

**Einführung in
SYSTAT für Windows
Version 11**

Seminar für Statistik
ETH Zürich

März 2005

Inhaltsverzeichnis

1	Kurze Beschreibung von SYSTAT	1
1.1	Allgemeines	1
2	Aufbau der Oberfläche	1
2.1	Der Viewspace	1
2.2	Der Workspace	2
2.3	Der Commandspace	2
3	Grundsätzliches zum Arbeiten mit SYSTAT	2
3.1	Arbeitsmodus	2
3.2	Hilfe	2
4	Import/Export	3
4.1	Import/Export von Daten	3
4.2	Export von Graphiken	3
4.3	Export ganzer Analysereports	4
5	Daten	4
5.1	Variablen-Formate	4
5.2	Daten eingeben	4
5.3	Fehlende Werte	4
5.4	SYSTAT Datenfile öffnen	5
5.5	Transformieren von Variablen	5
5.6	Auswählen von Beobachtungen und Variablen	5
5.7	Häufigkeitsangaben	6
6	Graphische Möglichkeiten	6
6.1	Selection Tools (Nur für Scatterplots)	6
6.2	Dynamic Explorer	6
6.3	Quick Graphs	7
6.4	Graphik drucken	7
7	Statistik	7
8	Programmiersprache	7
9	Diverses	8
9.1	Global Options	8
9.2	Taschenrechner	8
9.3	Neuerungen in Version 11	8

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 sind durch die Menüsteuerung zugänglich, so dass die Bedienung relativ komfortabel ist. Es 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, wie z.B. die Programmiersprache BASIC, mit der Daten eingelesen und behandelt werden können. Allerdings sind diese Möglichkeiten nicht mit denjenigen grösserer Produkte wie R/S-Plus oder SAS vergleichbar.

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

Besonders erwähnenswert sind 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.

Weitere Informationen sind auf <http://www.systat.com> erhältlich.

2 Aufbau der Oberfläche

Beim Aufstarten von SYSTAT erscheint das Hauptfenster, das aus drei Teilfenstern (Workspace links, Viewspace rechts und Commandspace unten), einem Menübalken und einem Quickbutton-Balken besteht. Die Menüs und Quickbuttons erlauben, die wichtigsten Operationen einfach auszuführen.

2.1 Der Viewspace

Im Viewspace rechts erscheinen die Resultate der ausgeführten Operationen im Blatt **Output Pane**. Durch Scrollen kann man ältere Resultate wieder hervorholen. Text (z.B. Titel) kann beliebig hinzugefügt oder gelöscht werden.

Die Resultate einer Graphik-Operation (z.B. Histogramme) erscheinen normalerweise auch im Blatt **Output Pane**. Mit einem Doppelklick auf solche Graphiken wird das Blatt **Graph Editor** geöffnet. Jetzt kann man sie z.B. genauer betrachten (Ausreissersuche), die Plot-Parameter verändern (Blatt **Dynamic Explorer** im Workspace) oder separat abspeichern.

Um die Daten in Tabellenform darzustellen, wählt man das Blatt **Data Editor**.

2.2 Der Workspace

Im **Output Organizer** des **Workspace** (linkes Teilfenster) erscheint ein Inhaltsverzeichnis des **Outputs**. Indem man einen entsprechenden Eintrag im Inhaltsverzeichnis anklickt, kann man sich mit dem **Output Organizer** bequem im **Output Pane** bewegen. Weiter kann man mit **Cut**, **Copy**, **Paste** die Resultate reorganisieren. Die Änderungen wirken sich auch auf das **Output Pane** aus.

2.3 Der Commandspace

Das untere Teilfenster (**Commandspace**) besitzt drei Blätter: **Interactive**, **Untitled** und **Log**. Im **Interactive** Blatt können Befehle via Tastatur eingegeben werden, sie werden nach drücken der **<Enter>**-Taste ausgeführt. Im mittleren Blatt können ganze Programmdateien (.**sys**) 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.

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. Dies ist die übliche Art der Arbeit mit SYSTAT.
- Interaktiv als Zeilenbefehle im **Interactive** Blatt im **Commandspace**.
- Durch Ausführen eines sogenannten **Command-Files** (= **Batch-Job**) mit dem Befehl **File/Submit File...** oder **File/Submit Window**, wobei das Blatt **Untitled** (bzw. Name des **Command-Files**) ausgewählt sein muss.

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

3.2 Hilfe

SYSTAT stellt wie unter **Windows** üblich drei Arten von Hilfen zur Verfügung: Den Inhalt erreicht man via **Help/Contents...**, den Index via **Help/Index...** und eine noch etwas weiter gehende Suchhilfe via **Help/Search for Help on...** Zudem erhält man durch Anklicken von "?" in der linken unteren Ecke jeder Dialogbox direkt Informationen zum gewählten Befehl.

Zu bemerken ist an dieser Stelle ausserdem, dass die **Manuals** von SYSTAT gut und kompakt aufgebaut sind.

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 und vollständig sein.
- Es dürfen keine Formeln im Worksheet enthalten sein.
- Sind in einem Worksheet / Workbook mehrere Tabellen vorhanden, wird man nach der gewünschten Tabelle gefragt.

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

Genauso gut können Daten nach Excel mittels **File/Save As...** exportiert werden (bei **Save as type:** "Excel Files (.xls)" auswählen).

Andere Formate

Ganz analog können auch andere Formate importiert/exportiert werden. SYSTAT akzeptiert unter anderem 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		

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 **Graph Editor** in diversen Graphikformaten abgespeichert werden (**Save as type:**). Innerhalb von Windows empfiehlt sich das Windows Metafile Format (.wmf). Dieses Format kann von anderen Windows-Programmen eingelesen werden. Das Bild kann auch ins "Clipboard" kopiert werden (**Edit/Copy Graph**) 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 Viewspace im Output Pane. Neben dem Output Pane steht im Output Organizer (im Workspace) eine Art Inhaltsverzeichnis des Outputs zur Verfügung. Das Output Pane 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 z.B. mit Microsoft Word weiter verarbeitet werden.

5 Daten

Die Daten werden im Data Editor (im Viewspace) 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, ...). **Variablenamen** müssen mit einem Buchstaben beginnen und dürfen maximal 12 Zeichen lang sein. Als einziges Sonderzeichen ist der Underscore "_" erlaubt. Variablenamen werden automatisch gross geschrieben. Verweist man in Skripts auf eine Variable, so spielt die Gross- und Kleinschreibung keine Rolle.

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 Zeichen lang sein. Sie dürfen auch Sonderzeichen enthalten. Zwischen Gross- und Kleinbuchstaben wird unterschieden. **Der Name einer Textvariablen muss mit einem \$ enden (z.B. LAND\$)!**

5.2 Daten eingeben

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

5.3 Fehlende Werte

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

5.4 SYSTAT Datenfile öffnen

Ein SYSTAT Datenfile (.syd) wird mit dem Menü **File/Open/Data...** geöffnet. In der Dialogbox muss der Pfad und der Filename angegeben werden. SYSTAT speichert die Namen der zuletzt gebrauchten Datenfiles. Sie werden im Menü **File/Recent Data** aufgelistet und können durch Anklicken direkt geöffnet werden.

5.5 Transformieren von Variablen

Das Menü **Data** 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.

Rank: Die Variablenwerte werden durch ihre Ränge ersetzt.

Achtung: die ursprünglichen Variablenwerte gehen verloren, falls keine Kopie in einer neuen Spalte erzeugt wurde!

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, falls keine Kopie in einer neuen Spalte erzeugt wurde!**

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 **Data Editor** 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 Off** wieder rückgängig gemacht werden.

Scatterplot Selection Tools: Vgl. Kapitel 6.1.

By Groups: Falls man Daten aus verschiedenen Gruppen hat und eine bestimmte Analyse oder Graphik für jede Gruppe separat machen will, so benützt man **Data/By Groups...** Dort kann man die Variablen angeben, welche die Beobachtungen in Gruppen einteilen. Falls man wieder alle Beobachtungen für eine Analyse benützen will, muss man die Gruppierung ausschalten (via **Data/By Groups...** und dann **Turn 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 **Output Pane**. Um sie weiter zu bearbeiten, muss eine Graphik mit einem Doppelklick im **Graph Editor** geöffnet werden. Die Möglichkeiten zur Nachbearbeitung sind (im Vergleich zu anderen Programmen) sehr begrenzt. Es lohnt sich deshalb, die Parameter vorwiegend sorgfältig zu wählen.

Vergrößert man die Graphik (anklicken und "auseinanderziehen"), muss ev. auch der für die Graphik reservierte Raum vergrößert werden. Dazu wählt man **View/Page View**. Durch Anklicken der Graphik wird der reservierte Raum sichtbar und kann auch verändert werden.

6.1 Selection Tools (Nur für Scatterplots)

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

Fadenkreuz: Wird mit dem Fadenkreuz (in der Toolbar wählbar) ein Punkt im Scatterplot (im **Graph Editor**) angeklickt, so wird die betreffende Beobachtung im **Data Editor** markiert (highlighted) und es wird in der Statuszeile die Beobachtungs-Nummer angezeigt.

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

Quadrat mit Punkten: Dieses Tool ist vor allem 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 der **Graph Editor** geöffnet wird. Mit ihm lassen sich unter anderem 3-dimensionale Graphiken rotieren oder

die Anzahl Balken in einem Histogramm variieren. Zudem können die Koordinatenachsen transformiert 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...` aus dem Graph Editor heraus gedruckt werden.

7 Statistik

Jede einzelne Statistikfunktion hier zu erklären, würde wenig Sinn machen. Wenn man weiss, welche Methode man anwenden möchte, so findet man sich mit Hilfe der Menüs, der Help-Files und der Manuals recht schnell zurecht.

Hier soll nur kurz der Befehl `Analysis/General Linear Model (GLM)/Estimate Model...` erwähnt werden. Dieser Befehl kann beliebige lineare Modelle rechnen, insbesondere auch alles, was die Befehle `Analysis/Regression/Linear/Least Squares...` (kann nur Kovariablen behandeln) und `Analysis/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 `Commandspace` im Blatt `Interactive` direkt eingeben. Programme können im Blatt `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 `Commandspace` im Blatt `Log` angeschaut werden kann. Dieses "Log-File" kann abgespeichert werden (es erhält 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 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 *Interactive* Blatt des Commandspace lassen sich mit F9 ältere Befehle wiederholen.

9 Diverses

9.1 Global Options

In *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. Ab Version 11 stehen zwei “Taschenrechner” zur Verfügung:

Über *Utilities/Calculator...* kann man übliche mathematische Operationen ausführen, z.B. $\text{Log}(123)$. Das Resultat wird im *Output Pane* angezeigt: 4.812.

Über *Utilities/Probability Calculator/...* lassen sich Eigenschaften wie Dichte, Verteilungsfunktion, Quantile etc. von vielen univariaten Verteilungen berechnen.

9.3 Neuerungen in Version 11

Die aktuelle Version 11 enthält etliche Erweiterungen und Verbesserungen:

- Einheitliche Benutzeroberfläche und gut strukturiertes Menüsystem.
- Gut strukturiertes Hilfesystem mit vielen Beispielen.
- Wesentliche Erweiterungen zur Unterstützung von Simulationsstudien und Teilstichprobenanalysen (Bootstrap, Jackknife).
- Zusätzliche Methoden im Bereich Versuchsplanung und Prozessoptimierung.
- Weitgehende Vereinheitlichung von Dialogfenstern für Graphiken.

Ein ausführlicher Vergleich zwischen den Versionen 11 und 10.2 findet man auf <http://www.systat.com> unter *Home:Products:Systat:Upgrade Comparison*.