Serie 8

1. Vier Medikamente zur Erhöhung des Blutdruckes werden miteinander verglichen. Die Medikamente M1 und M2 sind bekannte Standards, die Medikamente M3 und M4 sind neu. Die Versuchsdaten in der Datei `...../med-blut.dat` können mit `read.table` (**ohne** `header=T`) gelesen werden.

a) Zeichnen Sie die Box-Plots für jedes Medikament mit `plot(..)`.

b) Führen Sie die Einweg-Varianzanalyse durch und prüfen Sie, ob Unterschiede zwischen den Medikamenten vorliegen.

c) Wenden Sie die Methode der multiplen Vergleiche an, um eventuell vorhandene Unterschiede zwischen je zwei Medikamenten zu finden (Bonferroni Methode).

Tip: Sie können mit zwei `for`-Schlaufen arbeiten

```
imedi _ as.integer(Medi)
for (i in 1:3) {
 for (j in (i+1):4) {
    print(wilcox.test(Wirkung[imedi==i], Wirkung[imedi==j]))
    print(t.test(.....))
}}
```

oder Gruppen explizit auswählen, z.B. `Wirkung[Medi=="M1"]`.

d) Prüfen Sie, ob ein Unterschied zwischen (M1 + M2) und (M3 + M4) vorliegt (2-Gruppen Problem !)

e) Prüfen Sie den Unterschied zwischen den vier Medikamenten auch nichtparametrisch (ohne Annahme der Normalverteilung) mit Hilfe des Kruskal-Wallis-Tests.
S-plus: `kruskal.test`(*y*, *grp.fact*).

2. Der Datensatz `d.particles` enthält Messungen des Partikelgehalts der Luft für 4 Zeitpunkte und 5 Orte, siehe `doc(d.particles)`. Wir wollen prüfen, ob die Orte sich unterscheiden. Das ist ein Problem der zweidimensionalen ("2-Weg") Varianzanalyse.

a) Überlegen Sie sich graphische Darstellungen (→ Bilder; Funktionen `interaction.plot(....)`, `plot.design` sind nützlich). Sehen Sie sich auch an, was
`> plot( conc ~ t + s , data= d.particles)` produziert.

b) Führen sie die Varianzanalyse durch (`aov`(...)) und beantworten Sie die Hauptfrage. Stimmen die Modell-Voraussetzungen (Residuen !)?

c) Nebenübung (*): Schreiben Sie die Daten als Tabelle (Matrix), in der die Konzentrationen stehen: In den Zeilen die Konzentrationen für gleiche Zeitpunkte, in den Kolonnen für gleiche Orte. (Funktion `matrix` und evt. `dimnames`).

3. In der Datei `....datasets/ausbeute-chem.dat` (kein , `header=T` nötig) befinden sich die Ergebnisse eines 2^5-Versuches, d.h. 5 Faktoren mit je zwei Niveaus (ohne Replikate). In diesem Beispiel der chemischen Prozesssteuerung sind die Faktoren tatsächlich *quantitative* Variable, welche bei der Versuchsplanung auf nur zwei verschiedene Stufen, `L` („Low") und `H` („High"), festgelegt wurden:

Zufuhr	:	Zufuhr (Liter/Min)
Kat	:	Katalysator (in %)
Geschw	:	Rührgeschwindigkeit (Umdrehungen pro Min.)
Temp	:	Temperatur (^oC)
Konz	:	Konzentration (in %)
Ausbeute:		(Menge), → „$$“. Soll maximal sein.

a) Sehen Sie die Rohdaten (nur 32 Beob.) an (“`d.ausb`”, wenn der `data.frame` so heisst). Mit `plot.design(`*d.ausb*`)` bekommen Sie einen ersten Eindruck über die Effekte der 5 Faktoren.
Führen Sie die fünffach („5-weg“) VA durch, einmal nur Haupteffekte, dann auch mit allen 2-fach WW. In S-plus kann das letztere mittels `aov(Ausbeute ~ . ∧2, data= d.ausb)` erreicht werden. [Immer mit `summary(`*aov–Resultat*`)` anschauen!]

b) Untersuchen Sie die Residuen. Was muss man bei den Residuen beachten, wenn nur Haupteffekte im Modell sind?

c) Passen Sie jetzt das „endgültige“ Modell an, indem Sie nur statistisch gesicherte Haupteffekte und zweifach WW verwenden. Was ändert sich gegenüber a) ?

4. * *freiwillig — Beispiel aus Vorlesung —*
Auf der Datei `...../weizen-3w.dat` sind die Resultate eines Weizenbau-Versuches gespeichert: 6 verschiedene Weizensorten wurden je an 5 Orten während dreier Jahre angepflanzt (und wir hoffen, randomisiert). Es interessiert uns die Weizenqualität, die hier als `E.Grad` := Erweichungsgrad von Teig aus Weizenmehl (in Konsistenzeinheiten zwischen 0 und 20) gegeben ist. Für den gemessenen Wert `E.Grad` nehmen wir das folgende Modell an:

$$E.Grad_{ijk} = \mu + Sorte_i + Ort_j + Jahr_k + Fehler_{ijk}$$

Die Wechselwirkung (=: WW) zwischen je zwei der drei Faktoren wäre ebenfalls bestimmbar; wir nehmen sie vorläufig als unwichtig an. Der Fehlerterm enthält deshalb drei Zweifach-WW und die Dreifach-WW `Sorte*Ort*Jahr`.

a) Mit `plot.design(`*d.weizen....*`)` bekommen Sie einen ersten Eindruck über die Effekte der 3 Faktoren.
Führen Sie die Dreifach-VA (mit nur Haupteffekten) (“`Y ~ .`”) durch.

b) Untersuchen Sie die Residuen auf Normalität und konstante Varianz (gegen „Fitted“ *und* die einzelnen Faktoren).

c) Falls ein auffälliges Residuum vorkommt, ersetzen Sie den kritischen Wert durch einen “unauffälligen” Wert an dieser Stelle und führen Sie die VA nochmals durch.
Oder „ehrlicher“, ersetzen Sie den Ausreisser durch `NA`, und rufen Sie
`> .... <- aov( ......, na.action = na.omit)` auf.

d) Gehen Sie von c) aus und erweitern Sie die VA durch die WW zwischen je zwei Faktoren (“`Y ~ .^2`”) . Lohnt sich dies? Welches scheint das beste endgültige Modell?

Abgabe: Mo, 13. Jan. 97.
Präsenz, Hilfe: Im SOL F5 (Sonneggstr.33), gemäss Vereinbarung: Tel. (63) 2-3408 oder via e-mail To: `maechler@stat.math.ethz.ch`.

Übung via WWW in `ftp://ftp.stat.math.ethz.ch/Teaching/maechler/Stat-Meth/96-8.ps`