

STATISTIK 2 (107.325) WS 2010
COMPUTERSTATISTIK (107.258) WS 2010

Übung 10**18. Jänner 2012****Dutter**

37. Betrachten Sie wieder die Daten `werner_bcd`, bereinigen sie (2 Ausreißer und fehlende Werte) und betrachten die einfache, lineare Abhängigkeit von Cholesterin von den anderen Variablen.

Welche Parameter des linearen Modells erscheinen wesentlich (Variable ID ausgeschlossen)?

Verwenden Sie Mallows- C_p um ein „ideales“ Modell zu finden. Illustrieren Sie dies entsprechend grafisch mit dem C_p -Plot.

38. Testen Sie formal (F-Test) das oben gefundene „ideale“ Modell gegen jenes mit allen sinnvollen 8 unabhängigen Variablen. Bitte um textliche Schlussfolgerung.

39. Verwenden Sie wie im Beispiel 35. die Daten des Vienna City Marathons im Jahr 2011. Betrachten Sie die einfache, lineare (?) Abhängigkeit der Endzeit von der Zwischenzeit und zeichnen wieder die Werte und das geschätzte Modell.

Schätzen Sie nun ein einfaches lineares Modell, aber in einer „robusten Weise“, d.h. besondere Werte sollten auch speziell behandelt werden oder jene, die nicht zur Masse der Beobachtungen gehören, erhalten geringere Gewichte.

Hinweis: Verwenden Sie statt `lm` die Funktion `ltsReg` aus dem \mathbb{R} -Paket `robustbase`.

Für Fortgeschrittene: Man überlege sich, wie man hier „robuste Konfidenz- und Toleranzbereiche“ angeben könnte und wie man diese interpretieren könnte.

40. Bereinigen Sie die Daten des Vienna City Marathons wie im Beispiel 36., indem Sie die Nullen aus den Zwischenzeiten entfernen und weiters alle offensichtlich unsinnigen Werte wie solche, wo z.B. das Verhältnis von Endzeit zu Zwischenzeit kleiner als 1.8 ist. Führen Sie sonst die gleichen Rechnungen wie vorhin durch und diskutieren Sie die Ergebnisse.

Für Fortgeschrittene: Überlegung wie oben.

Bitte, stellen Sie die Ausarbeitung in Form eines pdf-Files (nicht mehr als 3 Seiten) mit den Resultaten (Outputs plus textliche Kommentare) und Kurz-Listing des Programmcodes (Funktion) in die TUWEL-Seite

<https://tuwel.tuwien.ac.at/course/view.php?id=2604>

bis incl. 23. Jänner 2012.

Bitte den Namen des pdf-File folgendermaßen:

name_exer_10.pdf

wobei 'name' für den Familiennamen steht. Bitte im pdf-File Name, Datum und Seitennumerierung nicht vergessen!

Empfehlenswert ist es, den Bericht mit „Sweave“ zu erstellen. (Vorlage siehe <http://www.statistik.tuwien.ac.at/public/dutt/vorles/>.)