

Tests / Devianz

Aufgabe 1: Tests

An einer Münchner Schule wird ein Jahr lang täglich die Anzahl krankheitsbedingt fehlender Schüler ermittelt. Ferner werden täglich **Temperatur**, **Luftdruck** und **Witterung** gemessen. Nun sollen geeignete Modelle gefittet und verglichen werden.

Finden Sie heraus und begründen Sie, in welchen der folgenden Situation ein likelihood-basierter Test möglich ist. Wann kann ein AIC-Vergleich vorgenommen werden?

Situation 1: Zunächst fitten Sie ein Poisson-Modell unter Berücksichtigung aller Einflussgrößen, anschließend nehmen Sie die Variable **Luftdruck** aus dem Modell heraus.

Situation 2: Sie fitten ein Poissonmodell unter Berücksichtigung aller Einflussgrößen und wollen es mit einem reinen Intercept-Modell vergleichen.

Situation 3: Zunächst fitten Sie ein Poisson-Modell mit Id-Link, anschließend mit natürlichem Link. Dabei werden jeweils alle drei Einflussgrößen beachtet.

Situation 4: Betrachten Sie wiederum Situation 3, wobei nun beim Fit des Poissonmodells mit natürlichem Link dem **Luftdruck** keine Rechnung mehr getragen werden soll.

Situation 5: Sie fitten ein Poissonmodell und ein klassisches lineares Modell - beide mit natürlichem Link und unter Berücksichtigung aller drei Einflussgrößen.

Situation 6: Betrachten Sie wiederum Situation 5, wobei nun beim Fit des klassischen linearen Modells ebenfalls der Log-Link verwendet wird.

Situation 7: Betrachten Sie wiederum Situation 5, wobei nun beim Fit des klassischen linearen Modells die Variable **Luftdruck** aus dem Modell genommen wird.

Aufgabe 2: Poissonmodelle, Tests, Devianz

Im vorliegenden Poissonmodell wird die Faktorvariable **visits** (Anzahl Arztbesuche pro Jahr) durch **gender** ("male" und "female"), **income** (metrisch: Einkommensklasse), **illness** (metrisch), **reduced** (metrisch), **health** (metrisch) und **freepoor** ("no" und "yes") erklärt.

```
glm(formula = visits ~ . + freepoor:income,  
family = "poisson", data = DoctorVisits)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.8635	-0.6748	-0.5840	-0.4925	5.6146

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.871070	0.076621	-24.420	< 2e-16	***
genderfemale	0.201205	0.054543	3.689	0.000225	***
income	-0.216862	0.078901	-2.749	0.005986	**
illness	0.206153	0.017327	11.898	< 2e-16	***
reduced	0.131131	0.004840	27.095	< 2e-16	***
health	0.029825	0.009898	3.013	0.002584	**

```
freepooryes      0.307779   0.252737   1.218 0.223307
income:freepooryes -3.794185   1.050238  -3.613 0.000303 ***
```

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 5634.8 on 5189 degrees of freedom
Residual deviance: 4381.9 on 5182 degrees of freedom
AIC: 6729.5
```

Number of Fisher Scoring iterations: 6

Beantworten Sie folgende Fragen:

- Wie lautet die zugrundeliegende Modellgleichung in Link- und Responseform? Warum wäre die Verwendung eines klassischen linearen Modells hier nicht anzuraten?
- Interpretieren Sie den Einfluss des Haupteffektes `income`!
- An welcher/-n Stelle(n) könnte dieses Modell eventuell reduziert werden und warum?

Als Alternative steht Ihnen folgendes Modell zur Verfügung:

```
glm(formula = visits ~ . , family = "poisson",
data = DoctorVisits)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.8674	-0.6757	-0.5805	-0.4889	5.6263

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.856158	0.076572	-24.241	< 2e-16	***
genderfemale	0.203423	0.054596	3.726	0.000195	***
income	-0.243983	0.078924	-3.091	0.001992	**
illness	0.206199	0.017331	11.898	< 2e-16	***
reduced	0.130270	0.004834	26.946	< 2e-16	***
health	0.030290	0.009898	3.060	0.002213	**
freepooryes	-0.586256	0.172844	-3.392	0.000694	***

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 5634.8 on 5189 degrees of freedom
Residual deviance: 4398.2 on 5183 degrees of freedom
AIC: 6743.7
```

Number of Fisher Scoring iterations: 6

- Testen Sie zum Niveau $\alpha = 5\%$, ob dieses Modell im Vergleich zum vorherigen noch ausreichend informativ ist. Sprechen die AICs für Ihr Testergebnis?