

## Angewandte Statistik für Biometrie

(Abgabe: Di., 07.06.2011, vor den Übungen)

1. Lies die Datensätze `data8a.dat`, `data8b.dat` und `data8c.dat` von der Homepage mit R ein. Sie enthalten in der zweiten Spalte Messzeitpunkte  $t_i$  und in der ersten Spalte die zu den Zeitpunkten  $t_i$  gemessenen Werte  $X_i$ . Finde heraus nach welchen Modellen die Datensätze erzeugt wurden und begründe Deine Aussage. Verwende zur Begründung (u.a.) die Ergebnisse von `lm()`, Plots und QQ-Plots der Residuen, sowie die Ergebnisse von Durbin-Watson-Tests der Residuen auf Autokorrelation.

*Hinweis:* Die Funktion `dwtest()` führt einen Durbin-Watson Test durch. Sie ist im Paket `lmtest` enthalten, das evtl. zuerst installiert werden muss. Es kann im Internet unter <http://www.r-project.org> heruntergeladen oder direkt in R über das Menü „Pakete und Daten“ (bzw. „Pakete“ unter Windows) oder mit dem Befehl `install.packages()` installiert werden. Im Skript wird das installierte Paket dann mit `library(lmtest)` geladen.

(6 Punkte)

2. Die Ergebnisse der letzten Klausur wurden analysiert. Dabei wurde der Zusammenhang zwischen dem Bestehen der Klausur ( $X_i = 1$  falls Student  $i$  bestanden hat,  $X_i = 0$  falls nicht) und den zuvor erreichten Übungspunkten ( $t_{i,1}$ ) sowie dem Lernaufwand in Stunden ( $t_{i,2}$ ) mittels binärer Regression betrachtet. Als Link-Funktion wurde die logit-Funktion  $g(y) = \frac{e^y}{1+e^y}$  gewählt, es handelt sich also um das Modell

$$X_i = g(\beta_1 + \beta_2 t_{i,1} + \beta_3 t_{i,2}) + \varepsilon_i$$

mit  $E(\varepsilon_i) = 0$ . Die Regression ergab  $\hat{\beta} = (-2.94, 0.025, 0.12)^\top$ .

- (a) Wie hoch sind damit die Wahrscheinlichkeiten die Klausur zu bestehen bei folgenden vier Strategien:
  - Weder Bearbeitung der Übungsaufgaben noch Vorbereitung auf die Klausur
  - Nichts gelernt, aber 140 Punkte in den Übungen.
  - Ignorieren der Übungen, aber 25h gelernt
  - 92 Übungspunkte und 21h Klausurvorbereitung
- (b) Finde eine Funktion  $f$  mit der man den Lernaufwand  $t_{i,2}$  in Abhängigkeit der Übungspunkte  $t_{i,1}$  berechnen kann (also  $f(t_{i,1}) = t_{i,2}$ ), um mit einer Wahrscheinlichkeit von  $p_i = 0.95$  zu bestehen.

(3+3 Punkte)