Einführung in SYSTAT für Windows Version 10

Seminar für Statistik
ETH Zürich

Juni 2002

Inhaltsverzeichnis

1 Kurze Beschreibung von SYSTAT

1.1 Allgemeines

SYSTAT ist ein umfangreiches ‘general purpose’ Statistik-Paket, das einen grossen Teil der bekannten statistischen und graphischen Methoden abdeckt. Fast alle Möglichkeiten des Programms lassen sich durch Menüsteuerung ausschöpfen, so dass die Bedienung relativ komfortabel ist. Andererseits können aber auch alle Methoden über Befehle aufgerufen werden, die sich z.B. zur Automatisierung von Routineanalysen in Commandfiles zusammenfassen lassen. Systat bietet eine Reihe von Programmiermöglichkeiten (unter anderem auch eine Programmiersprache (genannt BASIC), mit der Daten eingelesen und behandelt werden können). Allerdings sind diese Möglichkeiten nicht mit denjenigen grösserer Produkte wie R/S-Plus oder SAS zu vergleichen.

Wir empfehlen SYSTAT wegen der Bedienungsfreundlichkeit und der Methodenvielfalt für Anwender, die ein relativ zügig erlernbares Programm suchen, ohne auf ein breitgefächertes Angebot statistischer Methoden verzichten zu müssen. Für Anwender, die eigene Funktionen schreiben möchten/müssen, oder kompliziertere Datenmanagementaufgaben lösen müssen, ist SYSTAT weniger geeignet.

Besonders erwähnenswert sind auch die ansprechenden und gut durchdachten graphischen Fähigkeiten von SYSTAT. Im ersten Kapitel des SYSTAT Handbuchs ’Graphics’ erfährt man mehr über die visuelle Informationsverarbeitung beim Menschen und über das Design informativer und seriöser Graphiken.

SYSTAT wurde 1995 von der Firma SPSS übernommen. Die Fachleute sind sich nicht ganz einig, ob sich diese Übernahme positiv oder negativ ausgewirkt hat. Unserer Meinung nach wurde SYSTAT durch die Firma SPSS deutlich verbessert und weiterentwickelt. Im April 2002 wurde SYSTAT an eine indische Firma namens “Systat Software Inc.” weiterverkauft. Für diese Firma soll SYSTAT eines der Hauptprodukte sein, so dass man eine vernünftige Weiterentwicklung erwarten kann.

1.2 Verfügbarkeit an der ETH

Systat V10 (Windows) ist via IDES kostenfrei verfügbar. Diese Lizenz gilt auch für Forschungsanstalten. Alle Hinweise sind unter

`https://ides.ethz.ch/swdb/produkt.phtml?nr=39`

und die Manuals als pdf-Files unter

`ftp://ides.ethz.ch/systat/win/v10.0/manuals`

zu finden.

2 Aufbau der Oberfläche

2.1 Das Hauptfenster

Beim Aufstarten von SYSTAT erscheint das Hauptfenster, das aus drei Teilfenstern, einem Menübalken und einem Quickbutton-Balken besteht. Die Menüs und Quickbuttons erlauben, die wichtigsten Operationen einfach auszuführen.

Im Output Teilfenster (Output pane) rechts erscheinen die Resultate der ausgeführten Operationen. Durch Scrollen im Output Teilfenster kann man ältere Resultate wieder hervorholen. Text (zB. Titel) kann beliebig hinzugefügt oder gelöscht werden.

Im Output Organizer (linkes Teilfenster) erscheint ein Inhaltsverzeichnis des Output Teilfensters. Mit dem Output Organizer kann man sich einerseits bequem im Output Teilfenster bewegen (einfach ensprechenden Eintrag im Inhaltsverzeichnis anklicken), und andererseits die Resultate reorganisieren (`Cut`, `Copy`, `Paste` im Output Organizer bewirken parallele Resultate im Output Teilfenster).

Das untere Teilfenster (Command Teilfenster) besitzt drei Blätter (Tabs), `Interactive`, `Batch(Untitled)` und `Log`. Im `Interactive` Blatt können Befehle via Tastatur eingegeben werden, sie werden nach drücken der `<Enter>`-Taste ausgeführt. Im `Batch` Blatt können ganze Programmdateien eingegeben oder geladen und dann via `File/Submit Window` ausgeführt werden. Der Name der Programmdatei ersetzt die Bezeichnung `Untitled`. Schliesslich wird im `Log` Blatt ein "Logbuch" der bisher ausgeführten Operationen angezeigt.

2.2 Das Datenfenster

Das Datenfenster erscheint beim Öffnen einer Datei oder durch `View/Data`. In diesem Fenster werden die Daten der Datei in Tabellenform dargestellt.

2.3 Das Graphikfenster

Die Resultate einer Graphik-Operation (z.B. Histogramme) erscheinen im Hauptfenster. Mit einem Doppelklick auf solche Graphiken wird das Graphikfenster geöffnet, d.h. es entsteht ein neues Fenster, das nur die Graphik enthält. Jetzt kann man sie genauer betrachten (Ausreissersuche), deren Plot-Parameter verändern ('Dynamic Explorer') oder separat abspeichern.

2.4 Menübalken

Jedes Fenster hat einen spezifischen Menübalken, wobei diejenigen des Haupt- und des Datenfensters fast identisch sind. Es können somit die Graphik- und Statistikbefehle von beiden Fenstern aufgerufen werden. Wenn Sie einen Befehl eines Menüs mit der Endung **...** anklicken, erscheint ein Dialogfenster, worin Variablen und Optionen ausgewählt werden können. Gewisse Menüs (gekennzeichnet mit dem Symbol ►) enthalten auch Submenüs.

2.5 Toolbars (Quick Buttons)

Unter dem Menübalken im Hauptfenster befinden sich verschiedene Symboltasten (Buttons). Sie stellen einen bequemen Zugriff auf die wichtigsten Prozeduren der Menüs `Graph` und `Statistics` und zu den Options her.

3 Grundsätzliches zum Arbeiten mit SYSTAT

3.1 Arbeitsmodus

Es gibt drei verschiedene Möglichkeiten, in SYSTAT Befehle auszuführen:

- Mittels Menüs und dem folgenden Ausfüllen der entsprechenden Dialogboxen. (Übliche Art der Arbeit mit SYSTAT.)
- Interaktiv als Zeilenbefehle im untersten Teilfenster des Hauptfensters, dessen Menübalken auf `Interactive` gestellt sein muss.
- Durch Ausführen eines sogenannten Command-Files (= Batch-job) mit dem Befehl `File/Submit File...` im Hauptfenster oder `File/Submit Window` im untersten Teilfenster des Hauptfensters, dessen Menübalken auf `Untitled` (bzw. Name des Command-Files) gestellt sein muss.

Je nach Zeitaufwand ist die eine oder andere Variante zu wählen. Für grössere Routine-Jobs empfiehlt sich die Batch-Variante.

3.2 Help

SYSTAT stellt zwei Arten von Hilfe zur Verfügung: Eine Grundsätzliche, wo man über den Inhalt, einen Index oder eine noch etwas weiter gehende Suchhilfe vorgehen kann (`Help/Contents...` oder `Help/Search for Help on...`) und eine zweite Hilfe, die auf typische Probleme eingeht, die viele Benutzer antreffen (`Help/Knowledge Base (FAQ)...`).

Zudem bekommt man durch Anklicken von `Help` in einer Dialogbox direkt Information zum gewählten Befehl.

Anzumerken ist an dieser Stelle ausserdem, dass die Manuals von SYSTAT gut und kompakt aufgebaut sind. Leider werden die Beispiele nur mittels der Kommando-Sprache vorgeführt (und nicht via Menüs).

4 Import/Export

4.1 Import/Export von Daten

Excel

SYSTAT kann Excel-Files lesen. Keine Probleme sind zu erwarten, wenn folgende Regeln beachtet werden:

- Die Variablennamen sollten in der ersten Zeile des Excel-Worksheets stehen und die Systat-Konventionen erfüllen. D.h. sie müssen mit einem Buchstaben beginnen, dürfen keine Sonderzeichen enthalten und maximal 12 Zeichen lang sein. Sollten die Variablennamen länger als 12 Zeichen lang sein, werden einfach nur die ersten 12 genommen.
- Die erste Datenzeile sollte direkt unter der Variablennamenzeile stehen, vollständig sein und keine negativen Werte enthalten (sonst wird die Variable als Textvariable verkannt, was aber in SYSTAT problemlos wieder rückgängig gemacht werden kann).
- Es dürfen keine Formeln oder Formate im Worksheet enthalten sein.
- Die Daten müssen in einem Format gespeichert werden, das Systat versteht, d.h. Excel 4 Worksheet oder Excel 97 Workbook.
- Sind in einem Worksheet / Workbook mehrere Tabellen vorhanden, wird automatisch die erste (vorderste) Tabelle eingelesen.

Danach können die Daten mittels `File/Open/Data` im Hauptfenster importiert werden, wenn bei `Files of type:` "Microsoft Excel (*.xls)" ausgewählt wurde.

Genausogut können Daten nach Excel mittels `File/Save As...` im Datenfenster exportiert werden (bei `Save as type:` "Excel Files (*.xls)" auswählen).

Andere Formate

Ganz analog können auch andere Formate importiert/exportiert werden. SYSTAT akzeptiert folgende "Extensions":

ASCII	.txt	SAS Files	.sd2
Excel	.xls	SAS Transport Files	.xpt oder .tpt
DBase	.dbf	BMDP Data (TXT)	.por
SPSS	.sav	BMDP Data (BIN)	.sav
ArcView	.shp	BMDP Commands	.inp

Mit Hilfe von ODBC (Open Database Connectivity) können auch noch andere Formate verwendet werden (`File/Database Capture...`).

4.2 Export von Graphiken

Graphiken können mit `File/Save as...` im Graphikfenster in diversen Graphikformaten abgespeichert werden (Benutzte `Save as type:`). Innerhalb von Windows empfielt sich das Windows Metafile Format .wmf. Dieses Format kann von anderen Windows-Programmen eingelesen werden. Das Bild kann auch ins "Clipboard" kopiert (copy) werden und von dort aus in ein anderes Programm übernommen werden (paste).

4.3 Export ganzer Analysereports

Der Output von Statistik-Prozeduren und Graphiken erscheint im Output Teilfenster des Hauptfensters. Neben dem Output Teilfenster steht im Output Organizer eine Art Inhaltsverzeichnis des Outputs zur Verfügung. Das Output Teilfenster des Hauptfensters kann als Analysereport in folgenden Formaten gespeichert werden: SYSTAT Output (.syo), Rich Text Format (.rtf) und HTML Format (.htm). Dateien im .rtf-Format können dann mit Microsoft Word weiter verarbeitet werden.

5 Daten

Die Daten werden in einem Datenfenster angezeigt. Dabei entspricht jede Zeile einer Beobachtung und jede Spalte einer Variablen. Die Anzahl Variablen (Spalten) und die Anzahl Beobachtungen (Zeilen) werden nur durch den Speicherplatz limitiert.

5.1 Variablen-Formate

Durch Doppelklicken auf die Spaltenbezeichnung (z.B. "(VAR00002)") können die "Variable Properties" verändert werden (Name, Typ,...). **Variablennamen** müssen mit einem Buchstaben beginnen und dürfen maximal 12 Buchstaben lang sein. Als einziges Sonderzeichen ist der "Underscore" _ erlaubt und Variablennamen werden automatisch gross geschrieben.
SYSTAT unterscheidet zwei verschiedene Variablentypen:

In **numerischen Variablen** werden Zahlen bis zu 15 Dezimalstellen gespeichert. Die Zahlen können auch in Exponentialform eingegeben werden: 1.5E4 = $1.5 \cdot 10^4$.

Werte von **Textvariablen** ("String") können maximal 12 Buchstaben lang sein. Sie dürfen auch Sonderzeichen enthalten. Zudem unterscheidet SYSTAT zwischen Gross- und Kleinbuchstaben. **Der Variablenname einer Textvariablen muss mit einem $ enden (z.B. LAND$) !!**.

5.2 Fehlende Werte

Fehlende Werte (Missing Values) in einer numerischen Variablen werden als Punkt "." eingetragen. Bei einer Textvariablen wird die entsprechende Zelle einfach leer gelassen

5.3 Daten eingeben

Mit `File/New/Data` wird ein leeres Datenfenster geöffnet. Hier können die Daten eingetippt oder vom "Clipboard" hineinkopiert (paste) werden. Die Daten können mit dem Menü `File/Save as...` im Datenfenster gespeichert werden.

5.4 SYSTAT Datenfile öffnen

Ein SYSTAT Datenfile wird mit dem Menü `File/Open/Data` im Hauptfenster oder mit `File/Open...` im Datafenster geöffnet. In der Dialogbox muss der Pfad und der

Filename angegeben werden. SYSTAT speichert die Namen der zuletzt gebrauchten Datenfiles. Sie werden unten im Menü `File/Recent Data` aufgelistet und können durch Anklicken direkt geöffnet werden.

5.5 Transformieren von Variablen

Das Menü `Data` im Datenfenster offeriert mehrere Möglichkeiten, um Daten zu transformieren.

Transform/Let: Kreiert neue Variablen aus alten.
Z.B. `Let ALCOHOL = BEER + WINE + SPIRITS` oder
`Let L_INCOME = log(INCOME)`
Falls die Variable schon existiert, wird sie überschrieben. Sonst wird die neue Variable als letzte Spalte ins Datenfenster geschrieben.

Transform/If Then Let: Kreiert neue Variablen mit Bedingungen.
Z.B. `If TEMP < 5 then let KLIMA$ = 'kalt'`
Die neu generierte Variable wird am Schluss des Files angehängt.
ACHTUNG:Werte von nominalen Variablen müssen in Anführungs- und Schlusszeichen eingegeben werden! Sonst stürzt SYSTAT ab!

Rank: Die Variablenwerte werden durch ihre Ränge ersetzt. **Achtung: die ursprünglichen Variablenwerte gehen verloren!**

Sort: Das File wird nach einer oder mehreren Variablen sortiert.

Standardize: Die Werte einer Variablen werden durch die standardisierten Werte ersetzt. **Achtung: die ursprünglichen Variablenwerte gehen verloren!**

5.6 Auswählen von Beobachtungen und Variablen

Select cases: Den Befehl `Select cases...` findet man im Menü `Data`. Im Dialogfenster kann ein Selektionskriterium angegeben werden, z.B. `Select Temp>20`, oder `Select case<>12` (ungleich). Im Datenfenster werden die gewählten Beobachtungen mit einem * markiert. Für die weiteren Analysen und Graphiken werden anschliessend nur die ausgewählten Beobachtungen berücksichtigt. Die Selektion kann in der Dialogbox von `Data/Select cases...` durch Anklicken des Kästchen `Turn Select Off` wieder rückgängig gemacht werden.

Scatterplot Selection Tools: Vgl. Kapitel 6.1.

By Groups: Falls Sie Daten von verschiedenen Gruppen haben und eine bestimmte Analyse oder Graphik für jede Gruppe separat machen wollen, so benützen Sie `Data/By Groups...` im Hauptfenster. Dort können Sie die Variable angeben, welche die Beobachtungen in Gruppen einteilt. Beachte: Falls Sie wieder alle Beobachtungen für eine Analyse benützen wollen, müssen Sie die Gruppierung ausschalten (via `Data/By Groups...` und dann `Turn By Groups Off` anklicken).

5.7 Häufigkeitsangaben

Ist im Datenfile nicht jede Beobachtung einzeln gegeben, sondern ist nur angegeben, wieviele Beobachtungen in die einzelnen Klassen fallen (z.B. mit einer Variablen “Anzahl”), kann mit `Data/Frequency...` angegeben werden, welche Variable eine Häufigkeitsangabe ist. SYSTAT rechnet dann so, wie wenn jede Zeile mehrmals vorkommen würde (die Variable “Anzahl” gibt dann an, wie oft).

6 Graphische Möglichkeiten

Graphiken lassen sich mit Hilfe des `Graph`-Menüs und des Ausfüllens der darauf erscheinenden Dialogbox erstellen. Die Graphiken erscheinen jeweils im Hauptfenster. Um sie weiter zu bearbeiten, muss die Graphik mit einem Doppelklick in einem eigenen Graphikfenster geöffnet werden.

Vergrössert man die Graphik (anklicken und “auseinanderziehen”), muss auch der fuer die Graphik reservierte Raum vergrössert werden. Dazu wählt man `View/Page`. Durch Anklicken der Graphik wird der reservierte Raum sichtbar und kann auch vergrössert werden.

6.1 Selection Tools (Nur für Scatterplots)

Die folgenden Werkzeuge stellen eine Verbindung zwischen einem Scatterplot (oder Scatterplotmatrix) im Graphikfenster und dem Datenfenster her. Die nachfolgend beschriebenen Tools sind sehr wertvoll, um z.B. ’Ausreisser’ zu identifizieren.

Fadenkreuz: Wird mit dem Fadenkreuz ein Punkt im Scatterplot (im Graphikfenster) angeklickt, so wird die betreffende Beobachtung im Datenfenster markiert (highligthed) und es wird in der Statuszeile der Graphik die Beobachtungs-Nummer angezeigt.

Gestricheltes Rechteck und Lasso: Damit können mehrere Punkte markiert werden. Im Datenfenster sieht man, welche Beobachtungen den Punkten entsprechen. Für die folgenden Analysen werden nur die mit dem Rechteck oder Lasso ausgewählten Beobachtungen verwendet (genau wie bei `Select Cases...`). Die Selektion wird gelöscht, wenn das Lasso oder Rechteck irgendwo im Graphik-Hintergrund angeklickt wird, oder indem `Turn Select Off` in der Dialogbox des Befehls `Data/Select cases...` angekreuzt wird.

Quadrat mit Punkten: Dieses Tool ist vorallem für Scatterplotmatrizen sinnvoll. Klickt man auf dieses Kästchen leuchten die vorher mit dem Rechteck oder Lasso ausgewählten Beobachtungen in allen Plots auf.

6.2 Dynamic Explorer

Der Dynamic Explorer erscheint automatisch, wenn das Graphikfenster geöffnet wird. Mit ihm lassen sich unter anderem 3-dimensionale Graphiken rotieren oder die Anzahl

Balken in einem Histogramm variieren. Zudem können die Koordinatenachsen tranformiert werden, d.h. man kann graphisch überprüfen, ob z.B eine Log-Transformation angebracht wäre (Dabei bedeutet die Transformation “Power: 0” logarithmieren.).

6.3 Quick Graphs

Vgl. Kapitel 9.1.

6.4 Graphik drucken

Eine Graphik kann mit `File/Print...` im Graphikfenster gedruckt werden.

7 Statistik

Jede einzelne Statistikfunktionen hier zu erklären, macht wenig Sinn. Weiss man, welche Methode man anwenden möchte, findet man sich mit Hilfe der Menüs, der Help-Files und der Manuals recht schnell zurecht. Hier soll nur kurz der Befehl `Statistics/General Linear Model (GLM)/Estimate Model...` erläutert werden. Dieser Befehl kann beliebige lineare Modelle rechnen, insbesondere auch alles, was die Befehle `Statistics/Regression/Linear...` (kann nur Kovariablen behandeln) und `Statistics/Analysis of Variance (ANOVA)/Estimate Model...` (nimmt immer automatisch alle Wechselwirkungen ins Modell) können.

8 Programmiersprache

Manchmal ist es effizienter, nicht mit Menüs zu arbeiten. Zudem lassen sich gewisse Dinge nur via Programmiersprache erledigen. Einzelne Befehle lassen sich im unteren Teilfenster des Hauptfensters (Menübalken auf `Interactive`) direkt eingeben. Programme können im unteren Teilfenster des Hauptfensters (Menübalken auf `Untitled`) eingegeben und dann via `File/Submit Window` ausgeführt werden. Es besteht auch die Möglichkeit, direkt ein Programm-File via `File/Submit File...` auszuführen.

Die Befehle sämtlicher Auswertungen werden automatisch im “Log-File” mitprotokolliert, das im unteren Teilfenster des Hauptfensters (Menübalken auf `Log`) angeschaut werden kann. Dieses “Log-File” kann abgespeichert werden (bekommt die “Extension” .syc) und bietet oft einen guten Startpunkt für die Automatisierung des verwendeten Verfahrens durch ein Programm. Mit Hilfe des “Log-Files” kann auch bis zu einem gewissen Grad selbst die Syntax der Befehle studiert werden.

Commands sind folgendermassen aufgebaut: `Command Argument / Option List`
Der Befehl `PLOT var1*var2 / Label=var3$` erstellt einen Scatterplot der Variablen var1 und var2 und beschriftet die Punkte mit var3$ (Gross- und Kleinschreibung spielen keine Rolle bei der Eingabe von Commands). Genauere Anleitungen zur Syntax findet man in den SYSTAT-Manuals.

Im unteren Teilfenster des Hauptfensters (Menübalken auf `Interactive`) lassen sich durch Druck auf die Funktionstaste F9 ältere Befehle wiederholen.

9 Diverses

9.1 Global Options

Mit dem Befehl `Edit/Options...` lassen sich verschiedene Einstellungen vornehmen:

Im Blatt `Data` lässt sich unter `Data Editor Cursor` einstellen, ob beim Eingeben von Daten der "Cursor" nach rechts oder unten "springt", wenn die `<Enter>`-Taste gedrückt wird.

Im Blatt `Output` kann unter `Output Results` die Länge (Ausführlichkeit) des Outputs bestimmt werden. Zu empfehlen ist, `Display statistical quickgraphs` anzukreuzen. Dadurch werden bei vielen Statistikfunktionen gleich noch informative Graphiken erstellt (z.B. Tukey-Anscombe Plot bei der linearen Regression). Diese beiden Optionen lassen sich auch mit Quick Buttons einstellen.

9.2 Taschenrechner

Oft besteht der Wunsch 'husch husch' eine mathematische Operation auszuführen. Im unteren Teilfenster des Hauptfensters kann man in der Art eines Taschenrechners Rechnungen durchführen, sofern dessen Menübalken auf `Interactive` gestellt ist:

```
> calc Log(123)
```

Der Logarithmus von 123 wird ausgerechnet und angezeigt (im Output Teilfenster): 4.812

9.3 Neuerungen in Version 10

In SYSTAT Version 10 wurden unter anderem folgende Neuerungen eingeführt:

- Power Analysis: Beantwortet Fragen der Form: Wie gross muss die Stichprobe sein, damit ein t-Test eine Macht von 0.9 erreicht?
- Mixed Regression: Berechnet Regressionen mit festen und zufälligen Effekten.
- "Graph Express": Separates Programm, das den SAS Benutzern das Erstellen von SYSTAT Graphiken erleichtert.

9.4 Systemanforderungen

Systemanforderungen für SYSTAT 10 für Windows:

- Pentium/ clone or above
- Microsoft Windows 95/98/2000/NT
- 32 MB RAM (minimum)
- 30MB hard disk space
- SVGA monitor

9.5 Weitere Information

Mehr Infos auch unter

```
http://www.systat.com
http://www.stat.math.ethz.ch/~statsoft/stat.programme/systat.html
```

.