

NLNP Praktikum 5

Robin Baudisch, Merlin Kopfmann, Maximilian Neudert

Inhaltsverzeichnis

A1	2
a)	2
b)	2
c)	3
d)	4
A2	5
a)	5
b)	6
c)	6
d)	6

```
load("awards.RData")
load("DebTrivedi.RData")
```

A1

a)

```
data = awards

# poisson regression
plm = glm(num_awards ~ prog + math, data = data, family = poisson)
summary(plm)
```

Call:

```
glm(formula = num_awards ~ prog + math, family = poisson, data = data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.2043	-0.8436	-0.5106	0.2558	2.6796

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-5.24712	0.65845	-7.969	1.60e-15 ***
progAcademic	1.08386	0.35825	3.025	0.00248 **
progVocational	0.36981	0.44107	0.838	0.40179
math	0.07015	0.01060	6.619	3.63e-11 ***

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 287.67 on 199 degrees of freedom

Residual deviance: 189.45 on 196 degrees of freedom

AIC: 373.5

Number of Fisher Scoring iterations: 6

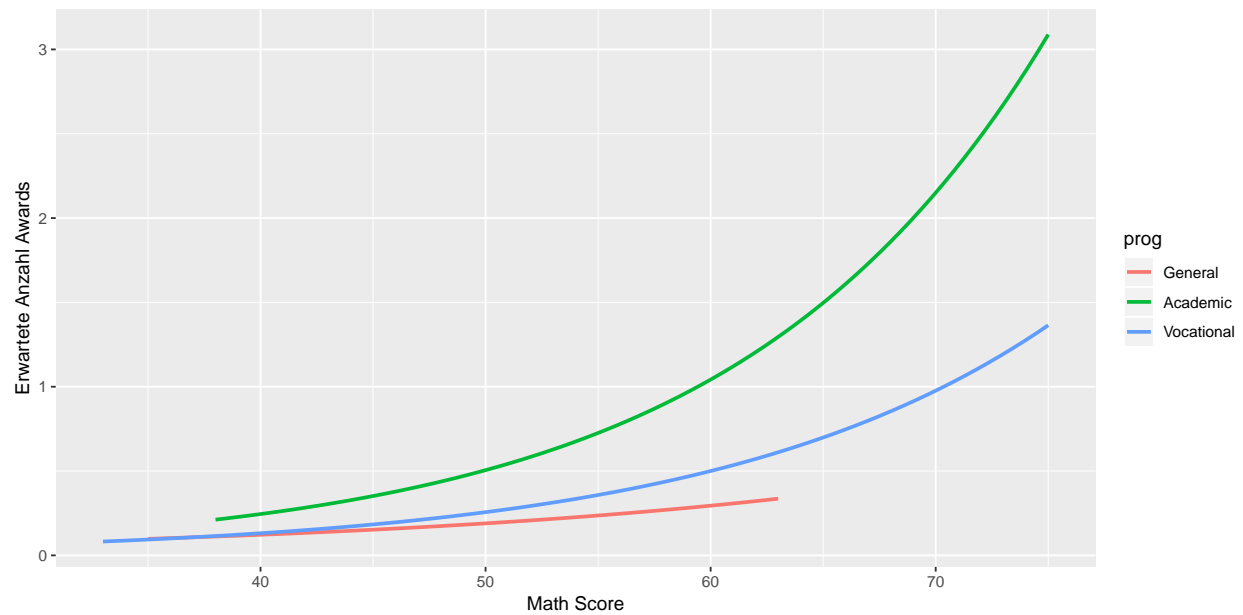
Die Summary zeigt die Poisson-Regressionskoeffizienten für jede der Variablen sowie die Standardfehler, Z-Scores, p-Werte und 95% Konfidenzintervalle für die Koeffizienten. Der Koeffizient für math liegt bei 0.07. Die Variable progAcademic vergleicht zwischen prog = "Academic" und prog = "General" mit einem Koeffizienten von 1.08. Die Variable prog.Vocational zeigt die erwartete Differenz in der Anzahl zwischen prog = "Vocational" und der Referenzgruppe (prog = "General") mit einem Koeffizienten von 0.37.

b)

```
library(ggplot2)

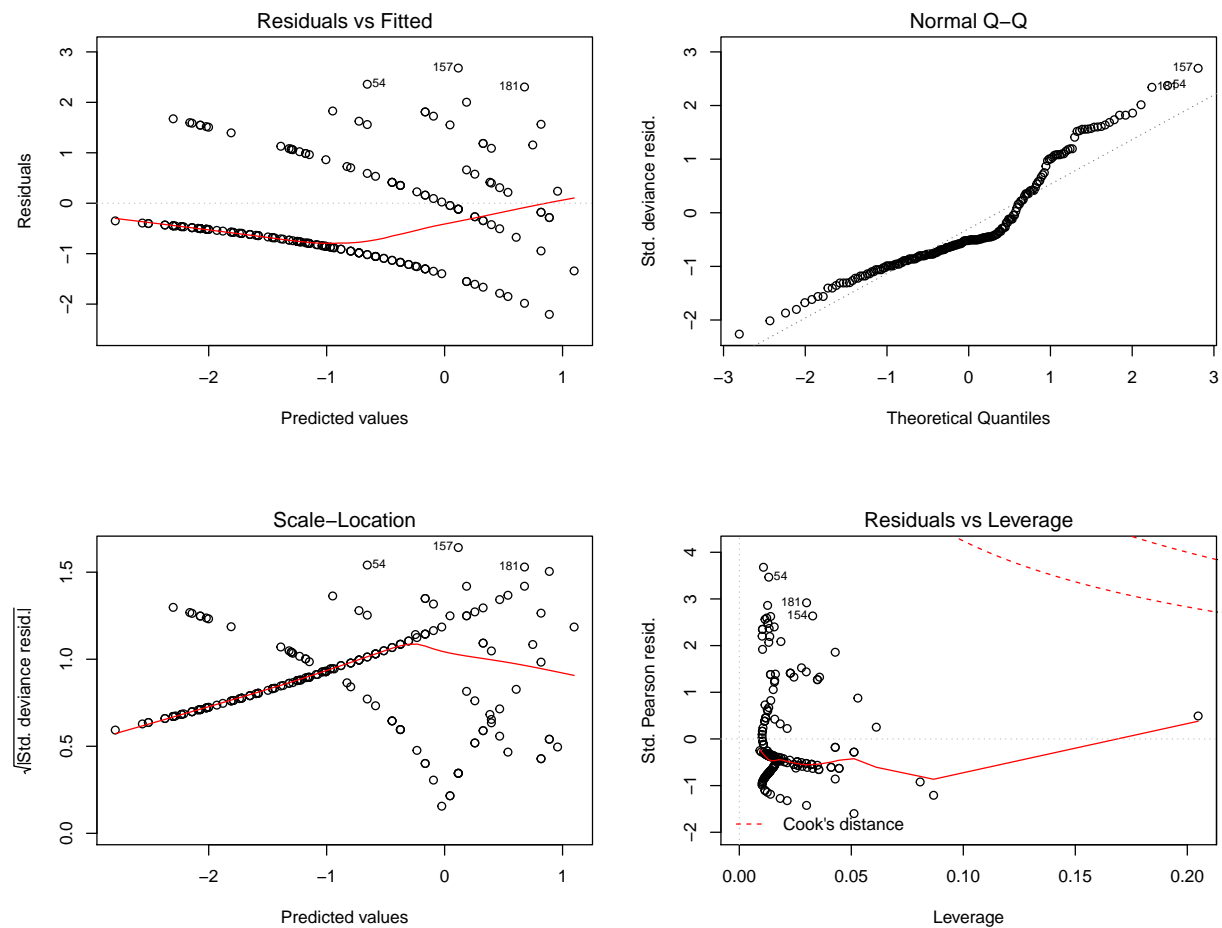
pred = predict.glm(plm, type = "response")

gg = ggplot(data = data, mapping = aes(x = math, y = num_awards, color = prog))
# gg = gg + geom_point()
gg = gg + geom_smooth(method = "glm", method.args = list(family = "poisson"),
  se = FALSE)
gg = gg + labs(x = "Math Score", y = "Erwartete Anzahl Awards")
gg
```



c)

```
par(mfrow = c(2, 2))
plot(plm)
```



Da Academics eine signifikante statistische Auswirkung hat, findet man an den in den Residuenplots auch Muster eines Offsets wieder. In den Plots Residuals vs Fitted und Scale Location lässt sich die Zuordnung der Werte in die einzelnen Prog Gruppen erkennen.

d)

```
lm = lm(num_awards ~ math + prog, data = data)
```

```
pAIC <- AIC(plm)
```

```
lAIC <- AIC(lm)
```

Poisson AIC	Linear AIC
373.5	532.25

Es zeigt sich ein niedrigerer AIC Wert bei der Poisson Regression als bei der linearen Regression. Ein niedriger AIC deutet auf ein besseres Modell hin als ein hoher AIC. Folglich ist die Poisson Regression als ein besseres Modell zu betrachten.

A2**a)**

```
load("DebTrivedi.RData")
plm = glm(ofp ~ health + numchron + hosp + married + medicaid, data = DebTrivedi,
          family = poisson)
summary(plm)
```

Call:

```
glm(formula = ofp ~ health + numchron + hosp + married + medicaid,
     family = poisson, data = DebTrivedi)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-5.2623	-2.0484	-0.6898	0.7949	16.1776

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.490539	0.013363	111.538	<2e-16 ***
healthpoor	0.338277	0.016536	20.457	<2e-16 ***
healthexcellent	-0.372756	0.030241	-12.326	<2e-16 ***
numchron	0.095171	0.006124	15.541	<2e-16 ***
hosp	0.500663	0.013982	35.809	<2e-16 ***
marriedyes	-0.019759	0.012860	-1.537	0.124
medicaidyes	0.032753	0.021101	1.552	0.121

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 26943 on 4405 degrees of freedom

Residual deviance: 24179 on 4399 degrees of freedom

AIC: 36968

Number of Fisher Scoring iterations: 5

Die Summary zeigt die Poisson-Regressionskoeffizienten für jede der Variablen sowie die Standardfehler, Z-Scores, p-Werte und 95% Konfidenzintervalle für die Koeffizienten.

„poor health“ und „excellent health“ sind jeweils neue Faktoren als Verhältnis zu „average health“.

Es zeigt sich dass „poor health“ positiv mit Arztbesuchen korreliert ist und „excellent health“ negativ korreliert. Zusätzlich weisen sowohl chronische Erkrankungen als auch die Notwendigkeit eines Krankenhausaufenthalts eine positive Korrelation auf.

Die Parameter „verheiratet“ und „staatliche Unterstützung“ weisen keine signifikante statistische Auswirkung auf.

b)

```

btx = plm$coefficients[1] + 2 * plm$coefficients[4] + plm$coefficients[6]
y = exp(btx)
y = round(y, 0)

```

Für die Poisson Regression gilt $E(Y_i|x_i) = \exp(b^T x)$, also Berechnen wir $b^T x$ mit den Koeffizienten aus dem Modell und exponieren anschließend für den Erwartungswert.

Man kann 5 Arztbesuche erwarten.

c)

	hosp.0	hosp.1
poor	2578	2351
average	14014	5323
excellent	1018	158

	hosp.0	hosp.1
poor	7.032564	11.814785
average	4.920448	8.266407
excellent	3.236101	5.436685

chisq
44.29864

- H_0 : {Erwartete und beobachtete Häufigkeiten sind gleichverteilt}
- H_1 : {Erwartete und beobachtete Häufigkeiten sind nicht gleichverteilt}

Der errechnete Wert des Chi-Quadrat Anpassungstests liegt über 11.07 und damit im Ablehnungsbereich. H_0 kann also verworfen und H_1 angenommen werden.

d)

```

par(mfrow = c(2, 2))
plot(plm)

```

