

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

**Übung 9**

**11. Jänner 2012**  
**Dutter**

33. Betrachten Sie wieder die Daten `werner_bcd`, bereinigen sie (2 Ausreißer und fehlende Werte) und betrachten die einfache, lineare Abhängigkeit von Cholesterin vom Alter (ohne Klasseneinteilung).

Sind die Parameter der linearen Abhängigkeit (der Modellgeraden) signifikant?

Um wieviel ändert sich statistisch (d.h. mit Angabe der ungefähren Genauigkeit von +/- entsprechend einem ungefähren Konfidenzintervall) der Cholesterin-Wert pro Jahr?

Zeichnen Sie die Werte, die geschätzte Gerade, den Konfidenzbereich für die Gerade und den Toleranzbereich für weitere (zukünftige, unabhängige) Beobachtungen (in Form von Hyperbeln).

34. Verwenden Sie die Patientendaten "Werner Blood Chemistry Data". Betrachten Sie den „Body Mass Index“, d.h.

$$bmi = (Gew. \text{ in } kg) / (Größe \text{ in } m)^2$$

(In unseren Daten könnten Sie die 'pounds' ungefähr mit 1/2 und die 'inches' mit 2.54/100 umrechnen, z.B. mit dem Befehl

```
bmi <- (werner_bcd[, "WEIGHT"]/2)/(werner_bcd[, "HEIGHT" ] * 2.54/100)^2
.)
```

Präsentieren Sie die Werte von `bmi` grafisch.

Hängt `CHOLSTRL` signifikant von `bmi` ab?

35. Verwenden Sie die Daten des Vienna City Marathons im Jahr 2011. Betrachten Sie die einfache, lineare (?) Abhängigkeit der Endzeit von der Zwischenzeit. Zeichnen Sie die Werte, die geschätzte Gerade, den Konfidenzbereich für die Gerade und den Toleranzbereich für weitere (zukünftige, unabhängige) Beobachtungen (in Form von Hyperbeln).

Konzentrieren Sie sich nun nur auf die ersten 100 Werte und zeichnen das Gleiche mit dem reduzierten Datensatz.

Es gibt noch immer einen Ausreisser (Zwischenzeit = 0). Entfernen Sie diesen und zeichnen das Resultat in die gleiche Grafik.

Die ersten etwa 26 Läufer verhalten sich etwas anders. Führen Sie Schätzungen auf der Basis dieser durch und zeichnen wieder das Resultat in die gleiche Grafik.

Bitte um Diskussion.

36. Bereinigen Sie die Daten des Vienna City Marathons, 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.

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?id=2604>

bis incl. 16. Jänner 2012.

Bitte den Namen des pdf-File folgendermaßen:

**name\_exer\_9.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/>.)