

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

Übung 9

10. Dezember 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? Ist diese in der Praxis häufig gehörte Frage sinnvoll gestellt? Oder haben Sie einen Verbesserungsvorschlag?

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
.)
```

Hängt `CHOLSTRL` signifikant von `bmi` ab?

Betrachten Sie als Modell Cholesterin linear auch abhängig von allen anderen Variablen. Welche Variablen sind in erster Linie von Einfluss und welche nicht?

35. Verwenden Sie die Daten des Vienna City Marathons im Jahr 2012. 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.

Die ersten etwa 21 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 Programmcodes (Funktion) in die TUWEL-Seite

<https://tuwel.tuwien.ac.at/course/view.php?idnumber=107258-2012W>
bis zum 14. Dezember 2012, 23:45 Uhr.

Bitte den Namen des pdf-File folgendermaßen:

name_exer_9.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> .