable (x) und das Geschlecht ($g = "M"$ oder $g = "W"$) erklärt werden.

Die Daten sind im data frame `dat` in folgendem rda-File gespeichert: `ueb202132.rda`.

Passen Sie ein logistisches Regressionsmodell mit Wechselwirkung an.

Welche der folgenden Aussagen sind korrekt?

- (a) Im data frame `dat` sind 332 Datenpunkte enthalten.
- (b) Der Effekt der Gruppe ist signifikant.
- (c) Die Wechselwirkung von g und x ist signifikant.
- (d) Verwenden Sie nun die Logistische Regression mit der Formel $y \sim g + x$ (unabhängig von Ihren vorherigen Resultaten). Wenn man von der Gruppe der Frauen in die Gruppe der Männer wechselt, verändern sich die odds für $y = 1$ um den Faktor 1.2422.
- (e) Verwenden Sie wieder die Logistische Regression mit der Formel $y \sim g + x$ (unabhängig von Ihren vorherigen Resultaten). Die Wahrscheinlichkeit für $y = 1$ für eine Frau mit $x = -0.2957$ wird als 0.6774 vorhergesagt.

Lösung

Im folgenden Code bezeichnet `fname` den Dateinamen der jeweiligen `.rda`-Datei (in Anführungszeichen), `xVal` bezeichnet den neuen Wert für x und `gVal` den neuen Wert für g (Aufgabenteil e).

```
> load(fname)
> fit <- glm(y ~ x * g, data=dat, family="binomial")
> summary(fit)
```

Call:

```
glm(formula = y ~ x * g, family = "binomial", data = dat)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.607	-0.450	0.149	0.447	2.243

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.1475	0.3163	3.63	0.00029	***
x	0.9394	0.1779	5.28	1.3e-07	***
gW	0.0146	0.4790	0.03	0.97562	
x:gW	0.2335	0.2678	0.87	0.38326	

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 315.44 on 239 degrees of freedom
Residual deviance: 163.64 on 236 degrees of freedom
AIC: 171.6

```

Number of Fisher Scoring iterations: 6

> fit2 <- glm(y ~ x + g, data=dat, family="binomial")
> summary(fit2)

Call:
glm(formula = y ~ x + g, family = "binomial", data = dat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.635  -0.460   0.154   0.470   2.374

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.271     0.303   4.19 2.8e-05 ***
x              1.057     0.133   7.96 1.7e-15 ***
gW            -0.217     0.395  -0.55  0.58
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 315.44  on 239  degrees of freedom
Residual deviance: 164.41  on 237  degrees of freedom
AIC: 170.4

Number of Fisher Scoring iterations: 6

> newdata <- data.frame(x=xVal, g=gVal)
> # für die log-odds
> predict(fit1, newdata=newdata, type="link")

      1
0.7417

> # für die odds
> exp(predVals[1])

[1] 2.1

> # für die Wahrscheinlichkeit
> predict(fit1, newdata=newdata, type="response")

      1
0.6774

(a) False. Im data frame dat sind 240 Datenpunkte enthalten.
(b) False. Der wahre p-Wert ist 0.9756.
(c) False. Der wahre p-Wert ist 0.3833.
(d) True. Die odds verändern sich um den Faktor  $\exp(-gW) = 1.2422$ .
(e) True. Der wahre Wert ist 0.6774.

```

2. Aufgabe

Bei einer Gruppe von Personen soll eine binäre Zielgrösse ($y = 0$ oder $y = 1$) durch eine erklärende Variable (x) und das Geschlecht ($g = "M"$ oder $g = "W"$) erklärt werden.

Die Daten sind im data frame `dat` in folgendem rda-File gespeichert: `ueb25423.rda`.

Passen Sie ein logistisches Regressionsmodell mit Wechselwirkung an.

Welche der folgenden Aussagen sind korrekt?

- (a) Im data frame `dat` sind 292 Datenpunkte enthalten.
- (b) Der Effekt der Gruppe ist signifikant.
- (c) Die Wechselwirkung von g und x ist signifikant.
- (d) Verwenden Sie nun die Logistische Regression mit der Formel $y \sim g + x$ (unabhängig von Ihren vorherigen Resultaten). Wenn man von der Gruppe der Männer in die Gruppe der Frauen wechselt, verändern sich die odds für $y = 1$ um den Faktor 0.3241.
- (e) Verwenden Sie wieder die Logistische Regression mit der Formel $y \sim g + x$ (unabhängig von Ihren vorherigen Resultaten). Die odds für $y = 1$ für einen Mann mit $x = 0.6327$ werden als 0.9077 vorhergesagt.

Lösung

Im folgenden Code bezeichnet `fname` den Dateinamen der jeweiligen `.rda`-Datei (in Anführungszeichen), `xVal` bezeichnet den neuen Wert für x und `gVal` den neuen Wert für g (Aufgabenteil e).

```
> load(fname)
> fit <- glm(y ~ x * g, data=dat, family="binomial")
> summary(fit)
```

Call:

```
glm(formula = y ~ x * g, family = "binomial", data = dat)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.3157	-0.3092	0.0152	0.4308	2.4945

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.087	0.259	4.19	2.8e-05 ***
x	0.766	0.128	5.99	2.1e-09 ***
gW	0.101	0.542	0.19	0.8520
x:gW	2.075	0.634	3.27	0.0011 **

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 394.75 on 291 degrees of freedom

Residual deviance: 171.02 on 288 degrees of freedom

AIC: 179

Number of Fisher Scoring iterations: 8

```
> fit2 <- glm(y ~ x + g, data=dat, family="binomial")
> summary(fit2)
```

```

Call:
glm(formula = y ~ x + g, family = "binomial", data = dat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.765  -0.469   0.128   0.392   2.207

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    1.553      0.305    5.09 3.6e-07 ***
x              1.159      0.131    8.87 < 2e-16 ***
gW            -1.127      0.379   -2.98 0.0029 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 394.75  on 291  degrees of freedom
Residual deviance: 195.41  on 289  degrees of freedom
AIC: 201.4

Number of Fisher Scoring iterations: 6

> newdata <- data.frame(x=xVal, g=gVal)
> # für die log-odds
> predict(fit1, newdata=newdata, type="link")

      1
2.286

> # für die odds
> exp(predVals[1])

[1] 9.835

> # für die Wahrscheinlichkeit
> predict(fit1, newdata=newdata, type="response")

      1
0.9077

```

- (a) **True.** Im data frame dat sind 292 Datenpunkte enthalten.
- (b) **False.** Der wahre p-Wert ist 0.852.
- (c) **True.** Der wahre p-Wert ist 0.001062.
- (d) **True.** Die odds verändern sich um den Faktor $\exp(gW) = 0.3241$.
- (e) **False.** Der wahre Wert ist 9.8347.

3. Aufgabe

Bei einer Gruppe von Personen soll eine binäre Zielgrösse ($y = 0$ oder $y = 1$) durch eine erklärende Variable (x) und das Geschlecht ($g = "M"$ oder $g = "W"$) erklärt werden.

Die Daten sind im data frame dat in folgendem rda-File gespeichert: ueb266961.rda.

Passen Sie ein logistisches Regressionsmodell mit Wechselwirkung an.

Welche der folgenden Aussagen sind korrekt?

- (a) Im data frame dat sind 270 Datenpunkte enthalten.
- (b) Der Effekt der Gruppe ist signifikant.
- (c) Die Wechselwirkung von g und x ist signifikant.
- (d) Verwenden Sie nun die Logistische Regression mit der Formel $y \sim g + x$ (unabhängig von Ihren vorherigen Resultaten). Wenn man von der Gruppe der Frauen in die Gruppe der Männer wechselt, verändern sich die odds für $y = 1$ um den Faktor 0.528.
- (e) Verwenden Sie wieder die Logistische Regression mit der Formel $y \sim g + x$ (unabhängig von Ihren vorherigen Resultaten). Die log-odds für $y = 1$ für einen Mann mit $x = 1.4187$ werden als 2.3 vorhergesagt.

Lösung

Im folgenden Code bezeichnet `fname` den Dateinamen der jeweiligen `.rda`-Datei (in Anführungszeichen), `xVal` bezeichnet den neuen Wert für x und `gVal` den neuen Wert für g (Aufgabenteil e).

```
> load(fname)
> fit <- glm(y ~ x * g, data=dat, family="binomial")
> summary(fit)
```

Call:

```
glm(formula = y ~ x * g, family = "binomial", data = dat)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.225	-0.538	0.180	0.506	2.575

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.790	0.360	2.19	0.028 *
x	0.923	0.194	4.76	1.9e-06 ***
gW	-0.543	0.490	-1.11	0.268
x:gW	0.199	0.305	0.65	0.514

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 220.58 on 159 degrees of freedom
 Residual deviance: 118.28 on 156 degrees of freedom
 AIC: 126.3

Number of Fisher Scoring iterations: 5

```
> fit2 <- glm(y ~ x + g, data=dat, family="binomial")
> summary(fit2)
```

Call:

```
glm(formula = y ~ x + g, family = "binomial", data = dat)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.318	-0.567	0.160	0.499	2.453

Coefficients:

```

              Estimate Std. Error z value Pr(>|z|)
(Intercept)    0.863      0.362    2.38   0.017 *
x              1.013      0.150    6.74  1.6e-11 ***
gW            -0.639      0.473   -1.35   0.177
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 220.58 on 159 degrees of freedom
Residual deviance: 118.71 on 157 degrees of freedom
AIC: 124.7

```

Number of Fisher Scoring iterations: 5

```

> newdata <- data.frame(x=xVal, g=gVal)
> # für die log-odds
> predict(fit1, newdata=newdata, type="link")

1
2.3

> # für die odds
> exp(predVals[1])

[1] 9.974

> # für die Wahrscheinlichkeit
> predict(fit1, newdata=newdata, type="response")

1
0.9089

```

- (a) **False.** Im data frame dat sind 160 Datenpunkte enthalten.
- (b) **False.** Der wahre p-Wert ist 0.2678.
- (c) **False.** Der wahre p-Wert ist 0.5139.
- (d) **False.** Die odds verändern sich um den Faktor $\exp(-gW) = 1.8938$.
- (e) **True.** Der wahre Wert ist 2.3.