

Übung 2

Ort und Zeit: HG E26.3, Montag, 16. Januar, 13.15-15.00 Uhr.

Abgabe: freiwillig am Montag, 23. Januar in der Vorlesung

Musterlösung: wird am 23. Januar 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. Von 4 Hybrid-Mais-Sorten, welche eine erhöhte Resistenz gegenüber Pilzbrand aufweisen, wurde das Ertragspotential ermittelt. Das Experiment wurde als vollständiger Blockversuch mit dem Blockfaktor Ort (5 Stufen) durchgeführt. In der folgenden Tabelle ist der Körnerertrag zusammengestellt:

Hybrid-Sorte	Ort					Mittel
	NW	NE	C	SE	SW	
FR-11	62.3	64.0	64.3	65.0	66.4	64.40
BCM	63.3	62.7	66.2	66.8	64.5	64.70
DBC	60.8	64.3	65.2	62.2	65.1	63.52
RC-3	55.4	56.0	59.8	58.0	58.8	57.60
Mittel	60.45	61.75	63.88	63.00	63.70	62.56

Die SYSTAT-Datei **Y:\users\s\sfs\datasets\Systat\hybmais.syd** enthält die Variablen **HYBRID\$, LOCATION\$** und **YIELD**. Die Datei kann mit **File/Open/Data** geöffnet werden.

- a) Stelle die Daten zunächst graphisch dar.

Quickbutton Scatterplot. **YIELD** als **Y-Variable(s)** und **HYBRID\$** als **X-Variable(s)** angeben. Damit jeder Ort in einer anderen Farbe und mit einem anderen Symbol dargestellt wird, kann man noch **LOCATION\$** als **Grouping variable** angeben und **Overlay multiple graphs into a single frame** ankreuzen.

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) Formuliere ein zum durchgeführten Versuch passendes Modell der Varianzanalyse.
- c) Berechne die ANOVA-Tabelle mit SYSTAT und speichere die Residuen. Gib die Nullhypothesen der durchgeführten F-Tests an. Welche Nullhypothesen werden verworfen?

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

Für die Residuenanalyse muss die Option *Save file* angekreuzt werden und am besten *Residuals/Data* ausgewählt werden. Speichere die Residuen im Directory **T:\systat**.

- d) Führe paarweise Vergleiche nach Fisher, Bonferroni und Tukey zwischen den verschiedenen Sorten durch. Welche Sorten unterscheiden sich signifikant? Kommentiere und diskutiere Unterschiede zwischen den verschiedenen Testverfahren.

Statistics / General Linear Model (GLM) / Pairwise comparisons..., bei *Groups HYBRID* angeben und den gewünschten Test wählen.

- e) Fasse die Testergebnisse in einer Tabelle gemäss Skript, S. 28 zusammen. Gibt es in dieser Darstellung Unterschiede zwischen den verschiedenen Testverfahren?
- f) Überprüfe die Modellannahmen mit Hilfe einer Residuenanalyse. Zeichne dazu einen Tukey-Anscombe-Plot sowie einen Normal-Plot der Residuen. Kommentiere!

Öffne den in c) gespeicherten Datensatz.

Tukey-Anscombe-Plot: Quickbutton Scatterplot. Trage die Variable RESIDUAL gegen ESTIMATE auf.

Normal-Plot: Graph / Plots / Probability Plot; dann RESIDUAL als X-Variable angeben und unter **Axes / All Axes Transpose X-Y Axes** markieren.

- g) Welcher Punkt im Tukey-Anscombe Plot gehört zum Ertrag der Hybrid-Sorte BCM am Ort SE?

2. Ein Unternehmen beschliesst, die Maschinen zur Fertigung eines bestimmten Teils zu ersetzen. Es stehen drei verschiedene Maschinentypen zur Auswahl. Das Management der Firma ordnet an, die Produktivität der neuen Typen mit firmeneigenem Personal zu testen. Dazu wird folgendes Experiment durchgeführt: Fünf Angestellte der Firma werden zufällig ausgewählt. Jede dieser fünf Personen arbeitet dreimal während einer bestimmten Zeit an jedem Maschinentyp. Für jede so geleistete Arbeitseinheit wird die Produktivität bestimmt, indem jede Arbeitseinheit mit einem Index, der sowohl die Quantität als auch die Qualität der gefertigten Teile berücksichtigt, bewertet wird.

Der Datensatz **Y:\users\s\sfs\datasets\Systat\maschine.syd** enthält die Variablen PROD (Produktivität), TYPE (Maschinentyp), PERSON und REP (Replikatsnummer).

- a) Betrachte die Daten zunächst graphisch. Zeichne einen Boxplot für jeden Maschinentypen.

Quickbutton Boxplot (5. Grafik-Icon). PROD als **Y-Variable(s)** und TYPE als **X-Variable(s)** angeben.

Kommentiere den Plot.

- b) Formuliere ein passendes Modell. Welche Faktoren sind fix, welche zufällig? Gibt es hierarchische Faktoren? Gibt es gekreuzte Faktoren?
- c) Berechne die Varianzanalyse mit SYSTAT. Fasse die Resultate kurz zusammen. Sind die Tests korrekt (siehe auch nächste Teilaufgabe)?

Statistics / General Linear Model (GLM) / Estimate Model..., PROD als *Dependent*, TYPE, PERSON und TYPE*PERSON als *Independent* angeben. 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 TYPE und PERSON numerisch kodiert sind, muss man diese bei **Categories...** als Gruppierungsvariablen angeben.

Für die Residuenanalyse (Teilaufgabe f) muss die Option *Save file* markiert werden und am besten *Residuals/Data* ausgewählt werden. Speichere die Residuen im Directory **T:\systat**.

- d) Wie muss man den Effekt des Maschinentyps und den Effekt der Person testen? Gib die Nullhypothesen an, führe diese zwei Tests durch und kommentiere die Ergebnisse.

Korrekte Tests kann man in Systat via **Statistics / General linear model (GLM) / Hypothesis Test...** berechnen, indem man bei **Effects** den zu testenden Effekt und bei **Between Subject(s) Effect(s)** den Fehlerterm angibt.

- 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. Kommentiere die beiden Plots!

Öffne den in c) 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.

Quelle: G. A. Milliken and D. E. Johnson, Analysis of Messy Data. Volume I: Designed Experiments, Van Nostrand Reinhold, New York (1992)