

StatwR Abschluss Klausur WS 2007-2008.

K. Utikal

1. Angenommen die Beobachtungen (y_i, x_i) , $i = 1, \dots, n$ stammen aus Pop(ulation)1 und das Modell

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

mit $\varepsilon_1, \dots, \varepsilon_n$ uiv $N(0, \sigma^2)$ gilt.

- a) Wie kann man am besten einen Ausreisser unter den Beobachtungen erkennen?
 - b) Welche Probleme kann ein Ausreisser verursachen?
 - c) Was kann man dagegen unternehmen?
2. Angenommen die Beobachtungen (y_i, x_i) , $i = n+1, \dots, n+m$ stammen aus Pop(ulation)2, sind unabhängig von den Beobachtungen aus Pop1, und das Modell

$$y_i = \delta + \beta x_i + \varepsilon_i$$

mit $\varepsilon_{n+1}, \dots, \varepsilon_{n+m}$ uiv $N(0, \sigma^2)$ gilt. Gegeben die (Null-) Hypothese

$$H_0 : \alpha = \delta$$

- a) Geben Sie ein konkretes Beispiel in dem diese Problem auftreten könnte.
 - b) Wie könnte man diese Hypothese testen? (Sie brauchen die test statistic nicht explizit aufschreiben, Tipp: Indikator-Variablen).
 - c) Verallgemeinerung: Wie könnte man testen ob die y-AchsenAbschnitte und die Steigungen in pop1 und pop2 gleich sind? Geben Sie ein geeignetes Modell, die NullHypothese und beschreiben Sie informell einen geeigneten Test.
3. Beobachtungen $(y_1, x_1), \dots, (y_n, x_n)$, y binär.
- a) Schreiben Sie das LogitModell auf .
 - b) Geben Sie eine Interpretation der Parameter.
 - c) Wann wäre es natürlicher, ein ProbitModell zu benutzen? (Tipp: latente Varibale) Geben Sie eine Motivation für das ProbitModell. (Sie brauchen keine mathematische Herleitung des Probit Modells zu unternehmen). Geben Sie ein Beispiel.
 - d) Wie würden Sie z.B. in einer Simulation Daten von einem Probit Modell erzeugen? (Es ist nicht unbedingt erforderlich, ein formales R-Programm zu schreiben).

4. Daten über AltersGruppe und KaufEntscheidung sind in der KontingenzTafel zusammengefasst.

	20-30	31-50	51-65
kein Kauf
Kauf

Wir wollen den Einfluss des Alters auf eine KaufEntscheidung untersuchen.

- a) Schreiben Sie ein geeignetes LogitModell.
- b) Wie könnten wir diesen Zusammenhang Testen? Stellen Sie die NullHypothese auf. Sie können entweder einen geeigneten Test beschreiben oder ein geeignetes R-Kommando angeben und erklären welches diesen Test implementiert.

Kommentare:

Aufgabe1: Man muss zwischen Ausreißern und einflussreichen Werten unterscheiden. Manchmal sind Ausreißer einflussreich aber manche Beobachtungen sind auch einflussreich wegen des Wertes der unabhängigen Variable. Einflussreiche Werte kann man mit Cook's distance erkennen. Ausreißer sieht man schon im StreuPlot oder besser noch im Residuen-Plot oder besser noch im standardisierten ResiduenPlot.

Was kann man in Anwesenheit von Ausreißern tun? Dafür haben wir zB die LAD-Regression.

Aufgabe2: Das beiden Pops gemeinsame Modell ist

$$y = \alpha_0 + \alpha_1 I(pop2) + \beta x + \varepsilon$$

wobei $I(pop2)$ die PopulationsIndikatorVariable ist. Auf einen gemeinsamen y-AchsenAbschnitt testen sie durch

$$H_0 : \alpha_1 = 0$$

mit einem t-test (im Output aufgeführt) Allgemeiner:

$$y = \alpha_0 + \alpha_1 I(pop2) + (\beta_0 + \beta_1 I(pop2)) x + \varepsilon$$

Auf gemeinsame Steigung testen sie mit

$$H_0 : \beta_1 = 0$$

und auf ein gemeinsames Model mit

$$H_0 : \alpha_1 = \beta_1 = 0$$

Diese Hypo können Sie mit dem F-test (**anova**) testen.

Aufgabe 3: Im LogitModell können sie die β -Koeffs interpretieren

1. als Änderung im LogOddsRatio ($\beta\Delta x$)
2. als Multiplum der Odds ($e^{\beta\Delta x}$)
3. als Prozentweise Änderung der Odds ($e^{\beta\Delta x} - 1 \approx \beta\Delta x$) wenn $\beta\Delta x$ klein.

Aufgabe4: Anwendung des LogitModells auf KontingenzTafeln:

$$\log Odds = \alpha + \beta_1 I(akI) + \beta_2 I(akII)$$

Teste auf AltersUnabhängigkeit:

$$H_0 : \beta_1 = \beta_2 = 0$$