

Abbildung 1: Der Startbildschirm von STATISTICA.

1 Erste Schritte in STATISTICA

Das Statistik-Programm STATISTICA umfasst eine sehr große Sammlung statistischer Verfahren und grafischer Methoden, ist aber trotzdem ein sehr überschaubares und leicht bedienbares Programm geblieben. Dazu trägt auch die benutzerfreundliche Oberfläche bei, die den meisten Computeranwendern ohnehin von EXCEL oder ähnlichen Tabellenkalkulationsprogrammen her vertraut erscheinen dürfte.



Wer gerne mit den seit Office 2007 bekannten Multifunktionsleisten („Ribbons“) arbeitet, der kann diese seit Version 9 im Menü *Ansicht* auswählen. Um von diesen Ribbons wieder zurück zur klassischen Menüstruktur zu gelangen, wählt man entweder ganz oben links den Punkt *Menüs* oder deaktiviert die Ribbons ganz oben rechts unter *Optionen*.

Wie in Abbildung 1 zu sehen, verfügt das Tabellenblatt von STATISTICA über ähnliche Funktionen wie das von EXCEL, so kann man beispielsweise auch hier eine Reihe von Zahlen mittels Mauszeiger fortsetzen.

Abweichend von EXCEL ist allerdings die Funktion von Zeilen und Spalten; während bei EXCEL Zeilen und Spalten prinzipiell gleichwertig sind, sind in STATISTICA, wie bei einer Datenbank, die Spalten den Zeilen übergeordnet. Jedoch erlaubt STATISTICA, bei Bedarf ein Tabellenblatt zu transponieren. Die Spalten, hier *Variablen* genannt, bezeichnen verschiedene Merkmale, in den Zeilen, den *Fällen*, werden die einzelnen konkreten Messwerte gesammelt. Ein Beispiel:

Beim innerdeutschen Vergleich der PISA-Studie (OECD PISA 2000, zusammengefasst vom Berliner Max-Planck-Institut für Bildungsforschung 2002) wurde u. a. die Fähigkeit im Lesen,

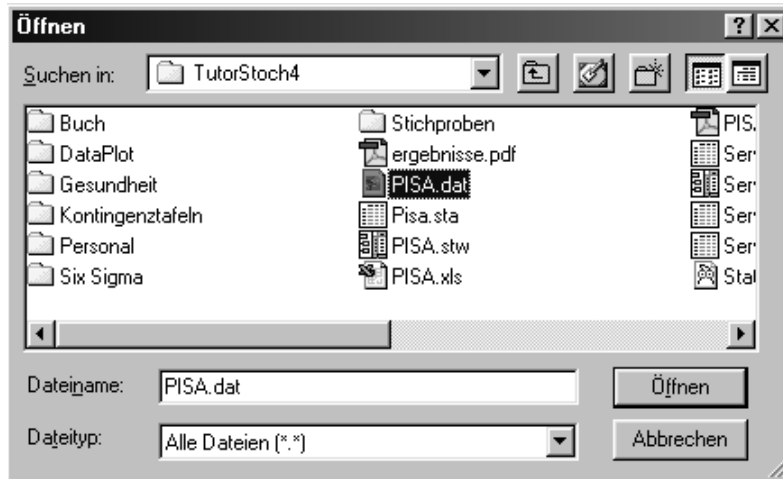


Abbildung 2: Öffnen der Datendatei.

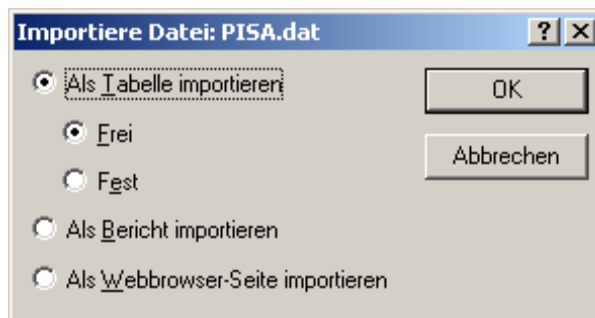


Abbildung 3: Die Datendatei als Tabelle importieren (Dialog der Version 8).

in der Mathematik und den Naturwissenschaften von Schülern der Jahrgangsstufe 9 in 14 deutschen Bundesländern (ohne Hamburg und Berlin) gemessen; in diesem Fall wären ‘Lesefähigkeit’, ‘Math. Fähigkeit’ und ‘Naturwiss. Fähigkeit’ die drei Variablen, die Fälle würden mit den einzelnen Bundesländern bezeichnet und die jeweiligen Werte enthalten. Nehmen wir dieses Beispiel, um uns ein wenig vertraut zu machen mit STATISTICA.

Die konkreten Daten befinden sich in der (Text-)Datei `PISA.dat` und müssen erst in ein STATISTICA-Tabellenblatt importiert werden.² Dazu wählt man wie gewohnt den Menüpunkt *Datei* → *Öffnen*, worauf ein Dialogfenster wie in Abbildung 2 erscheint; nachdem man bei *Dateityp* *Alle Dateien* eingestellt hat, wird die gesuchte Datei angezeigt und kann ausgewählt werden. Das weitere Vorgehen ist nun versionsabhängig, im Folgenden beschreiben wir die Vorgehensweise ab Version 8. Benutzer der Versionen 6 und 7 gehen aber prinzipiell analog vor, für Details sei auf Weiß (2006) verwiesen.

Im nächsten Dialog, vgl. Abbildung 3, wählt man *Als Tabelle importieren: Frei*. Nun öffnet sich der recht umfangreiche und etwas unübersichtliche Dialog *Text (Frei) Import-Optionen* aus Abbildung 4, mit dessen Hilfe man eine Reihe von Einstellungen bzgl. Trennzeichen, Zeilenzahl, u. Ä., treffen kann. Dabei wird dem Benutzer automatisch ein Vorschlag gemacht,

²Es sei an dieser Stelle nochmal erwähnt, dass alle in diesem Text besprochenen Datensätze von der im Vorwort angegebenen Seite heruntergeladen werden können.

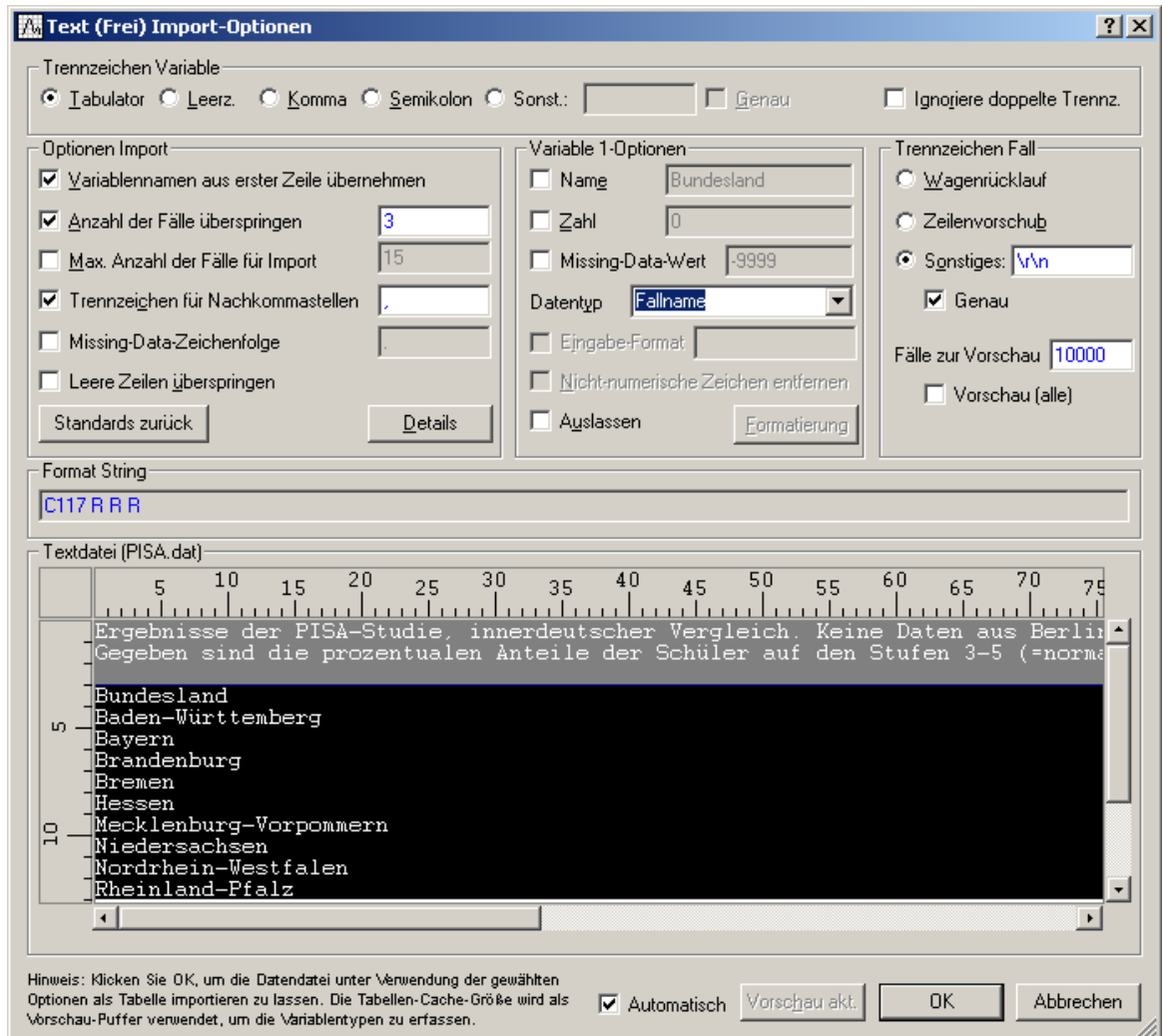


Abbildung 4: Die Optionen beim Import von Textdateien (Dialog der Version 8).

den man auf Wunsch aber auch abändern kann: Im obersten Teil des Dialogfensters kann man das Trennzeichen auswählen, und im Feld der Import-Optionen festlegen, bei welchem Fall der Import beginnen/enden soll (indem man die Zahl der zu Beginn zu überspringenden Fälle festlegt und die anschließend zu importierende Fallzahl), welches Dezimaltrennzeichen verwendet wird, u.Ä. Im unteren Teil des Fensters gibt es eine Vorschau, die sich auch automatisch aktualisiert, falls man ein Häkchen bei der entsprechenden Option ganz unten gesetzt hat. Nach Klick auf *OK* startet dann der Import. Für das hier betrachtete Beispiel der Pisa-Daten wären folgende Einstellungen passend: *Trennzeichen Variable: Tabulator* und Häkchen im Feld *Optionen Import* genau bei *Variablennamen aus erster Zeile* und *Anzahl ... überspringen: 3*. Ferner müsste in der Vorschau die erste Spalte (mit den Namen der Bundesländer) markiert und dann im Feld *Variable 1-Optionen* der *Datentyp: Fallname* gewählt werden, siehe Abbildung 4.

Nun ist es geschafft! Durch Klick in die linke obere Ecke kann die Tabelle markiert werden, durch erneuten Klick mit der rechten(!) Maustaste ins Tabellenblatt öffnet sich ein PopUp-Menü und dort kann unter *Format* der Befehl *Optimale Höhe/Breite* betätigt werden, damit unsere Daten und Randbeschriftungen tatsächlich gut lesbar sind. Hierzu vergleiche man

Abbildung 5. Die Breite der Fallnamen reduziert man mit Hilfe der Maus (linke Maustaste gedrückt halten), siehe Abbildung 6.

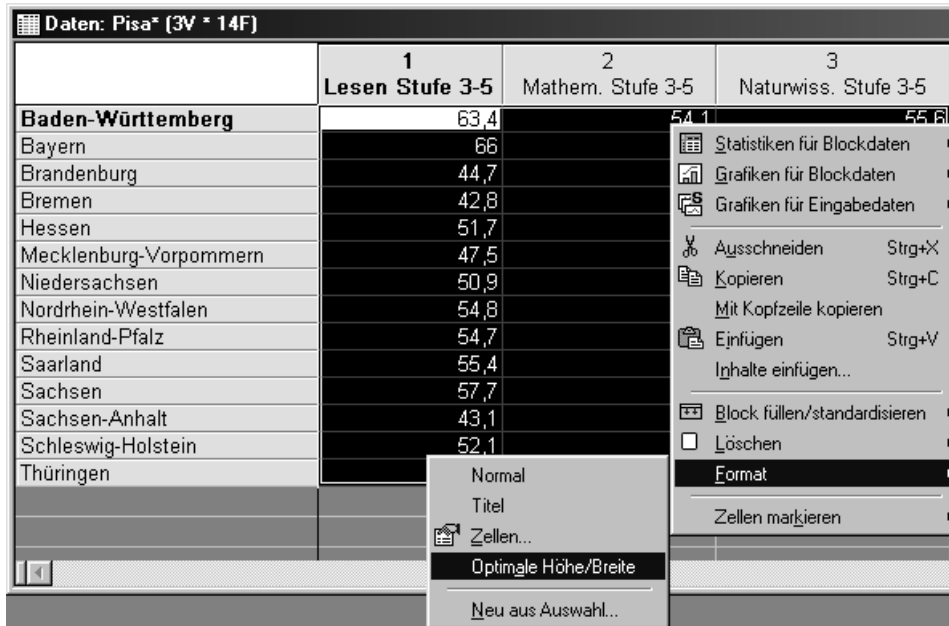


Abbildung 5: Die gewünschten Daten wurden nun als Tabellenblatt importiert.

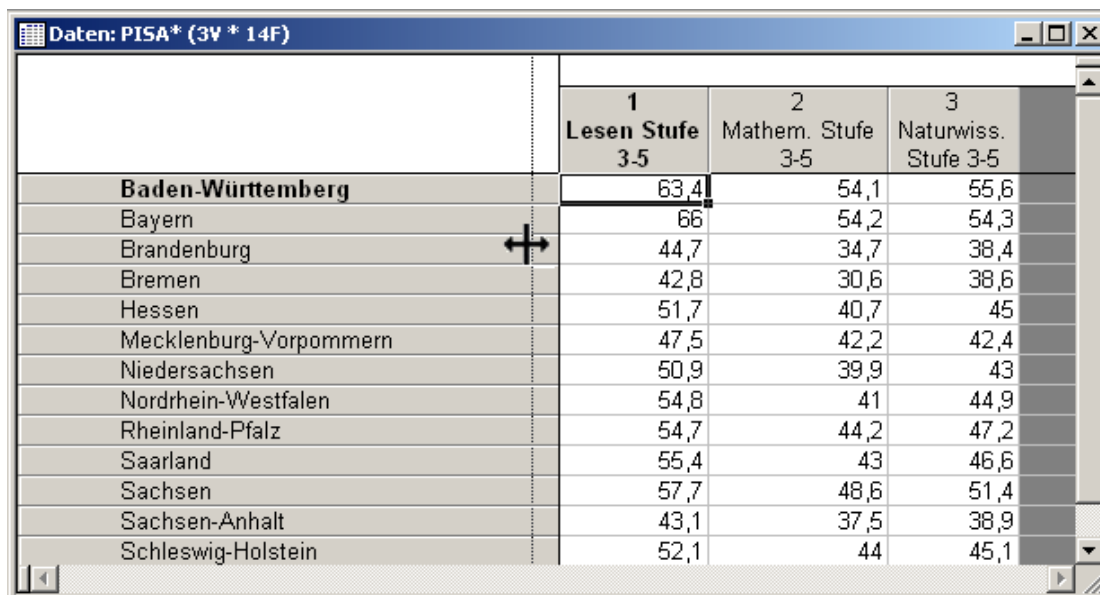


Abbildung 6: Die Breite der Fallnamen anpassen.

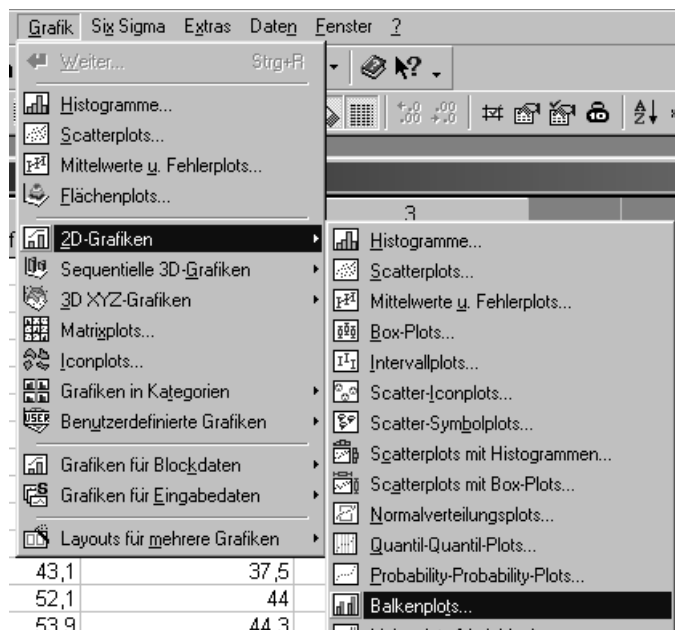


Abbildung 7: Wir erstellen ein Balkendiagramm der Daten.



Abbildung 8: Als Variable wählen wir 'Lesen Stufe 3-5'.

Betrachten wir die Daten genauer: In den Disziplinen Lesen, Mathematik und Naturwissenschaften wurde jeweils der Anteil von Schülern an den Stufen 3 bis 5 gemessen, das heißt im Klartext: Wenn wir unter der Rubrik Lesen bei Baden-Württemberg eine 63,4 finden, so bedeutet dies, dass 63,4 % aller dortigen Schüler der 9. Klasse über mittlere bis sehr gute Lesekenntnisse verfügen. Wir wollen uns die Ergebnisse der Variablen 'Lesen' grafisch veranschaulichen, und zwar mit Hilfe eines Balkendiagramms. Dazu wählen wir wie in Abbildung 7 den Menüpunkt *Grafik* → *2D-Grafiken* → *Balkenplots* und anschließend wie in Abbildung 8 als Variable 'Lesen Stufe 3-5', stellen horizontale Ausrichtung ein und betätigen dann den *OK*-Knopf.

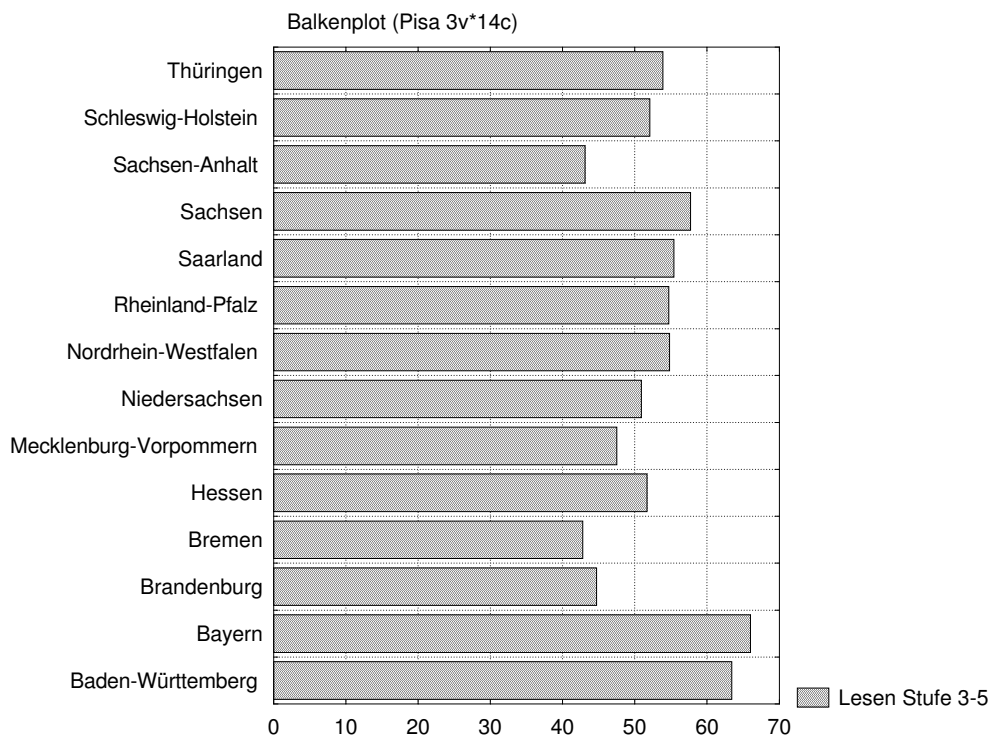


Abbildung 9: Ein Balkendiagramm der Lesedaten.

Nun wird eine *Arbeitsmappe* (*Workbook*) erstellt und die gewünschte Grafik in Selbige eingefügt. Das Ergebnis ist in Abbildung 9 zu sehen. Bis zu dem Zeitpunkt, an dem wir diese Arbeitsmappe zum ersten Mal schließen, werden alle weiteren Analyseresultate, auch wenn sie von anderen Datenblättern herrühren sollten, in dieser Arbeitsmappe abgelegt. Man erkennt dies zum Beispiel im Hintergrund von Abbildung 12. Beim Abspeichern der Arbeitsmappe bekommt diese die Dateiendung *.stw*.

Setzt man zu einem späteren Zeitpunkt Analysen fort, die man gern in einer früheren Arbeitsmappe ablegen möchte, gibt es drei Möglichkeiten:

1. Man analysiert wie gewohnt, öffnet am Ende die gewünschte Arbeitsmappe und zieht mit gehaltener Maustaste die Resultate in die Ziellarbeitsmappe.
2. Man wählt vor Beginn der Analysen den Punkt *Datei* → *Ausgabemanager*, woraufhin der *Optionen*-Dialog geöffnet wird, mit der Karte *Ausgabemanager* in Front. Dort wählt man *Alle Ergebnisse (...)* in → *Arbeitsmappen* → *Bestehende Arbeitsmappe* und in letzterem Feld die gewünschte Ziellarbeitsmappe. Nach Bestätigen mit *OK* werden alle weiteren Analyseresultate dort eingefügt. **Achtung:** STATISTICA merkt sich diese Einstellung. Solange Sie daran nichts ändern, werden von nun an immer Resultate in genau dieser Arbeitsmappe abgelegt.
3. Analog eben wählt man im *Ausgabemanager* *Arbeitsmappe enthält Datendatei* (seit Version 9 ist dies ohnehin die Standardeinstellung). Dann bleibt alles wie unter 1., es sei denn, der analysierte Datensatz befindet sich in der Arbeitsmappe selbst, siehe auch S. 13. Öffnet man in diesem Fall die Arbeitsmappe später erneut und wählt wieder den Datensatz der Arbeitsmappe aus, dann werden die Resultate auch automatisch in dieser Arbeitsmappe abgelegt.



Seit Version 8 wird übrigens zu erstellten Analyseresultaten ein Makro mit in die Arbeitsmappe abgelegt, so dass man einmal gemachte Analysen mit neuen/geänderten Daten wiederholen oder fortsetzen kann. Erkennbar ist dies am roten Pfeil der Arbeitsmappenordner, den man mit der rechten Maustaste anklickt. Wählt man im sich öffnenden PopUp-Menü den Punkt *Analyse wiederholen*, so wird die Analyse, die zu den gemachten Ausgaben führte, wiederholt, wobei man optional auch einen anderen Datensatz wählen kann. Ferner kann man wählen, dass die alten Resultate durch die neuen ersetzt werden. Die Wahl von *Analyse fortsetzen* erlaubt es einem dagegen, die Analyse mit geänderten Einstellungen zu wiederholen.

Abschließend sei bemerkt, dass man seit Version 7 sog. *Variablenbündel* definieren kann, und zwar immer vom Variablendialog einer Analyse aus. Bei aktuellen STATISTICA-Versionen verfügt dieser nämlich, im Gegensatz zu Abbildung 8, noch über den zusätzlichen Knopf *[Bündel]...*, nach dessen Betätigung sich der *Variablenbündel-Manager* öffnet. Hier kann man bestehende Bündel verändern, oder durch Klick auf *Neu...* ein neues definieren. Im letztgenannten Fall muss man zuerst einen Namen für das Bündel angeben, anschließend erscheint ein Variablenauswahldialog wie in Abbildung 8. Hier wählt man die Variablen für das Bündel aus und bestätigt.

Zukünftig werden, wenn die Tabelle im *.sta*-Format ab Version 7 gespeichert wird, im Variablendialog neben den Variablen selbst auch die definierten Bündel angezeigt. Klickt man auf ein solches, so werden auf einen Schlag genau jene Variablen ausgewählt, die im Bündel vermerkt sind. Gerade bei Dateien mit vielen Variablen, bei denen immer wieder die gleiche Auswahl einer Analyse unterzogen wird, ist dies eine große Erleichterung.