Lennart Oelschläger (lennart.oelschlaeger@uni-bielefeld.de)
310214 Praktische Übung zu "Einführung in die Mikroökonometrie"
Wintersemester 2023/2024

Aufgabenblatt 5

Überblick

In einem *binären Wahlmodell* ist die abhängige Variable binär. Wir modellieren sie nicht direkt, sondern stattdessen die Wahrscheinlichkeit, dass sie den Wert 1 annimmt (dass sie den Wert 0 annimmt, folgt dann direkt aus der Gegenwahrscheinlichkeit). Lineare Modelle weisen hier Nachteile auf: Prognosen können negativ sein oder oberhalb von 1 liegen und es liegt zwangsläufig Heteroskedastizität vor. Als bessere Alternative gibt es *Logit-* und *Probit-Modelle*, sie werden mit der Maximum-Likelihood Methode geschätzt. Mit *ROC* Kurven können die Klassifizierungsgüten verschiedener Modelle verglichen werden.

Und dann schauen wir uns noch eine neue Datensituation an, nämlich dass die abhängige Variable eine *Zählvariable* ist. In diesem Fall können wir die *Poisson-Regression* verwenden, um Einflüsse zu quantifizieren.

Aufgaben

1. Simulieren Sie Daten wie nachfolgend angegeben und schätzen Sie die Probit-Modelle

$$y = \beta_0 + \beta_1 x_1 + u \tag{1}$$
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{2}$$

   mit der Funktion `glm()`.

```
set.seed(1)
data <- MASS::mvrnorm(n = 50, mu = c(0, 0, 0), Sigma = diag(3))
data <- as.data.frame(data)
colnames(data) <- c("x_1", "x_2", "u")
beta_0 <- 1
beta_1 <- 0.2
beta_2 <- -0.5
y <- beta_0 + data$x_1 * beta_1 + data$x_2 * beta_2 + data$u
data$y <- ifelse(y >= 0.5, 1, 0)
```

2. Welches der Modelle (1) und (2) hat die "bessere" Null Deviance beziehungsweise Residual Deviance und was folgt daraus?
3. Bitte testen Sie für beide Modelle die Hypothese $H_0\colon \beta_k = 0, k = 1, \ldots, K$ zu $\alpha = 10\%$.
4. Zeichen Sie für beide Modelle die ROC Kurve mit der Funktion `pROC::roc()`. Welches Modell hat demnach die bessere Prognosequalität?
5. Es seien $y_1, \ldots, y_n$ unabhängig Poisson verteilt zum Parameter $\lambda > 0$. Zeigen Sie, dass $\bar{y}$ der MLE für λ ist.

6. Zeigen Sie, dass sowohl der Erwartungswert als auch die Varianz der Poisson-Verteilung jeweils der Parameter λ ist.
7. Für gegebene Beobachtungen y und X wollen wir $\lambda = \mathbb{E}(y \mid X)$ als Linearkombination der Regressoren X modellieren. Wie stellen wir sicher, dass $\lambda > 0$ erfüllt bleibt? Wie hoch ist gemäß dieses Modells dann die Wahrscheinlichkeit, dass $y = k$ ist für $k = 0, 1, 2, \dots$?
8. Nutzen Sie den Datensatz `wooldridge::crime1` und modellieren Sie die Anzahl `narr86` an Inhaftierungen durch das Einkommen `inc86`. Wie können Sie den Einkommenskoeffizienten interpretieren?