

Statistik II

Übungsblatt 3

Abgabe: Do. 13.11.2003, 10.00 Uhr, Briefkasten: Statistik II.

Bei jeder Aufgabe können maximal 5 Punkte erreicht werden.

1. "SMSA"-Datensatz:

Logarithmieren Sie die Stickoxidbelastung und ersetzen Sie dadurch in Ihrem Lieblingsregressionsmodell aus Aufgabe 3 von Blatt 2 die unlogarithmierte Stickoxidbelastung. Isolieren Sie die Beobachtung von York, indem Sie einen Selektor kreieren und in das Regressionsmodell einbeziehen. Untersuchen Sie für dieses Modell die Residuen, die Hebelwirkungen und die Cook-Distanzen. Verbessert sich das Modell durch den Selektor?

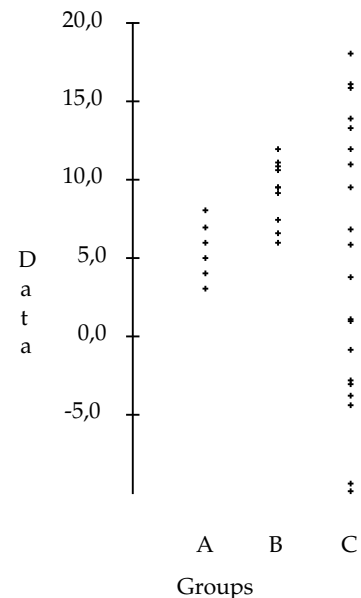
2. Verständnisfragen

(i) Bei linearen Modellen wird vielfach standardmäßig angenommen, daß die Beobachtungsfehler ϵ_i normalverteilt sind. Die Signifikanz eines Effekts wird dann üblicherweise mit einem t -Test für die Hypothesen, H_0 : Parameterschätzungen sind gleich 0, untersucht. Warum ist der t -Test hier gerechtfertigt?

(ii) Was ist der Unterschied zwischen dem Vergleich zweier Mittelwerte im Rahmen einer Varianzanalyse und dem direkten Vergleich von zwei Gruppen mittels t -Test?

(iii) Stellen Sie sich vor, Sie haben eine Varianzanalyse durchgeführt, bei der der Gesamt-F-Test signifikant ist. Was schlagen Sie vor, um herauszufinden, woran genau das liegt.

(iv) Betrachten Sie die nebenstehende Graphik. Der Unterschied zwischen den Gruppen A und B allein ist signifikant, sobald man jedoch Gruppe C hinzunimmt, ist der Unterschied zwischen A und B nicht mehr signifikant. Erläutern Sie die Gründe für diesen Effekt. Was schlagen Sie zur Lösung vor?



3. "Barley"-Datensatz:

Analysieren Sie den Barley-Datensatz mit Hilfe der Varianzanalyse. Ziehen Sie dabei alle Variablen in Betracht und möglicherweise auch Interaktionen zwischen den Variablen.

(a) Welches Modell halten Sie für "das Beste"? Wie haben Sie dieses Modell gefunden und auf welchen Überlegungen basiert Ihre Einschätzung, daß gerade dieses Modell "besser" als alle anderen ist?

(b) Geben Sie eine möglichst allgemeinverständliche Interpretation Ihres Modells.

(c) Welche Schlüsse lassen sich aus diesem Modell ziehen, welche nicht? Was würden Sie einem Landwirt raten, der auch Gerste anbauen will?

4. "Höglwald"-Datensatz Um die Wirkung von Beregnung und Kalkung im Wald studieren zu können, wurden im Forst "Höglwald" (Forstamt Aichach) Freilandexperimente gemacht. Erfasst wurden die

folgenden Einflußgrößen:

Berechnung (Kontrolle, zusätzliche saure Berechnung, zusätzliche normale Berechnung)

Kalkung (Kontrolle (ohne), zusätzliche einmalige Kalkung des Bodens (mit))

Baumnähe (Bodenprobe wurde in Baumnähe oder zwischen den Bäumen genommen).

Die dem Boden entnommenen Proben wurden dabei u. a. auf ihren pH-Wert hin untersucht (kleine pH-Werte = sauer, große Werte = alkalisch). (Dieser Datensatz ist Pruscha, H., (1989) *Angewandte Methoden der Mathematischen Statistik*, Teubner, Stuttgart, entnommen.)

- (a) Analysieren Sie den Datensatz mit Hilfe der mehrfaktoriellen Varianzanalyse, jedoch vorerst ohne Interaktionen. Interpretieren Sie Ihr Ergebnis. Machen die Koeffizientenschätzer "Sinn"? Welche Gruppen lassen sich deutlich voneinander unterscheiden, welche nicht? Können Sie die verschiedenen Berechnungsgruppen eindeutig identifizieren?
- (b) Pruscha empfiehlt eine Wurzeltransformation der pH-Werte, um die Forderung nach Varianzgleichheit in den Stichproben zu erfüllen. Nehmen Sie dazu Stellung und begründen Sie Ihre Ansichten.
- (c) Analysieren Sie den Datensatz **mit** Interaktionen. Welches Modell halten Sie diesmal für "das Beste"? Begründen Sie Ihre Auswahl und interpretieren Sie wieder Ihr Modell.

5. "Barley"-Datensatz:

Analysieren Sie den Barley-Datensatz mit Hilfe einer Varianzanalyse. Betrachten Sie dabei das Modell für Gerstenernte mit den Faktoren Ort und Jahr sowie deren Wechselwirkung

- (a) Erstellen Sie geeignete Graphiken, um den Einfluss der beiden Faktoren sowie der Interaktion darzustellen. Was fällt Ihnen bei dem Interaktionplot auf?
- (b) Berechnen Sie mittels Post-Hoc Tests nach Scheffé und Tukey, welche Orte und ob sich die Jahre signifikant zu einem Niveau von 5% unterscheiden. Gibt es unterschiedliche Ergebnisse für die beiden Verfahren?
- (c) Interpretieren Sie und vergleichen sie die Ergebnisse aus (a) mit denen aus (b).