

## Übung 3

**Ort und Zeit:** HG E26.3, Montag, 6. Februar, 13.15-15.00 Uhr.  
Uhr.

**Abgabe:** freiwillig nach der Übungsstunde.

**Musterlösung:** wird nach der Übungsstunde verteilt

Bei jeder Analyse werden auf Wunsch zusätzlich verschiedene Grafiken automatisch erzeugt. Für die Varianzanalyse brauchen wir diese im Moment nicht. Deshalb schalten wir sie aus via **Edit / Options** und dort bei **Output, Display Statistical Quickgraphs** inaktivieren (oder den Quickbutton **Statistical Quickgraphs** deaktivieren).

1. Mit einem Experiment sollte der Einfluss der Marinierung (3 Stufen) und der Bratdauer (4 Stufen) auf die Zartheit von Rindsbraten analysiert werden. Die rechten Rückenmuskeln von 6 zufällig ausgewählten Schlachtkörpern wurden in je 3 Bratenstücke zerlegt. Die 3 Bratenstücke desselben Tieres wurden unterschiedlich mariniert, wobei die Zuteilung nach einem Zufallsprinzip erfolgte. Während dem Braten wurden zu 4 vorgegebenen Zeiten mit einer Sonde kleine Teilstücke aus einer zufällig bestimmten Region entnommen und auf Zartheit analysiert.

Der Datensatz `Y:\users\s\sfs\datasets\Systat\braten.syd` enthält die Variablen **KOERPER** (Stufen 1 bis 6), **MARIN** (Stufen 1,2,3), **ZEIT** (Stufen 30, 36, 42 oder 48 Minuten) und **ZARTHEIT** (kleinere Zahlen bedeuten zartere Stücke).

- a) Welche Faktoren sind gekreuzt, welche geschachtelt?
- b) Weshalb liegt hier eine Spaltanlage vor? Welcher der drei Faktoren ist der Blockeffekt, welcher der "Main unit"-Faktor und welcher der "Split unit"-Faktor? Wie lautet das Modell?
- c) Stelle die **ZARTHEIT** gegen die **ZEIT** grafisch dar, mit unterschiedlichen Symbolen für jede Marinade. Was fällt auf?

Benutze **Quickbutton Scatterplot** und für das Plotsymbol **Appearance / Symbol and Label... / Symbols — Select Variable**.

- d) Berechne nun mit Systat die entsprechende Varianzanalyse-Tabelle. Das Vorgehen ist analog zum Beispiel auf Seite 24 im Skript Varianzanalyse II. Erstelle zuerst auf dem Papier die richtige ANOVA-Tabelle mit der Angabe der Faktoren, der Fehlerterme und der Freiheitsgrade.

Ergänze die Tabelle mit Hilfe des Systat-Outputs. Interpretiere die Resultate.

Zuerst müssen wir Systat mitteilen, dass es sich bei allen drei erklärenden Variablen um kategoriale Grössen handelt, obwohl sie mit Zahlen codiert sind: **Data / Category**.

Jetzt gilt es, das Modell mit allen benötigten Faktoren und Wechselwirkungen anzugeben: **Statistics / General Linear Model / Estimate Model**. Interaktionen müssen explizit angegeben werden, mit Hilfe von **Cross**. Beachte: Die Main Unit Abweichung entspricht der Interaktion von Block und Main Unit Faktor. (*Save file*: anklicken und am besten *Residuals/Data* auswählen, für Residuenanalyse später.)

Da Systat nichts von einem Split Plot weiss, wird getestet, als ob es sich um gewöhnliche gekreuzte Faktoren handelt. Benutze die Systat-Resultate, um die ANOVA-Tabelle korrekt auszufüllen! (Statt von Hand kann man den Main Unit Effekt auch mit Systat testen: **Statistics / General Linear Model / Hypothesis Test** und dort bei **Effects** den zu testenden Effekt und bei **Between Subject(s) Effect(s)** den Fehlerterm eintragen.)

- e) Überprüfe die Modellvoraussetzungen mit einer Residuenanalyse.

Lade den vorher gespeicherten Datensatz.

**Tukey-Anscombe-Plot:** **Quickbutton Scatterplot**, und die entsprechenden Variablen gegeneinander auftragen.

**Normal-Plot:** **Graph / Plots / Probability Plot ... RESIDUAL X→** und unter **Axes / All Axes** das Feld **Transpose XY-axes** anwählen.

- f) Betrachte auch den Interaktionsplot zwischen den Faktoren ZEIT und MARIN, um die signifikante Wechselwirkung ZEIT\*MARIN zu untersuchen.

Zuerst muss der Originaldatensatz wieder geladen werden. Dann **Quickbutton Line** und ZEIT als X-, ZARTHEIT als Y- und MARIN als Grouping Variable angeben. **Overlay Multiple Graphs** anklicken.

**Quelle:** S. Dowdy and S. Wearden, Statistics for Research (1991), John Wiley & Sons, New York.

2. In einem Ausgangsmaterial für die Farbenherstellung wurde der Wassergehalt mit folgender Erhebung analysiert: 15 Chargen wurden zufällig entnommen. Pro Charge sind anschließend je 2 zufällige Proben untersucht worden. Zu jeder Probe liegt eine Doppelbestimmung (2 Werte) vor.

Die SYSTAT-Datei **Y:\users\s\sfs\datasets\Systat\paint.syd** enthält die Zielgrösse **MOISTURE** sowie die Werte zu den Gruppierungsvariablen **CHARGE** (1 bis 15), **PROBE** (1 bis 2) und **REP** (1 bis 2). Mit **SAMPLE** (1 bis 30) sind alle Proben fortlaufend durchnummeriert.

- a) Stelle die Daten in einem Streudiagramm **MOISTURE** gegen **CHARGE** dar, mit unterschiedlichen Symbolen für die beiden Proben. Was fällt im Bezug auf die Proben auf? Gibt es ein Problem mit dieser Darstellung?

**Quick-Button: Scatterplot.** **MOISTURE** als **Y-Variable(s)** und **CHARGE** als **X-Variable(s)** angeben. Bei der Option **Appearances/Symbol and Label...** wählt man bei **Symbols** — **Select variable PROBE** aus.

- b) Formuliere das entsprechende lineare Modell. Weshalb sind alle Faktoren zufällig?  
c) Skizziere die Struktur der entsprechenden Varianzanalyse bezüglich Faktoren, Freiheitsgraden und Erwartungswert der einzelnen Durchschnittsquadrate.

**Hinweis:** Im Skript gibt es ein entsprechendes Beispiel.

- d) Analysiere die Daten mit der SYSTAT-Prozedur **GLM**. Vorsicht: **GLM** kennt keine Faktoren mit zufälligen Effekten; der angegebene Signifikanztest für den Faktor **CHARGE** ist deshalb unpassend!

Systat-Anleitung: ähnlich wie Aufgabe 1d):

Zuerst Systat mitteilen, dass die Erklärenden Faktoren sind. Dann den Befehl **GLM** verwenden. Den geschachtelten Faktor kann man mit **Nest** erzeugen: **PROBE(CHARGE)** bedeutet **PROBE** in **CHARGE** geschachtelt. Wiederum die Residuen für später abspeichern.

- e) Berechne von Hand die Varianzkomponenten. Wie sind diese zu interpretieren?  
f) Überprüfe die Modellannahmen mit Hilfe einer Residuenanalyse. Zeichne dazu einen Tukey-Anscombe-Plot sowie einen Normal-Plot der Residuen.

Öffne den in d) gespeicherten Datensatz.

**Tukey-Anscombe-Plot:** **Quickbutton Scatterplot**, und trage die entsprechenden Variablen gegeneinander auf. **Normal-Plot:** **Graph / Plots / Probability Plot**, dann **RESIDUAL** als X-Variable angeben und unter **Axes / All Axes** **Transpose X-Y Axes** markieren.

**Präsenz in den Ferien:** Bei Fragen zur Vorlesung oder den Übungen geben der Dozent Dr. H.-R. Roth (Tel. 632 32 45, roth@stat.math.ethz.ch) sowie der Assistent Nicolai Meinshausen (Tel. 632 53 19, nicolai@stat.math.ethz.ch) gerne Auskunft.