Prof. Dr. E. Spodarev
Judith Olszewski

SoSe 2015
25.06.2015

Statistik-Praktikum/WiMa-Praktikum II - Übungsblatt 9

Vorstellung der Ergebnisse in der Übung am 02.07.2015

Aufgabe 1

Es soll der lineare Zusammenhang zwischen der Nettomiete und den Eigenschaften einer Wohnung ermittelt werden. Lade dazu die Datei 'miete03.txt' von der Homepage herunter. Hier eine Beschreibung der Daten:

> Zahlreiche deutsche Städte erstellen sogenannte Mietspiegel, um Mietern, Vermietern, Mietberatungsstellen und Sachverständigen eine 'objektive' Entscheidungshilfe in Mietfragen zur Verfügung zu stellen. Die Mietspiegel werden dabei insbesondere zur Ermittlung der 'ortsüblichen Vergleichsmiete' (Nettomiete in Abhängigkeit von Wohnungsgröße, -ausstattung, -alter, etc.) herangezogen. Bei der Erstellung von Mietspiegeln wird aus der Gesamtheit aller in Frage kommenden Wohnungen eine repräsentative Zufallsstichprobe gezogen (im Fall der Stadt München durch Infratest), und die interessierenden Daten werden von Interviewern anhand von Fragebögen ermittelt. Der vorliegende Datensatz stellt einen Ausschnitt aus dem Mietspiegel München des Jahres 2003 dar und enthält die Daten von 2053 Wohnungen.

Die im Datensatz enthaltenen Variablen bezeichnen die folgenden Größen: **nm**: Nettomiete in EUR; **nmqm**: Nettomiete pro m^2 in EUR; **wfl**: Wohnfläche in m^2; **rooms**: Anzahl der Zimmer; **bj**: Baujahr; **bez**: Stadtbezirk; **wohngut**: Gute Wohnlage? (J=1, N=0); **wohnbest**: Beste Wohnlage? (J=1, N=0); **ww0**: Warmwasserversorgung vorhanden? (J=0, N=1); **zh0**: Zentralheizung vorhanden? (J=0, N=1); **badkach0**: Gekacheltes Badezimmer? (J=0, N=1); **badextra**: Besondere Zusatzausstattung im Bad? (J=1, N=0); **kueche**: Gehobene Küche? (J=1, N=0).

Betrachte den folgenden linearen Zusammenhang:

$$\begin{aligned}\text{nm} = {} & \beta_0 + \beta_1 \text{nmqm} + \beta_2 \text{wfl} + \beta_3 \text{rooms} + \beta_4 \text{bj} + \beta_5 \text{wohngut} + \beta_6 \text{wohnbest} + \beta_7 \text{ww0} \\ & + \beta_8 \text{zh0} + \beta_9 \text{badkach0} + \beta_{10} \text{badextra} + \beta_{11} \text{kueche}\end{aligned}$$

Führe eine multivariate Regressionsanalyse durch. Nimm dabei an, dass die Designmatrix vollen Rang besitzt. Teste außerdem, ob die Variablen wohngut und zh0 überhaupt relevant sind. Speichere das Ergebnis in der Datei 'regression.pdf'.
Hinweis: Für die Interpretation der Ergebnisse lohnt sich ein Blick auf die Hilfeseite der Prozedur REG (Displayed Output).

Aufgabe 2

Auf der Homepage findet sich die Datei 'challenger.txt', die folgende Werte enthält:

Temperatur	53	57	58	63	66	67	67	67	68	69	70	70
Ausfall	1	1	1	1	0	0	0	0	0	0	0	0
Temperatur	70	70	72	73	75	75	76	76	78	79	81	
Ausfall	1	1	0	0	0	1	0	0	0	0	0	

Die Variable ‘Temperatur’ ist die Außentemperatur (in Fahrenheit) beim Start der 23 Space-Shuttle-Flüge vor der Challenger-Katastrophe und die Variable ‘Ausfall’ gibt an, ob mindestens einer der Dichtungsringe wegen Materialermüdung ausgefallen ist (1) oder nicht (0).

(a) Untersuche mit Hilfe eines logistischen Regressionsmodells (Logit-Modell) den Einfluss der Temperatur auf das Auftreten solcher Materialermüdungserscheinungen. (Die Angaben, die SAS über die ‘Tolerance Distribution’ macht, müssen nicht näher analysiert werden.)

(b) Zeichne die Messdaten sowie die Kurve der laut Modell geschätzten Wahrscheinlichkeiten in Abhängigkeit von der Temperatur in ein gemeinsames Schaubild. (Tipp: GPLOT, OVERLAY)

(c) Welche Wahrscheinlichkeit wird für das Versagen mindestens eines Dichtungsringes prognostiziert, wenn die Außentemperatur wie am Unglückstag 31 °F beträgt? (Tipp: Dafür wird ein DATA-Step benötigt. Kann alternativ auch per Hand berechnet werden)

(d) Wiederhole Teil (a) und (b) mit einem entsprechenden Probit-Modell.

Nutze folgenden Beispielcode als Hilfestellung:

In der Prozedur REG kann durch das Statement TEST ein Test auf gemeinsame Relevanz verschiedener Variablen durchgeführt werden.

Für die Berechnung mit dem Logit/Probit-Modell sollten trials angelegt werden, damit SAS merkt, dass es sich um numerische Werte handelt. Sonst sieht es ‘1’ und ‘0’ als Kategorien wie ‘gut’ und ‘schlecht’. Das wäre nicht weiter schlimm, bedeutet aber, dass z. B. bei Schaubildern die Bedeutung von 0 und 1 vertauscht wird. (Das liegt daran, dass zuerst eine 1 auftritt, deshalb kommt die Kategorie ‘1’ vor Kategorie ‘0’.) Nutze außerdem das Statement OUTPUT um einen Datensatz mit den Ausgangsvariablen und den laut angepasstem Modell geschätzten Wahrscheinlichkeiten zu erstellen.

```
DATA ...
    trials=1;
RUN;

PROC PROBIT;
    MODEL .../trials=... / DISTRIBUTION=...;
    OUTPUT OUT=... P=...;
RUN;
```

Mehrere Plots können z.B. wie folgt erzeugt werden:

```
PROC GPLOT DATA=...;
    PLOT (a b)*c / OVERLAY;
RUN;
```