

**STATISTIK 2 (107.325) WS 2012**  
**COMPUTERSTATISTIK (107.258) WS 2012**

**Übung 10**

**17. Dezember 2012**  
**Dutter**

37. Betrachten Sie wieder die Daten `werner_bcd`, bereinigen sie (2 Ausreißer und fehlende Werte) und schätzen ein Modell, in dem Cholesterin von allen anderen Variablen (außer Patientenummer) linear abhängt.
- Welche geschätzte Parameter sind von null signifikant verschieden (siehe summary-Tabelle)?
  - Bestimmen Sie aus den Einträgen in der 'summary-Tabelle' approximativ Konfidenzintervalle für die Parameter.
  - Wie können Sie aus den Konfidenzintervallen schnell die Signifikanz erkennen?
38. Prüfen Sie (grafisch) bei den obigen beiden Modellen die Verteilung der Residuen. Ist die Annahme der Normalverteilung gerechtfertigt?  
 Gibt es beim Residuen/ $\hat{y}$ -Plot Auffälligkeiten? Welche Aussagen über das Modell könnte man mit diesem Plot machen?
39. Betrachten Sie wieder wie oben die gereinigten Daten `werner_bcd` 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.
40. Testen Sie formal (F-Test) das oben gefundene „ideale“ Modell gegen jenes mit allen sinnvollen 8 unabhängigen Variablen. Bitte um textliche Schlussfolgerung.

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 Programmkodes (Funktion) in die TUWEL-Seite

<https://tuwel.tuwien.ac.at/course/view.php?idnumber=107258-2012W>  
 bis zum 4. Jänner 2013, 23:45 Uhr.

Bitte den Namen des pdf-File folgendermaßen:

**name\_exer\_10.pdf**

wobei 'name' für den Familiennamen steht.

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

**Kreuzen Sie außerdem bitte im Internet** jede Nummer des Übungsbeispiels an, das Sie dann in der Übungsstunde an der Tafel (mit Beamer-Unterstützung) vorrechnen wollen und können. Der Termin ist üblicherweise 3 Stunden vor der Übung, d.h. Mo., 11:00. Siehe <http://www.statistik.tuwien.ac.at/cgi-bin/uebbsp/bspstart.cgi> .