

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

Übung 7

26. November 2012
Dutter

25. Betrachten Sie wieder den Datensatz „werner.bcd“ und entfernen Sie die beiden Ausreisser von Cholesterin. Untersuchen Sie die Abhängigkeit der Variablen Cholesterin von den Klassen von `bmi`, wie es im Übungsblatt 5 beschrieben ist (Wiederholung: Die geordneten `bmi`-Werte wurden in etwa 4 gleich grosse Teile geteilt). Konstruieren Sie nochmals genau wie in der Vorlesung Abschnitt „Lineare Modelle“ besprochen, die Versuchsplanmatrix \mathbf{X} und betrachten dann das lineare Modell

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

Man versuche den unbekannt Parametervektor $\boldsymbol{\beta}$ mit der kleinsten Quadrate-Methode zu schätzen (konkrete Rechnung bitte, nicht mit höherer Funktion `lm` oder Ähnlichem)¹.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$$

26. Setzen Sie Beispiel 25 fort, indem Sie die Varianz-Kovarianz durch

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}$$

mit

$$s^2 = \widehat{\sigma^2} = \frac{1}{n-r} \|\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2$$

berechnen.

- (a) Diskutieren Sie die Werte von $\hat{\boldsymbol{\beta}}$ und deren Signifikanzen (mit Daumenregel: geschätzter Wert ± 2 *Standardfehler).
- (b) Was sind dabei eigentlich die 0-Hypothesen?
27. Führen Sie die gleiche Rechnung wie im Beispiel 25 durch, mit der Versuchsplanmatrix erzeugt durch die `R`-Funktion `model.matrix`.
28. Berechnen Sie auch die Varianz-Kovarianz und diskutieren die beiden Punkte aus Beispiel 26 auf der Basis der anderen Versuchsplanmatrix aus Beispiel 27.

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 30. November 2012, 23:45 Uhr.

Bitte den Namen des pdf-File folgendermaßen:

name_exer_7.pdf

wobei 'name' für den Familiennamen steht.

¹Hinweis: `R`-Funktion `t` für 'transponiert', `R`-Operator `%*%` für 'Matrizenmultiplikation', `R`-Funktion `solve` für 'invertieren'

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> .