

STATISTIK 2 (107.325) WS 2009
COMPUTERSTATISTIK (107.258) WS 2009

Übung 9

11. Jänner 2010
Dutter

33. Betrachten Sie die cissik-Daten (File ‘cissik_split.txt’ auf <http://www.statistik.tuwien.ac.at/public/dutt/vorles/>).
- Kurzbeschreibung: Experiment bezüglich Produktion von gasförmigem Stickstoff im menschlichen Körper. Die gemessenen Werte hängen vermutlich von der Art der Diät und weiteren Faktoren (Zeitpunkt (m/a) ...) ab. Gibt es durchschnittliche Unterschiede bezüglich den Faktoren ‘Personen’, ‘Zeitpunkt’ (morgens/abends) bzw. bei der ‘Art der Diät’? Gibt es Wechselwirkungen?
34. Wie 33, aber ‘Person’ wird als Faktor mit zufälligen Effekten interpretiert.
- Hinweis: Man könnte einfach mit festen Effekten arbeiten und dann händisch laut Skriptum die Teststatistiken berechnen. In der Funktion ‘aov’ kann allerdings in diesem Fall ‘Person’ auch als Fehlerterm interpretiert und in die Modellformel eingegeben werden (+ `Error(Person)`).¹
35. Verwenden Sie die Daten des Vienna City Marathons im Jahr 2009. 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).²³

¹Fortgeschrittene können auch die Funktion `lmer` aus dem Paket `lme4` versuchen. (Vorsicht: Verwendung von S4-Klassen.)

²Ein Konfidenzintervall mit der Konfidenzzahl α für den Mittelwert $\mu_{y,x}$ an der Stelle x erhält man mit der Formel

$$\hat{y}_x - t_{n-2;1-\frac{\alpha}{2}} S \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)s_x^2}} < \mu_{y,x} < \hat{y}_x + t_{n-2;1-\frac{\alpha}{2}} S \sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)s_x^2}}.$$

Ein Toleranzintervall ist folgendermaßen:

$$\hat{y}_x - t_{n-2;1-\frac{\alpha}{2}} S \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)s_x^2}} < y < \hat{y}_x + t_{n-2;1-\frac{\alpha}{2}} S \sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)s_x^2}}.$$

³: siehe `predict`.

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.

Bitte senden Sie die Ausarbeitung in Form eines pdf-Files (nicht mehr als 3 Seiten) mit den Resultaten (Outputs plus textliche Kommentare) und Listing des Programmkodes (Funktion) an

R.Dutter@tuwien.ac.at

bis zum Samstag-Abend, 16 Jänner 2010.

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/> .)