

Aufgaben zur Vorlesung

Multivariate Verfahren

Übungsblatt VIII

1. Beantworten Sie folgende Fragen bzw. bearbeiten Sie folgende Arbeitsanweisungen:
 - (a) Erklären Sie drei wesentliche Schwächen des linearen Wahrscheinlichkeitsmodells.
 - (b) Zeigen Sie: Für das Logit Modell gilt $P(y = 0|x) = \frac{1}{1+e^x}$.
 - (c) Wie lauten die Likelihoodfunktion und log-Likelihoodfunktion für das Logit bzw. Probit Modell?
 - (d) Was testet der Likelihood-Ratio-Test?
2. Eine Bank möchte ein Modell für die Wahrscheinlichkeit, dass ein Kunde im Bestand einen Kredit für den Hausbau beantragt. Als treibende Variablen für einen Hausbau ergab eine Umfrage bei den Kundenberatern der Bank das Nettoeinkommen des Kunden und ob der Kunde Kinder hat oder nicht. Die Bank modelliert darauf ein lineares Wahrscheinlichkeitsmodell mit folgendem Ergebnis:

$$P(y = 1|x') = 0.5 + 0.0001x_1 + 0.1x_2,$$

mit $y = 1$, falls ein Kredit für den Hausbau beantragt wurde, und $y = 0$ sonst sowie x_1 als Nettoeinkommen des Kunden in EUR und $x_2 = 1$, falls der Kunde Kinder hat, und $x_2 = 0$, falls nicht.

- (a) Berechnen Sie die Wahrscheinlichkeit für $y = 1$ für einen kinderlosen Kunden mit einem Nettoeinkommen von $x_1^* = 0$, $x_1^{**} = 1000$ und $x_1^{***} = 7000$. Was sind die jeweiligen Wahrscheinlichkeiten für $y = 0$?
 - (b) Sind die Werte aus (a) aus Ihrer Sicht plausibel? Was würden Sie der Bank empfehlen, welches Modell Sie anstelle des linearen Wahrscheinlichkeitsmodells wählen sollte?
3. Einen Monat später ist die Bank aus Aufgabe 2 nach dem Einsatz einer Beratungsfirma zu dem Schluss gekommen, dass das lineare Wahrscheinlichkeitsmodell nicht adäquat für die Modellierung des Sachverhalts ist. Die Statistikexperten der Bank schätzen daraufhin folgendes Probit-Modell:

$$P(y = 1|x') = \Phi(1.96 + 0.0005x_1 + 0.75x_2).$$

Die Notation ist wie in Aufgabe 2. Die Standardfehler der geschätzten Regressionskoeffizienten sind 0.5, 0.0001 und 1.

- (a) Sind die Regressionskoeffizienten zum Niveau $\alpha = 0.05$ signifikant?

- (b) Der Wert der log-Likelihood-Funktion aus der Schätzung beträgt $\ln L(\hat{\beta}) = -20.88$. Es flossen insgesamt 1000 Kundendatensätze in die Schätzung ein, von denen 975 einen Kredit beantragt haben. Führen Sie einen Likelihood-Ratio-Test durch.
4. Betrachten Sie nochmals Beispiel 8.1 aus dem Skript.
- (a) Leiten Sie eine Formel für die Odds und log-Odds her.
- (b) Sei nun $x_1 = (1, 1000, 1, 0)$. Berechnen Sie $P(y = 1|x'_1)$. Wie groß sind die Odds? Interpretieren Sie diese. Wie groß sind die log-Odds?
5. Rechnerübung: Verwenden Sie den Datensatz *psi* aus Ilias. Der Datensatz enthält Werte einer Studie zu Schülern im Bildungssystem PSI (Personalized System of Instruction). Schätzen Sie ein Logit-Modell auf der abhängigen Variablen „GRADE“, die angibt, ob sich die Note des Schülers verbessert hat (Kodierung 1) oder nicht (Kodierung 0). Verwenden Sie als Regressoren einen konstanten Term, sowie die Variablen „PSI“ (Dummyvariable zur Teilnahme am PSI), „TUCE“ (Resultate eines Vortests) und „GPA“ (Grade Point Average, entspricht der deutschen Abiturnote).