

## Übung 1

**Ort und Zeit:** HG E23, Dienstag, 18. November, 10.15 bis 11.55 Uhr.

**Abgabe:** freiwillig am Dienstag, 25. November in der Vorlesung.

**Musterlösung:** wird nach der Übungsstunde verteilt.

**Starten von SYSTAT:** Menü **Start\Programs\Statistics\SYSTAT 10\SYSTAT 10**

**Pfad für Datensätze:** Datensätze werden im Verzeichnis **Z:\users\s\sfs\datasets\Systat** zur Verfügung gestellt.

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. In einem Experiment sollte untersucht werden, ob die Zugabe eines bestimmten Enzyms die Fleischqualität derart beeinflusst, dass das Fleisch zarter wird. Im Rahmen des durchgeführten Experimentes wurden vier verschiedene Enzymbehandlungen (TREATM) verglichen. Jede Behandlung wurde an fünf Fleischstücken durchgeführt. Mit Hilfe einer sogenannten Warner-Bratzler Scherapparatur konnte schliesslich für jedes Fleischstück die „Zartheit“ (TENDER) bestimmt werden.

Die Daten stehen im Systat-File **meat.syd** zur Verfügung.

- a) Zunächst wollen wir die Daten graphisch analysieren. Zeichne dazu einen Dot Density Plot der Daten.

Menü **Graph / Density Displays / Dot Density** anklicken. TENDER als Y-Variable und TREATM als X-Variable angeben, **OK** klicken, dann auf Grafik doppelklicken, damit sie im separaten Grafikkfenster erscheint. Dort **Graph / Appearance / Symbol and Label** anklicken und *Symbol Size* auf 0.9 setzen.

Kommentiere den Plot. Gibt es extreme Beobachtungen? Ist zu erwarten, dass signifikante Unterschiede zwischen den Gruppen bestehen? Gibt es Hinweise, dass Voraussetzungen für die Durchführung der Varianzanalyse verletzt sein könnten?

- b) Führe eine einfaktorische Varianzanalyse durch. Unterscheiden sich mindestens zwei Behandlungen signifikant?

**Statistics / Analysis of Variance (ANOVA) / Estimate model...**

dann entsprechend der Aufgabenstellung die eine Grösse als *Dependent* und die andere als *Factors* angeben.

Um die Residuen zu speichern (für nächste Teilaufgabe), muss man *Save file:* markieren und am besten *Residuals/Data* auswählen. Speichere die Residuen im Directory **T:\systat** (zuerst kreieren) ab.

- c) Überprüfe die Modellannahmen mit Hilfe einer Residuenanalyse. Zeichne dazu einen Normal-Plot der Residuen und einen Tukey-Anscombe-Plot.

Öffne den vorher gespeicherten Datensatz.

**Tukey-Anscombe-Plot:** **Graph / Plots / Scatterplot** oder einfach **Quickbutton Scatterplot** (6. Grafik-Icon), und trage die richtigen Variablen gegeneinander auf.

**Normal-Plot:** **Graph / Plots / Probability Plot** dann RESIDUAL als X-Variable angeben und die Achse transponieren (**Axes/All Axes, Transpose X-Y Axes**).

Gibt es aufgrund der beiden Plots Hinweise, dass bestimmte Modellvoraussetzungen nicht erfüllt sind? Falls ja, welche?

- d) Öffne wieder den Original-Datensatz (**File / Recent Data**). Berechne eine neue Variable  $\text{LOGTEND} = \log(\text{TENDER})$  und speichere den veränderten Datensatz unter dem Namen **meat2.syd** ins Directory **T:\systat**. Führe die Teilaufgaben **a)** und **b)** für die neue Zielgrösse LOGTEND durch.

**Data / Transform / Let ...** : und dort **Let LOGTEND = log(TENDER)** eingeben.

Gibt es Unterschiede zum Modell mit der Zielgrösse TENDER?

- e) Betrachte nun wiederum die Residuen. Zeichne die gleichen Plots wie in **c)** für die Zielgrösse LOGTEND. Sind die Modellvoraussetzungen nun eher erfüllt als vorher?

(Quelle: *Gacula, Maximo C., and Jagbir Singh (1984). Statistical Methods in Food and Consumer Research.*)

2. Die Datei **weizen.syd** enthält Messungen des Ertrags (Variable **GEWICHT**, in kg/ha) von Winterweizen. Die 3 Faktoren sind: 5 Sorten (**SORTE\$**), 2 Saatlängen (**SAAT**, 150 bzw. 400 Körner pro Quadratmeter) und 4 Felder (**FELD**). Zu jeder der  $5 * 2 * 4 = 40$  Kombinationen liegen Werte von 2 Parzellen vor.

- a) Führe eine multifaktorielle Varianzanalyse für den Ertrag durch. Welche Interaktionen sind auf dem 5%-Niveau signifikant?

**Statistics / Analysis of Variance (ANOVA) / Estimate Model...**

- b) Führe eine Varianzanalyse für den Ertrag durch, wobei nur die Haupt- und Wechselwirkungseffekte berücksichtigt werden, die sich in **a)** als signifikant erwiesen haben. Erzeuge auch einen Scatterplot der Residuen gegen die angepassten Werte (Tukey-Anscombe-Plot), sowie einen Normal-Plot der Residuen. Kommentiere die Resultate.

**Statistics / General Linear Model (GLM) / Estimate model...**

Eine Interaktion gibt man wie folgt ein: füge die erste Variable der Interaktion nochmals bei *Independent* an und wähle dann die zweite Variable in der Liste aus. Mit *Cross* wird die Interaktion nun bei *Independent* erzeugt.

**Wichtig:** Da SAAT und FELD numerisch kodiert sind, muss man diese bei **Categories...** als Gruppierungsvariablen angeben.

Für die Residuenanalyse muss man die Residuen explizit speichern (Option *Save file* markieren und am besten *Residuals/Data* auswählen, dann den gespeicherten Datensatz öffnen).

- c) Erzeuge eine Graphik der mittleren Erträge für SAAT= 150 und SAAT= 400 gegen FELD. Warum ist diese Graphik (Interaktionsplot) von Interesse?

**Graph / Summary Charts/ Line ...** oder einfach **Quickbutton Line** (2. Grafik-Icon), und GEWICHT als **Y-Variable(s)**, FELD als **X-Variable(s)** und SAAT als **Grouping-Variable(s)**. Dazu noch *Overlay multiple graphs into a single frame* anklicken.

- d) Führe pro Stufe des Faktors SAAT eine Varianzanalyse für den Ertrag mit den Faktoren FELD und SORTE durch. Vergleiche mit **c)** und kommentiere.

**Data / By Groups ...**

**Statistics / Analysis of Variance (ANOVA) / Estimate model....**