

Übungen zu Ökonometrie - Blatt 6

(Abgabe: Donnerstag, 22.06.2011, vor den Übungen)

Aufgabe 1 (8 Punkte)

Es sei ein verallgemeinertes lineares Regressionsmodell mit natürlicher Linkfunktion gegeben. Wir wollen untersuchen, wie sich der Erwartungswert der Zielvariablen unter kleinen Änderungen der einzelnen erklärenden Variablen verhält. Dazu müssen wir

$$\frac{\partial \mathbb{E}(Y)}{\partial x_i} = \frac{\partial g^{-1}(x^T \beta)}{\partial x_i} = \frac{\partial g^{-1}(\beta_1 + \beta_2 x_2 + \dots + \beta_m x_m)}{\partial x_i}, \quad i = 2, \dots, m$$

berechnen.

Führe eine solche Analyse für

- (a) eine gammaverteilte Zielvariable mit (unbekanntem) Parameter $b > 0$ und bekanntem Parameter $\lambda_0 > 0$ (4)
- (b) eine Poisson-verteilte Zielvariable mit (unbekanntem) Parameter λ (4)

durch und interpretiere das Ergebnis. (Hinweis: Die natürliche Linkfunktion im Falle der Gammaverteilung wurde in Aufgabe 3, Blatt 4 berechnet.)

Aufgabe 2 (12 Punkte)

Auf der Homepage der Vorlesung befindet sich die Datei `apprentice.dat` mit Daten über die Anzahl der Lehrlinge, die für ihre Ausbildung von anderen schottischen Regionen nach Edinburgh gezogen sind. Es gibt insgesamt 33 Beobachtungen für 5 Variable plus die Namen der Regionen in der (unbenannten) Spalte 1:

- Distance: Die Entfernung von Edinburgh (die Einheit ist unbekannt, vermutlich Meilen)
- Apprentices: Anzahl der Lehrlinge, die von der angegebenen Region nach Edinburgh gezogen sind
- Population: Die Bevölkerung (in Tausend) der angegebenen Region
- Urbanization: Der Grad der Urbanisierung, gemessen anhand des Prozentsatzes der Bevölkerung, der in städtischen Gebieten wohnt
- Location: Die Himmelsrichtung der Region relativ zu Edinburgh: Entweder Norden, Westen oder Süden.

	Distance	Apprentices	Population	Urbanization	Location
Midlothian	21	225	56	18.8	South
WestLothian	24	22	18	37.9	West
⋮	⋮	⋮	⋮	⋮	⋮

- (a) Betrachte das verallgemeinerte lineare Regressionsmodell mit der Poissonverteilung für die Zielvariable "Apprentices" und natürlicher Linkfunktion. Regressiere "Apprentices" auf die Variablen "Distance", "Population", "Urbanization" und "Location" und interpretiere den Output der R-Funktion `summary()` nach Anwendung auf das Ergebnis von `glm()`. Lese hierzu die Daten mit dem Befehl (4)

```
> daten <- read.table("C:\\...\\apprentice.dat", header=TRUE)
```

ein. Benutze für die Interpretation auch das Ergebnis aus Aufgabe 1.

- (b) Diskutiere, ob es Sinn machen könnte, gewisse erklärende Variablen aus dem Modell zu entfernen. (2)
- (c) Wir ändern nun in dem Datensatz bei allen Zeilen, bei denen die Region "West" ist, den Wert auf "North": (2)

```
> detach(daten)
> daten[daten[,5]=="West",5] <- "North"
> attach(daten)
```

Wir machen also aus den drei Gruppen "South", "West" und "North" zwei Gruppen, wobei "North" für alle Regionen steht, die nicht im Süden liegen.

Entscheide mit Hilfe des AIC-Kriteriums, ob diese Abänderung im Datensatz zur Verbesserung des Modells beiträgt.

- (d) Teste mit Hilfe des Likelihood-Quotiententests aus der Vorlesung, ob mindestens eine der erklärenden Variablen im "neuen" Datensatz von Teil (c) Einfluss auf die Anzahl der Lehrlinge hat und interpretiere das Testergebnis. (2)
- (e) Welche Anzahl von Lehrlingen, die nach Edinburgh ziehen, erwartet man, wenn die Entfernung der Region, aus der sie kommen, 50 beträgt, die Bevölkerung der Region 40000 ist, der Grad der Urbanisierung 50% beträgt und die Himmelsrichtung "South" ist? Verwende hierzu das Modell, das aus dem modifizierten Datensatz aus Teil (c) hervorgeht. (Hinweis: R kodiert die Variable "Location" mit "South"=0 und "North"=1) (2)