Datenanalyse und Modellierung mit STATISTICA

Von
Christian Weiß

Oldenbourg Verlag München Wien

Dipl. Math. Christian Weiß ist Wissenschaftlicher Mitarbeiter am Lehrstuhl für Statistik der Julius-Maximilians-Universität Würzburg.

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Rosenheimer Straße 145, D-81671 München
Telefon: (089) 45051-0
oldenbourg.de

Lektorat: Margit Roth
Herstellung: Anna Grosser
Umschlagkonzeption: Kochan & Partner, München
Gedruckt auf säure- und chlorfreiem Papier
Druck: Grafik + Druck, München

ISBN 3-486-57959-2
ISBN 978-3-486-57959-8

Das Wahrscheinliche (daß bei 6 000 000 000 Würfen mit einem regelmäßigen Sechserwürfel annähernd 1 000 000 000 Einser vorkommen) und das Unwahrscheinliche (daß bei 6 Würfen mit demselben Würfel einmal 6 Einser vorkommen) unterscheiden sich nicht dem Wesen nach, sondern nur der Häufigkeit nach, wobei das Häufigere von vornherein als glaubwürdiger erscheint. Es ist aber, wenn einmal das Unwahrscheinliche eintritt, nichts Höheres dabei, keinerlei Wunder oder Derartiges, wie es der Laie so gerne haben möchte. Indem wir vom Wahrscheinlichen sprechen, ist ja das Unwahrscheinliche immer schon inbegriffen und zwar als Grenzfall des Möglichen, und wenn es einmal eintritt, das Unwahrscheinliche, so besteht für unsereinen keinerlei Grund zur Verwunderung, zur Erschütterung, zur Mystifikation.

(Max Frisch, *Homo Faber*)

Vorwort

STATISTICA ist ein weit verbreitetes Softwarepaket der Firma StatSoft, welches die statistische und grafische Analyse von Datenmaterial erlaubt. Bereits das Basismodul „STATISTICA Standard“ stellt eine Reihe grundlegender Verfahren der deskriptiven wie auch der induktiven Statistik zur Verfügung, zudem zahlreiche Werkzeuge zur grafischen Analyse. Durch den gezielten Erwerb zusätzlicher Module lässt sich das Repertoire statistischer Datenanalyseverfahren nach Bedarf erweitern; einen Überblick über die angebotenen Module findet der Leser in Anhang E. Trotz seiner Popularität gibt es im deutschsprachigen Raum jedoch erstaunlich wenig Literatur zu STATISTICA, was ein Grund war, das vorliegende Buch zu verfassen.

Die Notwendigkeit statistischer Datenanalyse besteht in vielen Berufs- und Forschungszweigen. Zu nennen sind hier etwa das Handels- und Dienstleistungsgewerbe, die Fertigungsindustrie, sowie Biologie, Medizin, Chemie, Physik, Geographie, Wirtschafts-, Sozialwissenschaften, u. v. m. Für einen großen Teil der dort anfallenden Fragestellungen bietet STATISTICA, zumindest nach Erwerb geeigneter Module, Lösungsmöglichkeiten an. Ein besonderer Vorzug von STATISTICA gegenüber manch anderem Produkt ist dabei die klare und leicht bedienbare Benutzeroberfläche. Diese ist an herkömmlichen Tabellenkalkulationsprogrammen orientiert und erleichtert somit den Einstieg in die Verwendung von STATISTICA. Ferner erlaubt STATISTICA den Import und Export verschiedenster Datenformate, was die Praxistauglichkeit deutlich erhöht.

Bei dem vorliegenden Buch zur *Datenanalyse und Modellierung mit STATISTICA* handelt es sich um einen einführenden Text zu STATISTICA ab Version 6.0. Neuerungen, die erst mit Version 7 implementiert wurden, sind im Text explizit als solche gekennzeichnet, u. a. durch das am Rand stehende Symbol. Somit kann dieses Buch von einer breiten Leserschaft verwendet werden. Es wird dabei stets eine deutschsprachige Version von STATISTICA zu Grunde gelegt, ferner wurden Bildschirmausdrucke stets im klassischen Windows-Stil gemacht. Große Teile des Buches gehen auf Materialien zurück, die ich für Kurse entwickelt habe, welche die Stochastik-Vorlesungen von Herrn Prof. Dr. Rainer Göb begleiten und ergänzen. Ferner fanden frühere Versionen des Manuskriptes ihren Einsatz im Rahmen des Kompaktseminars *Qualitäts- und Risikomanagement* des MBA Weiterbildungsstudienganges *Business Integration*.

Der Text ist im Stile eines Lehrbuchs verfasst worden, darüberhinaus aber auch eingeschränkt als Nachschlagewerk verwendbar. Wer nur an speziellen technischen Details oder Beschreibung einzelner Menüs interessiert ist, der findet sicherlich in der Hilfe von STATISTICA vergleichbare Informationen. Ebenso kann das im Internet verfügbare, auf Englisch verfasste, elektronische Handbuch, siehe hierzu StatSoft (2004), von Nutzen sein. Die Idee des vorliegenden Textes ist es dagegen, den mit STATISTICA noch nicht oder nur partiell vertrauten Leser auf (hoffentlich) leicht verständliche und kurzweilige Weise an die Möglichkeiten von STATISTICA heranzuführen. Obwohl bei den besprochenen Verfahren auch immer die statistischen Hintergründe erläutert werden, handelt es sich beim vorliegenden Buch nicht um ein Lehrbuch zur Statistik, sondern primär um ein Buch für Anwender. Alle im Text vorgestellten Verfahren werden anhand von Beispielen illustriert. Die dabei verwendeten Datensätze, abgespeichert im `.sta`-Format für Version 6.0, finden sich bei den Informationen zum Buch auf der Verlags-Website `oldenbourg.de`. Diese werden auch für die Aufgaben zu den einzelnen Kapiteln benötigt. Es wird dem Leser empfohlen, diese Datensätze am besten vor Beginn der Lektüre herunterzuladen und im Laufe dieser alle Beispiele selbst durchzuführen.

Die Voraussetzungen, die an den Leser gemacht werden, sind, zumindest nach Auffassung des Autors, minimal – eine gewisse Routine im Umgang mit handelsüblicher Windows-Software sowie rudimentäre Vertrautheit mit Tabellenkalkulationsprogrammen wie z. B. EXCEL[1] sind hilfreich. Vertiefte statistische Kenntnisse sind zum Verständnis vieler Teile des Textes nicht zwingend notwendig, sollten aber zumindest bei späterer verantwortungsvoller Nutzung des Programms vorhanden sein. Das jeweils nötige Fachwissen ist entweder in Form eines *Hintergrundes* zusammengefasst, oder es wird auf einen Anhang verwiesen. Ferner werden an entsprechenden Stellen im Buch Hinweise auf einführende Texte wie etwa Basler (1994) oder Falk et al. (2002) gegeben, die zur weiteren Vertiefung herangezogen werden können. Gelegentlich ist eine solche Darstellung des statistischen Hintergrundes auch zur Erläuterung der Bezeichnungsweisen von STATISTICA und der Resultate der Analysen unumgänglich.

Generell ist das vorliegende Buch stark untergliedert, um die Übersichtlichkeit und Lesbarkeit zu erhöhen. Es werden eine Reihe von Umgebungen verwendet, wie der genannte *Hintergrund*, oder auch *Voraussetzungen* oder *Durchführung*. Diese sind nummeriert und durch Randsymbole gekennzeichnet. Damit sollte eine einfache Navigation möglich sein – der Leser kann gezielt die von ihm benötigten Informationen auf-

[1] An dieser Stelle eine Bemerkung: Immer wieder werden im Text Firmen- oder Produktnamen wie Microsoft, StatSoft, MySQL, etc. verwendet werden. Auch wenn nicht ein jedes einzelne Mal vermerkt, sei darauf hingewiesen, dass es sich dabei jeweils um eingetragene Markenzeichen handelt.

suchen. Die verwendeten Umgebungen und deren Symbole sind dabei die Folgenden:

Die *Durchführung*, zusätzlich noch grau unterlegt, beschreibt stets die konkreten Schritte, die in STATISTICA zu durchlaufen sind, um die gewünschte Analyse vollziehen zu können. Bezüglich der Menüführung wird dabei eine Schreibweise der Art *Datei* → *Öffnen* verwenden, welche auf den Punkt *Öffnen* im Menü *Datei* verweist.

Eine solche Durchführung wird gelegentlich durch kleine *Tipps* ergänzt, welche auf Besonderheiten im positiven oder negativen Sinne hinweisen.

Beispiele dienen der Veranschaulichung und Einübung der beschriebenen Verfahren; dabei werden zumeist reale oder realitätsnahe Datensätze verwendet. Auch die *Motivation* versucht durch Verweis auf reale Situationen den Nutzen von behandelten Methoden nahezubringen.

Der *Hintergrund* präsentiert, wie bereits erwähnt, statistisches Fachwissen in kompakter Form und ermöglicht somit ein vertieftes Verständnis der behandelten Themen.

Voraussetzungen an das zu untersuchende Datenmaterial stellen insbesondere Verfahren der induktiven Statistik. Nur wenn diese erfüllt sind, sind die gewonnenen Aussagen auch tatsächlich zuverlässig.

Gerade im Rahmen der explorativen Datenanalyse werden algorithmische Verfahren verwendet, deren prinzipieller Ablauf in Form eines *Algorithmus* zusammengefasst ist.

Häufig wird zudem eine der genannten Beschreibungen durch eine *Bemerkung* ergänzt.

Der Aufbau dieses Buches ist ausschließlich an inhaltlichen Aspekten orientiert, nicht an der Menüführung von STATISTICA. In Teil I werden wir uns vor allem mit der grundlegenden Bedienung von STATISTICA beschäftigen, wozu insbesondere auch Fragen der Datenhaltung und Datenverarbeitung gehören, siehe Kapitel 2 und 3. Anschließend werden in Teil II zahlreiche Verfahren der deskriptiven und explorativen Statistik vorgestellt. An dieser Stelle sei erwähnt, dass das partiell recht anspruchsvolle Kapitel 7, in welchem Verfahren wie Cluster-, Hauptkomponenten-, Diskriminanzanalyse, o. Ä., vorgestellt werden, nicht für das Verständnis der anschließenden Kapitel benötigt wird und deshalb auch einer späteren Lektüre vorbehalten werden kann.

Erst in Teil III zur induktiven Statistik kommt dann der Zufall ins Spiel. Hier werden eine Reihe wichtiger Testverfahren vorgestellt sowie Ansätze zur Datenmodellierung. In Teil IV werden schließlich einige Be-

sonderheiten von STATISTICA behandelt, welche dieses von anderen Programmpaketen abheben. Es sind dies in Kapitel 12 das Modul der statistischen Qualitätskontrolle, und in Kapitel 13 eine kleine Einführung in die Makroprogrammierung. Ergänzt wird der Text in Teil V durch eine Reihe von Anhängen, zu denen etwa Kompaktkurse zu MySQL und Visual Basic gehören, siehe die Anhänge B und C. Dadurch ist es zumindest theoretisch möglich, das vorliegende Buch ohne Begleitlektüre durchzuarbeiten.

Zu guter Letzt möchte ich noch einige Worte des Dankes aussprechen. Ein großer Dank geht an meinen Kollegen, Herrn Dr. René Michel, für die sorgfältige Durchsicht des Manuskriptes und die damit verbundenen wertvollen Änderungs- und Ergänzungsvorschläge. Nicht vergessen möchte ich Herrn Prof. Dr. Rainer Göb, der das STATISTICA-Projekt am hiesigen Statistiklehrstuhl begründete und mich mit STATISTICA überhaupt erst in Kontakt brachte, außerdem einzelne Abschnitte einer kritischen Durchsicht unterzog. Auch Herrn Fabian Müller von der Universität Freiburg gilt mein Dank für konstruktive Anregungen zu einer früheren Version des Manuskriptes. Nicht versäumen möchte ich es, mich bei den Herren Bernd-Uwe Loll, Thilo Eichenberg und Michael Busch von der Firma StatSoft zu bedanken für die Unterstützung während des Buchprojektes, zahlreiche wichtige Informationen und wertvolle Tipps zu STATISTICA, insbesondere bei Herrn Thilo Eichenberg auch für das Durchlesen des Manuskriptes und die daraus resultierenden Anregungen und Korrekturen. Ebenso gilt mein Dank dem R. Oldenbourg Verlag, vor allem meiner Lektorin, Frau Margit Roth, aber auch Herrn Dr. Rolf Jäger, Herstellungsleitung, für den Einsatz für dieses Buchprojekt und die sehr gute Zusammenarbeit. Und natürlich möchte ich meiner Frau Miia danken, für sprachliche Anregungen sowie vor allem für ihr Verständnis dafür, dass ich in den letzten Monaten meine Freizeit größtenteils der Arbeit am Computer widmen musste.

Würzburg, im Sommer 2006 Christian H. Weiß

Christian H. Weiß Universität Würzburg
Institut für Mathematik, Lehrstuhl für Statistik
Am Hubland
97074 Würzburg

christian.weiss@mathematik.uni-wuerzburg.de

Inhaltsverzeichnis

Teil I

Einführung in STATISTICA

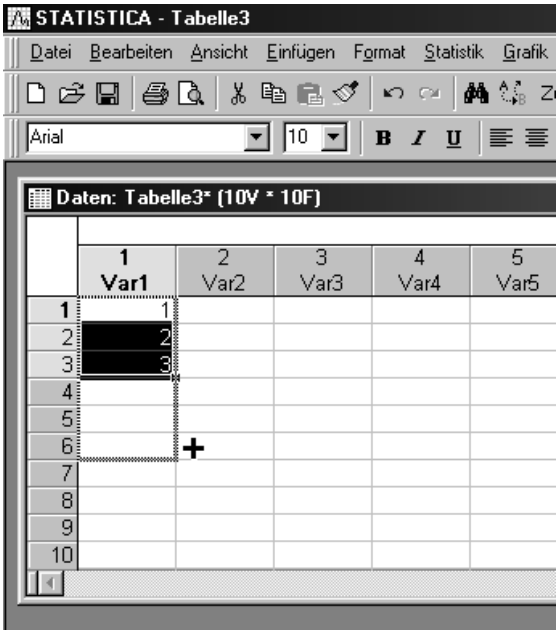


Abb. 1.1: *Der Startbildschirm von* STATISTICA.

1 Erste Schritte in STATISTICA

Das 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 durch EXCEL oder ähnliche Tabellenkalkulationsprogramme vertraut erscheinen dürfte.

Wie in Abbildung 1.1 angedeutet, 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.

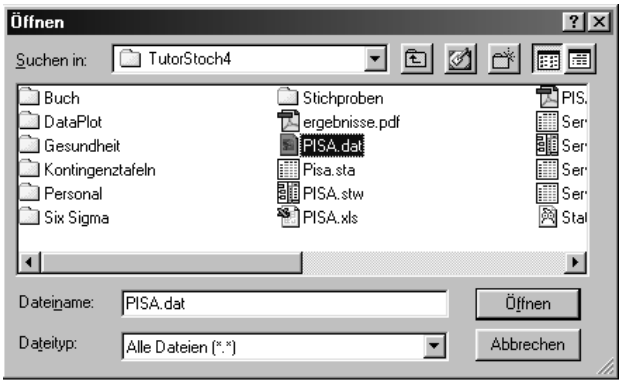


Abb. 1.2: *Öffnen der Datendatei.*

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, 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 jeweils erzielten Werte enthalten.

Nehmen wir dieses Beispiel, um uns ein wenig mit STATISTICA vertraut zu machen:

Die konkreten Daten befinden sich in der (Text-)Datei `PISA.dat` und müssen erst in ein STATISTICA-Tabellenblatt importiert werden.[1] Dazu wählt man wie gewohnt den Menüpunkt *Datei* → *Öffnen*, worauf ein Dialogfenster wie in Abbildung 1.2 erscheint. Nachdem man bei Dateityp *Alle Dateien* eingestellt hat, wird die gesuchte Datei angezeigt und kann ausgewählt werden.

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

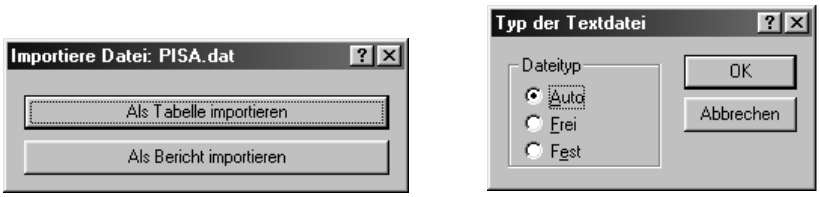


Abb. 1.3: *Die Datendatei als Tabelle importieren und Dateityp* Auto *wählen.*

Im nächsten Dialog, vergleiche Abbildung 1.3, wählt man *Als Tabelle importieren* und belässt es anschließend beim Dateityp *Auto*.

Nun öffnet sich das Fenster *Textdatei öffnen* aus Abbildung 1.4. Hier müssen einige Einstellungen gemacht werden. Ein Blick in die Datei `PISA.dat` zeigt, dass sich über dem eigentlichen Datensatz ein paar Zeilen mit einer kleinen Beschreibung befinden. Da der beschreibende Text nicht im Tabellenblatt erscheinen soll, beginnen wir den Import bei Zeile 4, welche die Bezeichnungen der späteren Variablen enthält. Wir machen einen Haken bei *Variablennamen aus erster Zeile*. Ebenso verfahren wir bei *Fallnamen aus erster Spalte*, damit die Fälle die Namen der einzelnen Bundesländer tragen. Schließlich darf bei *Feldtrennung(en)* lediglich *Tabulator* markiert sein. Eine Auswahl von *Komma* würde dagegen dazu führen, dass die Kommazahlen in mehrere Spalten zerlegt würden.

Falls der Computer auf deutsche Sprache eingestellt ist, verwendet STATISTICA die deutsche Schreibweise, insbesondere also das Komma als Dezimaltrennzeichen. In der Praxis hat man jedoch häufig Datensätze aus dem angloamerikanischen Raum vorliegen, die Punkte verwenden. Damit solche Daten importiert werden können, müssen bei den Versionen bis inkl. 6.1 vorher in einem Editor die Punkte durch Kommata ersetzt werden. Alternativ, aber wohl problematischer, kann in der Systemsteuerung unter Ländereinstellungen die Sprache geändert werden. Seit Version 7 werden auch Punkte als Dezimaltrennzeichen erkannt.

Neu ab Version 7 sind leicht modifizierte Dialoge zum Import von Dateien, verglichen mit denen aus den Abbildungen 1.3 und 1.4. Nützlich sind dabei insbesondere einige weitere Optionen im Dialog *Textdatei öffnen* aus Abbildung 1.4, wie etwa *Voranst. Leerzeichen abschneiden* oder *Leere Zeilen ignorieren*. Ferner erlaubt ein Häkchen bei *Ansicht/Bearbeiten von Spaltentypen vor Import* eine Vorschau auf die von STATISTICA erstellten Variablen. •

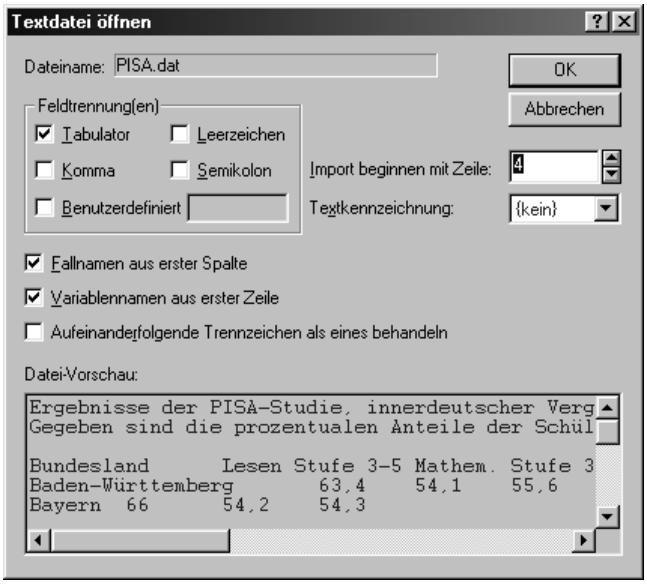


Abb. 1.4: *Unter* Textdatei öffnen *müssen eine Reihe wichtiger Einstellungen gemacht werden.*

Nun ist es geschafft, und durch Klick in das weiße Feld links oben in der Tabelle kann diese markiert werden. Durch erneuten Klick mit der *rechten* Maustaste ins Tabellenblatt öffnet sich ein PopUp-Menü. Dort kann unter *Format* der Befehl *Optimale Höhe/Breite* betätigt werden, damit die Daten und Randbeschriftungen tatsächlich gut lesbar sind, vergleiche hierzu Abbildung 1.5.

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, konkret: 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 Werte der Variablen *Lesen* grafisch veranschaulichen, und zwar mit Hilfe eines Balkendiagramms. Dazu wählen wir wie in Abbildung 1.6 den Menüpunkt *Grafik* → *2D-Grafiken* → *Balkenplots*. Anschließend stellen wir wie in Abbildung 1.7 *Lesen Stufe 3-5* als Variable und horizontale Ausrichtung ein, und betätigen dann den *OK*-Knopf. Nun wird eine *Arbeitsmappe* (*Workbook*) erstellt und die gewünschte Grafik in Selbige eingefügt. Das Ergebnis ist in Abbildung 1.8 zu sehen.

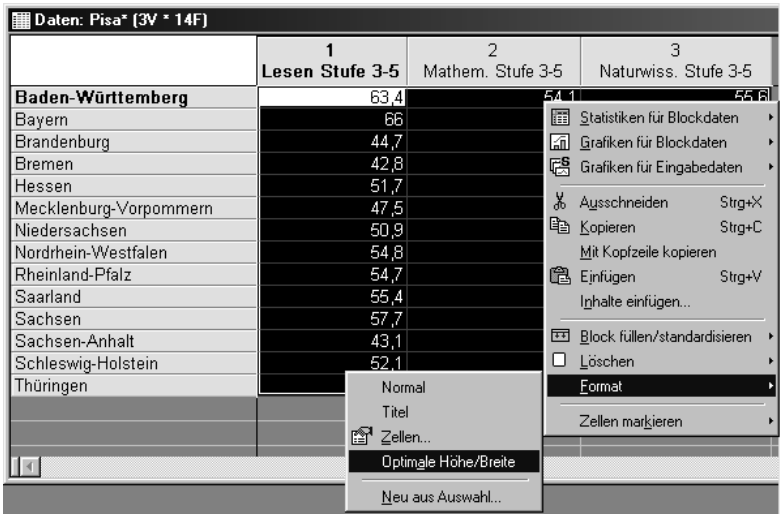


Abb. 1.5: *Die gewünschten Daten wurden nun als Tabellenblatt importiert.*

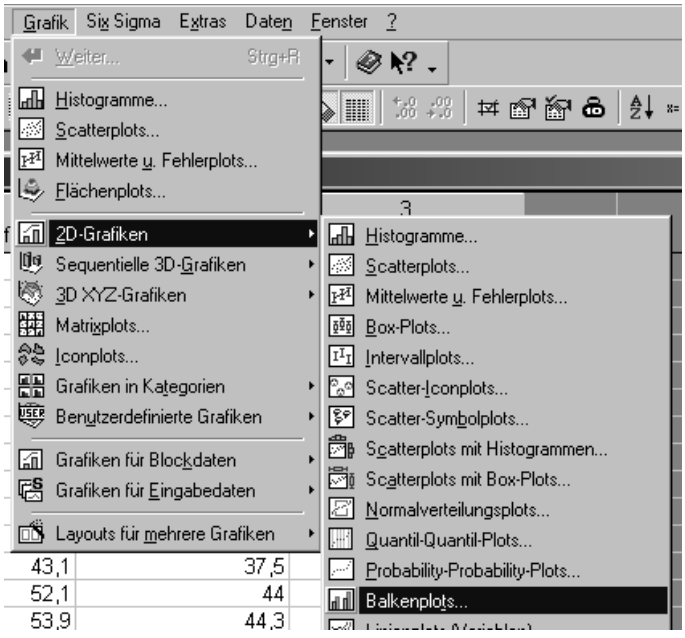


Abb. 1.6: *Wir erstellen ein Balkendiagramm der Daten.*

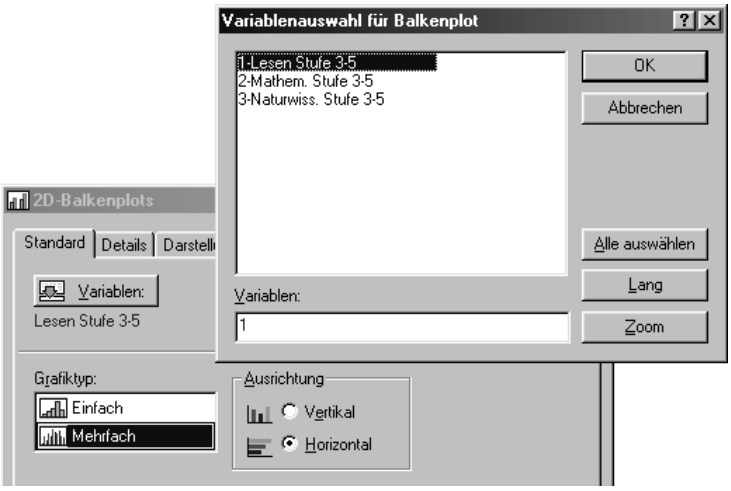


Abb. 1.7: *Als Variable wählen wir* Lesen Stufe 3-5.

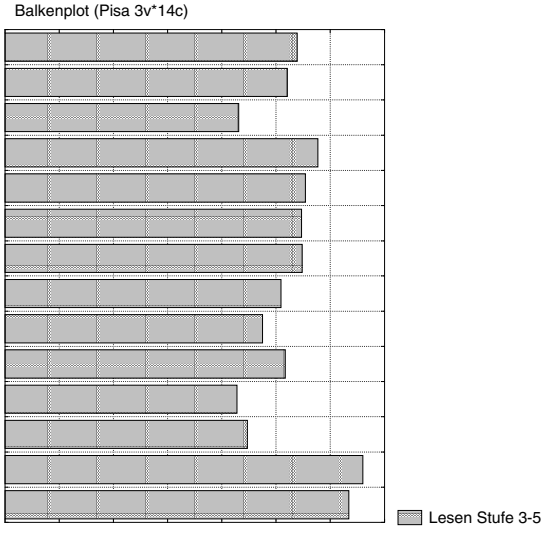


Abb. 1.9: *Aktives Grafikmenü – Reaktivierung per Knopfdruck.*

Links unten im STATISTICA-Fenster befindet sich nun ein Knopf, mit dem das Balkendiagramm-Menü mit den getroffenen Einstellungen wieder aktiviert werden kann, siehe Abbildung 1.9. Nach erneuter Aktivierung können weitere Diagramme, etwa alle drei Disziplinen in einem mehrfachen Balkendiagramm, erstellt werden können. Dies sei dem Leser zur Übung empfohlen. Parallel dazu können auch beliebige weitere Analysen gestartet werden, auch hier sei an die Experimentierfreude des Lesers appelliert.

Bis zu dem Zeitpunkt, an dem wir die erstellte 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 2.2 auf Seite 15. 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 zwei Möglichkeiten: Entweder man analysiert wie gewohnt, öffnet am Ende die gewünschte Arbeitsmappe und zieht mit gedrückter Maustaste die Resultate in die Zielarbeitsmappe. Oder 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 Zielarbeitsmappe. Nach Bestätigung 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.

Neu ab Version 7 ist die Möglichkeit, sog. *Variablenbündel* zu definieren, und zwar immer vom Variablendialog einer Analyse aus. Bei aktuellen STATISTICA-Versionen verfügt dieser nämlich, im Gegensatz zu Abbildung 1.7, 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 1.7. 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. •

2 Datenhaltung in STATISTICA

In diesem Abschnitt wollen wir uns mit Fragen auseinandersetzen, die in irgendeiner Weise mit der Datenhaltung in Zusammenhang stehen. Dazu zählt ein erster Überblick über die verschiedenen Dateitypen in STATISTICA in Abschnitt 2.1, sowie die Frage des Imports und Exports von Daten. Gerade Letzteres ist von großer praktischer Bedeutung, da einerseits Rohdaten gewöhnlich in Form von Textdateien oder gar Datenbanken vorliegen, andererseits Resultate von Analysen natürlich auch Benutzern zur Verfügung gestellt werden müssen, die nicht auf STATISTICA Zugriff haben. Deshalb werden wir uns in den Abschnitten 2.2 und 2.4 bis 2.6 mit dem Import von Daten verschiedenster Formate auseinandersetzen, und in Abschnitt 2.3 Exportmöglichkeiten besprechen.

Beginnen wir jedoch zuerst mit einem kurzen Überblick über die wichtigsten Dateitypen von STATISTICA.

2.1 Die unterschiedlichen Dateitypen in STATISTICA

Neben den Datentabellen, welche bei STATISTICA die Dateiendung `.sta` besitzen, sind folgende Dateitypen von Bedeutung:

- `.stw`: In der *Arbeitsmappe* (*Workbook*) werden alle Ergebnisse abgelegt, die im Laufe der Datenanalyse anfallen. Bei Bedarf können einzelne Teilergebnisse extrahiert und in verschiedenen anderen Dateiformaten abgespeichert werden, siehe hierzu auch Durchführung 2.3.1. Generell ist eine Arbeitsmappe eine Art von Behälter für ActiveX-Dokumente.
- `.stg`: In Durchführung 2.3.2 werden wir auf die *Grafikdateien* von STATISTICA zu sprechen kommen.
- `.smx`: Hierbei handelt es sich um *Matrixdateien*, wie sie etwa im Rahmen der Clusteranalyse (vgl. Abschnitt 7.1) vorkommen werden, wenn man dort eine Abstandsmatrix erstellen lässt.
- `.str`: Alternativ kann man Datentabellen und jegliche Art von Analyseergebnis auch zusätzlich in *Berichtsdateien* abspeichern

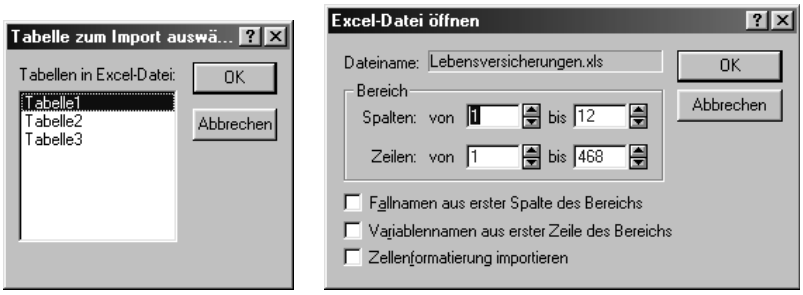


Abb. 2.1: *Auswahl und Einschränkung der zu importierenden Tabelle.*

bzw. einzelne Teilresultate aus einer Arbeitsmappe via Kopieren und Einfügen in einen Bericht übertragen. Ein Bericht ist im Prinzip ein gewöhnliches Textdokument, welches man auch als `rtf`-Datei speichern kann. Nur ist es hier möglich, im Gegensatz zur Arbeitsmappe, noch weitere Bemerkungen oder externe Daten anzubringen. Eine genauere Beschreibung ist in Abschnitt 3.4 zu finden.

- `.svb`: Hierbei handelt es sich um *Makrodateien*. STATISTICA erlaubt die Aufzeichnung und Programmierung von Makrodateien, wobei hier ähnlich wie bei vielen anderen Produkten ein Visual-Basic-Dialekt verwendet wird. Dieses Thema werden wir in Kapitel 13 vertiefen. Verwandte Dateitypen sind `.svc` (Klassenmodule), `.svo` (Objektmodule) und `.svx` (Codemodule).

2.2 Import von Daten

Nur selten wird man in der Natur auf Daten treffen, die bereits im `.sta`-Format abgelegt sind. Deshalb ist der Import von Daten in der Praxis sehr wichtig. Neben dem Import von Daten aus Datenbanken, auf welchen wir in den Abschnitten 2.4 bis 2.6 eingehen werden, wird man vor allem häufig auf zwei Fälle treffen: Daten, die im Textformat abgespeichert sind, die zugehörigen Dateien haben dann üblicherweise die Endung `.txt`, `.dat` oder gar keine Endung. Oder Daten, die in einer EXCEL-Datei mit Endung `.xls` vorliegen.

Der Import von Daten aus Textdateien wurde bereits in Kapitel 1 erläutert und soll deshalb hier nicht erneut vertieft werden. Betrachten wir hier also den Fall, bei dem die Daten in einer EXCEL-Datei vorliegen.

Durchführung 2.2.1

Um EXCEL-Dateien zu importieren, wählt man wie in Kapitel 1 den Menüpunkt *Datei* → *Öffnen* und dann im sich öffnenden Dialog den *Dateityp: Datendateien* oder direkt *Excel-Dateien*. Nachdem man die gewünschte Datei gewählt und *Öffnen* gedrückt hat, wird man seit Version 7 gefragt, ob man *Alle Tabellen in eine Arbeitsmappe importieren* oder eine *Ausgewählte Tabelle in eine Tabelle importieren* möchte; wir wählen Letzteres.

Anschließend erscheint ein Fenster wie im linken Teil von Abbildung 2.1. Dort wählt man das gewünschte Tabellenblatt aus, bestätigt, und trifft dann auf einen Dialog wie im rechten Teil von Abbildung 2.1. Hier ist der zu importierende Ausschnitt der Tabelle zu bestimmen, wobei der angegebene Vorschlag erst einmal den gesamten nichtleeren Bereich der Tabelle umfasst. Ähnlich wie bei Textdateien, vgl. Kapitel 1, kann man auch hier auf Wunsch Fall- und/oder Variablennamen übernehmen, und diesmal sogar Formatierungen. Durch Betätigung des *OK*-Knopfes wird der Import abgeschlossen.

•

Andere von STATISTICA unterstützte Dateitypen sind etwa dBase-, SPSS-, Lotus- oder Quattro-Pro-Dateien, seit Version 7 auch SAS-, Minitab- oder JMP-Dateien.

Ferner können in Arbeitsmappen auch andere ActiveX-Dokumente wie WORD- oder EXCEL-Dateien abgelegt werden. Dazu klickt man im Verzeichnisbaum an die gewünschte Stelle mit der rechten Maustaste und wählt im sich öffnenden PopUp-Menü *Einfügen ...*, dann *ActiveX-Dokumentobjekt* und klickt *OK*. Anschließend markiert man *Aus Datei erstellen* und sucht die betreffende Datei aus.

2.3 Export von Daten

Der Export von Daten verläuft im Prinzip auf umgekehrten Wege wie im vorigen Abschnitt 2.2 beschrieben. Um eine `.sta`-Tabelle in eines der bereits oben genannten Datenformate zu übertragen, wählt man *Datei* → *Speichern unter* und im sich öffnenden Dialog den gewünschten Dateityp.

Unter Umständen benötigt man diese Exportmöglichkeit aber auch, wenn man mit einer älteren Version von STATISTICA arbeiten will. Um z. B. eine Datei, welche mit STATISTICA 6.1 oder höher erstellt wurde, mit STATISTICA 6.0 öffnen zu können, muss diese im STATISTICA 6.0-Format abgespeichert werden. Und für Version 7 gilt, dass nicht nur

erneut die .sta-Dateien, sondern diesmal auch die .stw- und .stg-Dateien modifiziert wurden.

Die eben beschriebenen Exportmöglichkeiten gelten in gleicher Weise auch für .smx-Dateien, jedoch gibt es Unterschiede bei den übrigen Dateitypen. Bei den Berichtsdateien, Endung .str, handelt es sich ja eigentlich um höhere Textdateien, entsprechend sind hier ein Export nach .rtf oder .html möglich, seit Version 7 ist zudem auch nach XML oder PDF. Mehr dazu in Abschnitt 3.4.

Neu ab Version 7 ist die Möglichkeit, den Inhalt eines Arbeitsmappenordners auf einmal anzeigen zu lassen; dazu ist lediglich der entsprechende Ordner zu markieren. Vorsicht: Bei gut gefüllten Ordnern kann dieser Vorgang erheblich Zeit beanspruchen, so dass man unter Umständen den Fokus besser auf ein anderes Objekt setzen sollte. •

Bei .stw-Dateien ist ein direkter Export erst einmal gar nicht möglich, da es sich hierbei ja eigentlich um ein Sammelsurium verschiedenster Analyseresultate handelt. Die einzelnen Teile kann man aber wie folgt exportieren:

Durchführung 2.3.1

Im Wesentlichen befinden sich in einer Arbeitsmappe eigentlich nur zwei Arten von Analyseresultaten: Entweder Tabellen oder Grafiken. Betrachten wir die Arbeitsmappe aus Abbildung 2.2, welche sowohl eine Grafik als auch eine Tabelle enthält. Klickt man mit der rechten Maustaste auf das Tabellensymbol in der linken Hälfte der Arbeitsmappe, so öffnet sich das PopUp-Menü der Abbildung. Um die Tabelle zu extrahieren (und die Originalarbeitsmappe unverändert zu lassen), wählt man *Als Stand-alone Fenster extrahieren → Kopie*. Anschließend erscheint die Tabelle in einem eigenen Fenster, und STATISTICA erlaubt es, diese Tabelle als gewöhnliche .sta-Datei abzuspeichern oder in einem der anderen oben erwähnten Formate.

Um eine Grafik zu extrahieren, klickt man mit der rechten Maustaste direkt in die Grafik hinein, so dass sich ein PopUp-Menü wie in Abbildung 2.3 öffnet. Nun kann man entweder *Identische Grafik erzeugen* wählen, dann wird analog eben eine freistehende Kopie der Grafik erzeugt, oder man wählt direkt *Grafik speichern*.

Welchen Weg man immer auch gegangen ist, letztlich stehen einem die Exportmöglichkeiten zur Verfügung, wie sie für die Grafikdateien von STATISTICA, Endung .stg, gelten. Diese sind in Durchführung 2.3.2 beschrieben. •

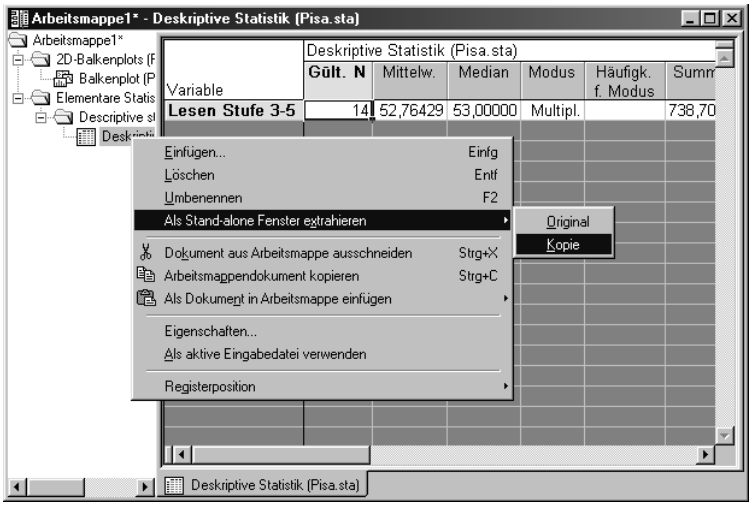


Abb. 2.2: *Extraktion einer Tabelle aus einer Arbeitsmappe.*

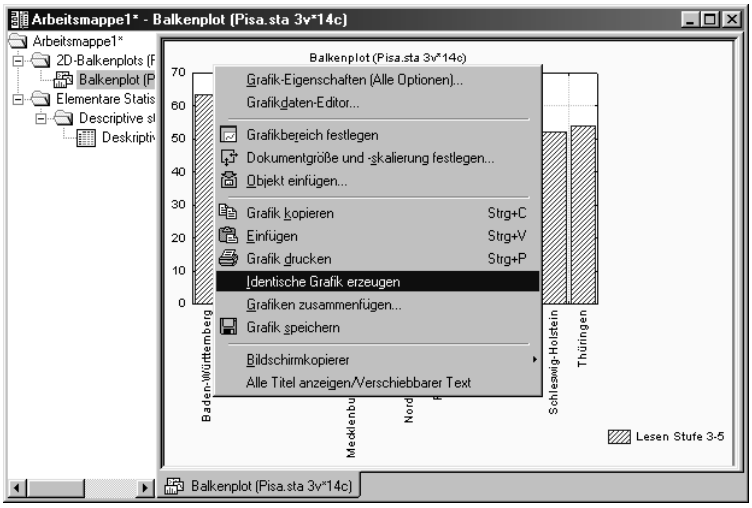
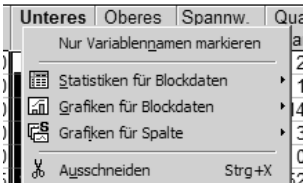


Abb. 2.3: *Extraktion einer Grafik aus einer Arbeitsmappe.*

(a) 

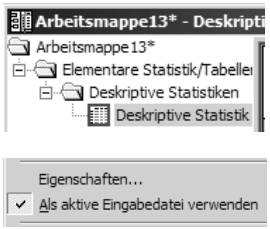
(b) 

Abb. 2.4: *Tabellen in Arbeitsmappen.*

Seit Version 7 kann man Grafiken und Tabellen aus Arbeitsmappen übrigens auch gleich über *Speichern unter* exportieren.

Sollen Grafiken oder Tabellen in einem Textverarbeitungsprogramm wie WORD o. Ä. direkt verwendet werden, kann man sich den Umweg über Speichern und später dann Grafik/Tabelle einfügen auch sparen. Man wählt in den in Durchführung 2.3.1 besprochenen PopUp-Menüs einfach den Punkt *Kopieren* und dann im Textverarbeitungsprogramm entsprechend *Einfügen*.

Liegt in einer Arbeitsmappe eine Tabelle vor, die weiter analysiert werden soll, stehen dazu im Prinzip zwei Möglichkeiten zur Verfügung:

- Man markiert die interessierenden Variablen, klickt diese mit der rechten Maustaste an, und wählt Analysewerkzeuge aus den Teilmenüs *Statistiken für Blockdaten*, *Grafiken für Blockdaten* oder *Grafiken für Spalte*, die im sich öffnenden PopUp-Menü zur Verfügung stehen, siehe Abbildung 2.4 (a). Allerdings ist die Auswahl an Analysen begrenzt.
- Oder man setzt den Fokus auf die Tabelle der Arbeitsmappe, indem diese im Navigationsfeld mit der rechten Maustaste angeklickt und im sich öffnenden PopUp-Menü *Als aktive Eingabedatei verwenden* gewählt wird. Dann erscheint die betreffende Tabelle im Navigationsfeld rot umrandet, siehe Abbildung 2.4 (b) oben. Nun kann man wie gewohnt Analysen ausführen. Allerdings werden nun tatsächlich alle Analysen mit dieser Tabelle durchgeführt. Will man wieder eine `.sta`-Datei analysieren, muss man die rot umrandete Tabelle erneut mit der rechten Maustaste anklicken und das Häkchen vor *Als aktive Eingabedatei verwenden* entfernen, siehe Abbildung 2.4 (b) unten.

Obwohl eine aus einer Arbeitsmappe extrahierte Tabelle, siehe Durchführung 2.3.1, die Endung `.sta` trägt, ist sie leider zuerst einmal keine gewöhnliche Datentabelle. Zwar ist die Eingabe von Formeln im Tabellenblatt bereits innerhalb der Arbeitsmappe möglich, siehe Abschnitt 3.2, doch die üblichen Analysen des Hauptmenüs können auf eine extrahierte Tabelle zunächst nicht direkt angewendet werden. Dies kann man jedoch einfach beheben, indem man, nachdem die Tabelle als aktives Fenster gewählt wurde, den Punkt *Daten* → *Eingabetabelle* wählt.

Einzelne Grafikdateien kann man in STATISTICA nicht nur als `.stg`-Datei abspeichern, es stehen auch hier eine Reihe von Exportmöglichkeiten zur Verfügung:

Durchführung 2.3.2

Will man eine einzelne Grafik speichern, stehen das hauseigene `.stg`-Grafikformat sowie `.bmp`, `.jpg`, `.png` oder auch `.wmf` zur Verfügung. Bei Version 6.0 verläuft der Export nach `.wmf` gelegentlich nicht zufriedenstellend. Seit Version 7 werden die Exportmöglichkeiten noch ergänzt durch `.pdf`, `.emf`, `.gif` und `.tif`, so dass nun kaum noch Wünsche unerfüllt bleiben dürften. •

Der Vorteil des `.stg`-Formats ist es, dass man die Grafik auch zu einem späteren Zeitpunkt in gewohnter Weise, siehe Kapitel 4, manipulieren kann. Der Nachteil ist natürlich, dass wohl kaum ein anderes Programm solche Dateien lesen kann.

Zumindest Benutzer von LaTeX sind vorwiegend an `.eps`-Grafiken interessiert. Um diese, sowohl von Tabellen als auch Grafiken einer Arbeitsmappe, in sehr guter Qualität zu erzeugen, empfiehlt sich die Installation eines geeigneten Druckertreibers (z. B. HP LaserJet 5/5M PostScript) der Windows-CD an den Anschluss FILE. Ist dies geschehen, markiert man in der linken Hälfte der Arbeitsmappe das gewünschte Objekt und wählt *Datei* → *Drucken*. Als Drucker wählt man den installierten PS-Drucker, welcher die Druckausgabe automatisch in eine `.eps`-Datei umleitet.

Neu ab Version 7 Nicht nur Berichte, auch Tabellen, Grafiken und Teile einer Arbeitsmappe allgemein, können nun als `.pdf` gespeichert werden. Diese Möglichkeit wird entweder im Menü *Datei* → *Speichern als PDF*, über den Knopf der Symbolleiste, oder über ein sich öffnendes PopUp-Menü angeboten. •

Zu erwähnen ist noch ein Merkmal, dass vor allem für Autoren von STATISTICA-Büchern sehr hilfreich sein dürfte. Unter Windows kann man einen Bildschirmausdruck gewöhnlich entweder nur mit Hilfe der *Druck*-Taste, dann wird der gesamte Bildschirm in die Zwischenablage kopiert, oder mit der Kombination *Alt+Druck* erstellen, dann wird nur die aktive Anwendung in die Zwischenablage kopiert. Bei STATISTICA dagegen kann man sich des Werkzeuges *Bildschirmkopierer* bedienen, das über das Menü *Bearbeiten* sowie über manche PopUp-Menüs erreichbar ist.

Dabei gibt es zwei Wahlmöglichkeiten: Klickt man auf *Rechteck kopieren*, so wandelt sich der Mauszeiger in ein Kreuz um, und man kann bei gedrückter linker Maustaste einen beliebigen rechteckigen Bereich des Bildschirms auswählen; nach dem Loslassen der Maustaste wird dieser automatisch in die Zwischenablage kopiert. Bei der Option *Fenster kopieren* dagegen ist immer eines der geöffneten Fenster komplett markiert. Nach Klick mit der linken Maustaste in das gewünschte Fenster wird dieses vollständig in die Zwischenablage kopiert.

Den Inhalt der Zwischenablage kann man dann mit Hilfe eines Bildbearbeitungsprogrammes wie beispielsweise PAINT im gewünschten Grafikformat speichern.

2.4 Anbindung an Datenbanken via OLE DB

STATISTICA erlaubt auch den direkten Zugriff auf Datenbanken; man ist dabei auf das Menü *Datei* → *Externe Daten* angewiesen. Prinzipiell benötigt man dazu entweder eine geeignete OLE DB[1]- oder ODBC[2]-Schnittstelle. In diesem Abschnitt wollen wir den Zugang via OLE DB beispielhaft demonstrieren, im anschließenden Abschnitt 2.5 den über ODBC. Schließlich soll in Abschnitt 2.6 der Sonderfall von ACCESS-Datenbanken angeschnitten werden, die ja im Hausgebrauch recht beliebt sind.

Zuerst wollen wir erörtern, wie man mit Hilfe von OLE DB auf Datenbanken zugreifen kann. Das notwendige Vorgehen soll beispielhaft am kostenlosen Datenbanksystem MySQL demonstriert werden. Bei Datenbanken anderer Datenbankhersteller ist ein analoges Vorgehen möglich. Informationen zu MySQL findet der interessierte Leser im Anhang B,

[1] ***O**bject **L**inking and **E**mbedding for **Data**bases*, eine Sammlung von Schnittstellen zum Zugriff auf Datenbanken, basierend auf dem *Component Object Model (COM)*.

[2] ***O**pen **D**ata**B**ase **C**onnectivity* ist eine standardisierte Schnittstelle zum Zugriff auf Datenbanken.

und ferner beispielweise auch bei Hinz (2002). Informationen zur Sprache SQL an sich gibt es etwa bei Throll & Bartosch (2004).

Um Daten von STATISTICA aus über OLE DB aus einer MySQL-Datenbank abfragen zu können, muss man zumindest MySQL selbst sowie MyOLEDB installieren. Zu finden sind diese auf der MySQL-Homepage `http://www.mysql.com`, siehe auch Anhang B.

Gehen wir im Folgenden vom Beispiel eines Handelsunternehmens aus. In der Datenbank `verkauf`[3] befindet sich u. a. die Tabelle `deckung`, welche zu jedem durchgeführten Auftrag den zugehörigen Deckungsbeitrag enthält. Ferner sind diese Daten um die zuständige Abteilung und den zuständigen Mitarbeiter ergänzt.

Nun wollen wir die ermittelten Deckungsbeiträge einer tiefergehenden Analyse unterziehen. Dazu werden wir die Daten nach STATISTICA importieren, so dass wir alle dort verfügbaren statistischen und grafischen Werkzeuge einsetzen können.

Hierzu wählen wir das Menü *Datei → Externe Daten → Neue Abfrage erstellen*. Es öffnet sich ein Dialog wie in Abbildung 2.5. Dort könnte man bereits fertige Datenbankverknüpfungen auswählen, in unserem Fall gilt es jedoch, zuerst einmal eine passende Datenbankverknüpfung zu erstellen. Deshalb wählen wir *Neu*, und es öffnet sich ein Menü wie in Abbildung 2.6.

Leider kennt STATISTICA weniger Datentypen als MySQL. Beispielsweise können `BIGINT`-Variablen nicht importiert werden, entsprechende Spalten wären nur mit Nullen gefüllt. Falls ohne Datenverlust möglich, sollte deshalb der Datentyp in MySQL verändert werden, z. B. zu `INT`.

Auf der Karte *Provider* wählen wir den *MySQL.OLEDB Provider* und klicken *Weiter ≫*. Auf der anschließenden Karte *Verbindung* tragen wir wie in Abbildung 2.7 bei *Datenquelle* die gewünschte Datenbank ein, in unserem Fall also `verkauf`. Bei *Speicherort* geben wir entweder *localhost* ein, so wahr sich die Datenbank auf dem lokalen Rechner befindet, oder ansonsten die IP-Adresse des entsprechenden Rechners. Wichtig ist, dass der MySQL-Server im Hintergrund aktiv ist, siehe Anhang B.

Nach dem Bestätigen mit OK wird im nun folgenden Dialog aus Abbildung 2.8 nach einem Namen für die Datenbankverknüpfung gefragt,

[3] Datenbanken werden bei MySQL als eigener Unterordner im Verzeichnis `C:\Programme\...\MySQL\data` abgelegt. In der Datei `verkauf.zip` befindet sich der gepackte Ordner `verkauf`, welcher in dieses Verzeichnis zu entpacken ist. Alternativ kann man das Beispiel in Anhang B durchlaufen, dabei wird die genannte Datenbank erzeugt.

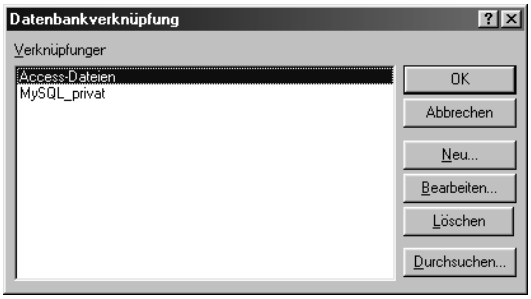


Abb. 2.5: *Erstellen einer neuen Datenbankverknüpfung.*

wir geben z. B. *MySQL_ verkauf* ein. Nun gelangen wir wieder in das allererste Menü, vergleiche die Abbildungen 2.5 und 2.9. Unterschied ist, dass jetzt auch die neue Verknüpfung angeboten wird. Diese wählen wir aus und gelangen in das Fenster aus Abbildung 2.10. Nachdem nun die Verknüpfung zur Datenbank definiert wurde, werden wir bei zukünftigen Zugriffen immer an dieser Stelle einsteigen.

Links in Abbildung 2.10 sehen wir alle Tabellen und Variablen der Datenbank `verkauf`, und es besteht die Möglichkeit, per Maus eine Abfrage zusammenzustellen. Wer seine Abfragen lieber selbst mit SQL formulieren möchte, kann den grafischen Modus verlassen, indem er das Häkchen vor dem Menü *Ansicht* → *Grafischer Modus* entfernt, vgl. Abbildung 2.10, oder den Knopf löst. Man wechselt dadurch in den Textmodus, wie er in Abbildung 2.11 zu sehen ist, und kann dort gewöhnliche SQL-Abfragen formulieren, etwa

```
SELECT * FROM deckung;
```

Anschließend drücken wir den Abspielknopf ganz rechts in der Symbolleiste, vgl. Abbildung 2.11 oben, und geben der Abfrage im folgenden Menü, siehe Abbildung 2.12, einen Namen. Ein Häkchen bei *Angepasst an Datentabelle* bewirkt, dass die Tabelle genau so viele Felder enthält, wie auch importiert werden. Schließlich erhalten wir die Daten in Form einer `.sta`-Tabelle, wie sie in Abbildung 2.13 unten zu sehen ist. Ferner erkennen wir in der Abbildung 2.13 auch, dass wir im Hintergrund via MySQL die Daten unserer Datenbank weiterbearbeiten und durch Auswahl des Menüpunktes *Daten aktualisieren* die Änderungen automatisch in STATISTICA übernehmen können.

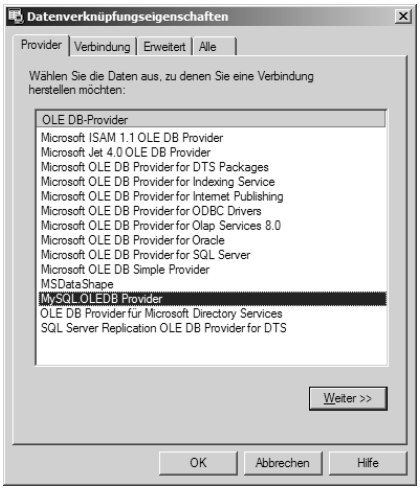


Abb. 2.6: *Auswahl des MySQL.OLEDB-Treibers.*

Abb. 2.7: *Auswahl von Datenbank und 'localhost'.*

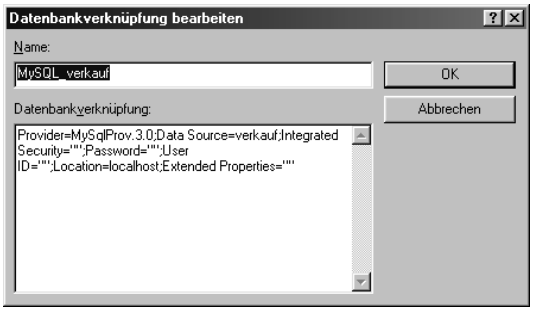


Abb. 2.8: *Benennung der fertigen Datenbankverknüpfung.*

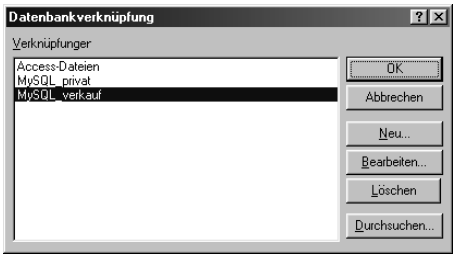


Abb. 2.9: *Ausführen der Datenbankverknüpfung.*

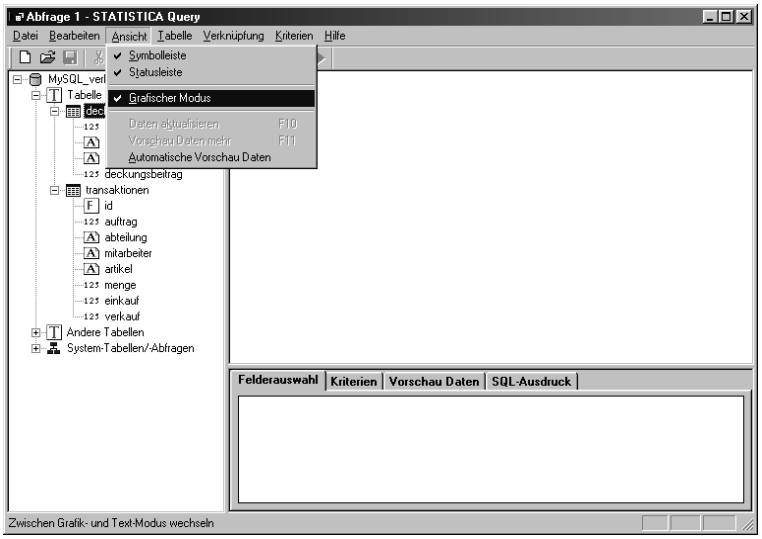


Abb. 2.10: *Abfragen formulieren im grafischen Modus.*

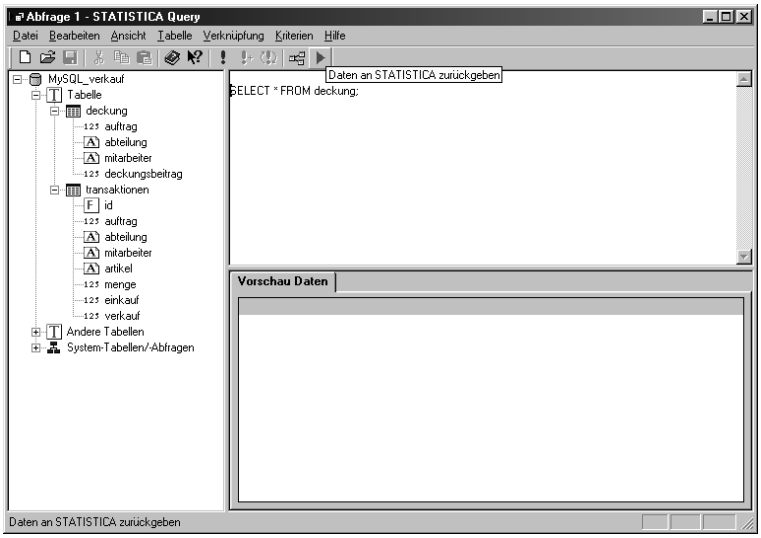


Abb. 2.11: *Abfragen formulieren im Textmodus.*

Abb. 2.12: *Abfrage abspeichern.*

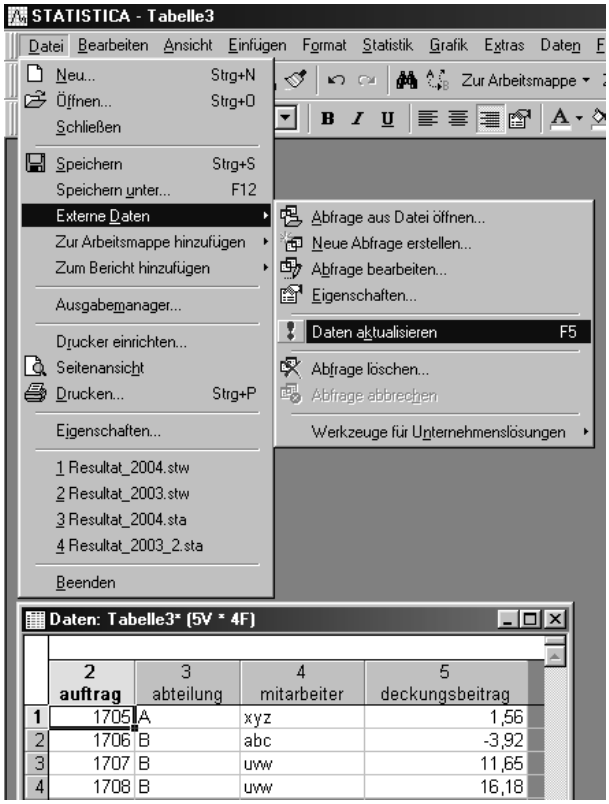


Abb. 2.13: *Die von MySQL importierten Daten.*

2.5 Anbindung an Datenbanken via ODBC

Eine zweite Möglichkeit, an eine Datenbank anzudocken, besteht darin, zuerst einmal eine ODBC-Schnittstelle zu dieser Datenbank einzurichten, und dann über diese Schnittstelle auf die Datenbank zuzugreifen. In diesem Abschnitt soll erläutert werden, wie man eine solche Schnittstelle allgemein einrichtet, unabhängig davon, ob überhaupt auf dem Rechner STATISTICA installiert ist. Anschließend wollen wir dann mittels STATISTICA über diese Schnittstelle an die Datenbank andocken. Jedoch sei schon hier erwähnt, dass die Erstellung einer solchen Schnittstelle auch direkt von STATISTICA aus möglich wäre, worauf wir im anschließenden Abschnitt 2.6 eingehen werden.

Als Beispiel verwenden wir wieder die MySQL-Datenbank `verkauf` aus dem vorigen Abschnitt 2.4. Damit das nun beschriebene Verfahren funktioniert, muss zusätzlich zu MySQL auch *MyODBC* installiert sein, vgl. Anhang B.

Um für diese Datenbank eine ODBC-Schnittstelle einzurichten, wechseln wir ins *Start*-Menü von Windows, und dort zu *Einstellungen* → *Systemsteuerung* → *Verwaltung* → *Datenquellen (ODBC)*, siehe Abbildung 2.14. Es öffnet sich der Dialog aus Abbildung 2.15, welcher bereits eine Reihe von fertigen ODBC-Schnittstellen anbietet.

Zur Erstellung einer neuen Schnittstelle betätigen wir den Knopf *Hinzufügen* und gelangen zum Dialog der Abbildung 2.16. Hier wählen wir den installierten MyODBC-Treiber aus, vgl. Anhang B, und klicken auf *Fertig stellen*. Im anschließenden Dialog, siehe Abbildung 2.17, müssen wir im Feld *Data Source Name* einen Namen für die Verknüpfung eintragen. Unter diesem Namen wird die Verknüpfung dann als `.dsn`-Datei gespeichert. Bei *Server* trägt man *localhost* oder die IP-Adresse des Rechners ein, auf dem die Datenbank lagert, bei *User* einfach *Yes*, und schließlich bei *Database* den Namen der Datenbank.

Nach Bestätigung mit *OK* kommt man wieder in das Ausgangsfenster aus Abbildung 2.14 zurück, mit dem Unterschied, dass nun die gerade erstellte Verknüpfung sichtbar wird, siehe Abbildung 2.18. Die Einrichtung der ODBC-Schnittstelle ist abgeschlossen.

Um mit STATISTICA über die neu geschaffene ODBC-Schnittstelle auf die Datenbank `verkauf` zuzugreifen, gehen wir erst einmal genauso vor, wie im vorigen Abschnitt 2.4 beschrieben, bis wir zu dem Dialog aus Abbildung 2.6 gelangen. Hier wählen wir aber nun, im Unterschied zu vorher, den Eintrag *Microsoft OLE DB Provider for ODBC Drivers* aus, siehe Abbildung 2.19. Wir klicken auf *Weiter* ≫ und gelangen zum Dialog aus Abbildung 2.20. Hier können wir unter *1.* bei *Datenquellenname*

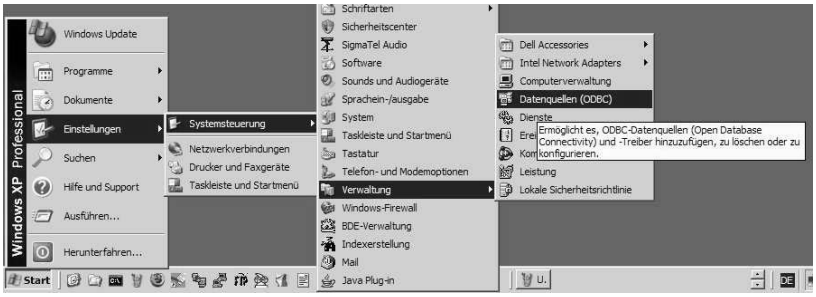


Abb. 2.14: *Menü zum Anlegen einer ODBC-Schnittstelle.*

Abb. 2.15: *Benutzerdatenquelle hinzufügen.*

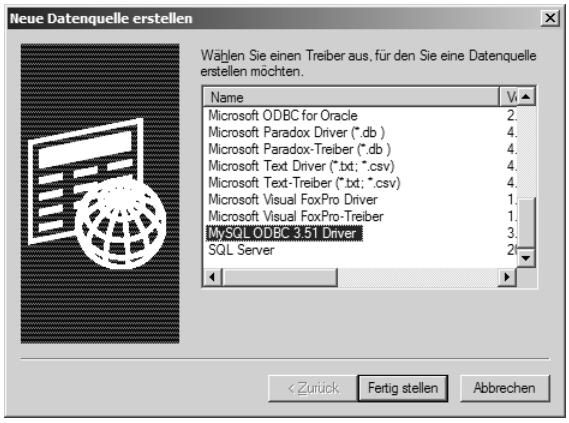


Abb. 2.16: *Datenbanktreiber auswählen.*

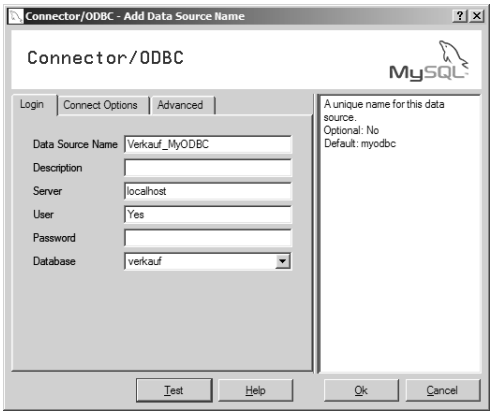


Abb. 2.17: *Datenbankverbindung bestimmen.*

Abb. 2.18: *Die neu erstellte Datenbankverknüpfung.*

verwenden die eben eingerichtete ODBC-Schnittstelle auswählen, siehe Abbildung 2.20. Nach einem Klick auf *OK* verfahren wir völlig analog wie im vorigen Abschnitt 2.4 ab Abbildung 2.8.

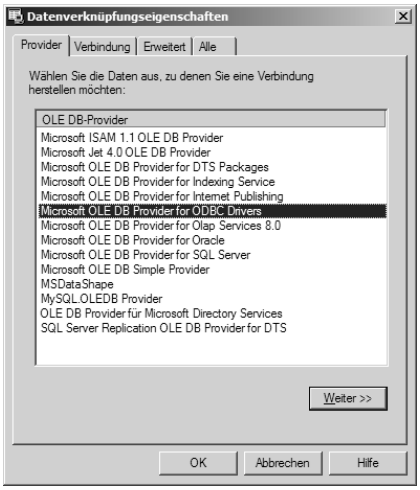


Abb. 2.19: *ODBC-Anbieter auswählen.*

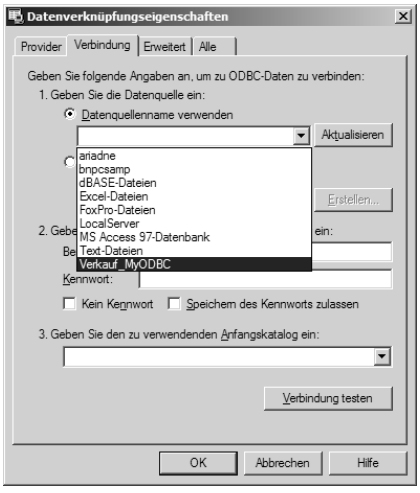


Abb. 2.20: *Datenbankverknüpfung auswählen.*

2.6 Anbindung an ACCESS-Datenbanken

Wie bereits in Abschnitt 2.5 erwähnt, kann man eine ODBC-Schnittstelle auch direkt von STATISTICA aus einrichten. Das dazu notwendige Vorgehen soll beispielhaft am Datenbanksystem ACCESS demonstriert werden. Bei Datenbanken anderer Hersteller ist ein analoges Vorgehen möglich.

Gehen wir auch wieder vom Beispiel eines Handelsunternehmens aus. In der ACCESS-Datenbank `Verkauf.mdb` befinden sich die gleichen Daten wie in der MySQL-Datenbank `verkauf`, siehe Abschnitt 2.4.

Um auf diese Datenbank zuzugreifen, gehen wir erst einmal wie in Abschnitt 2.4 bis Abbildung 2.5 vor. Im Dialog *Datenverknüpfungseigenschaften*, Karte *Provider*, wählen wir dann den *Microsoft OLE DB Provider for ODBC Drivers* aus, vgl. Abbildung 2.19, und klicken *Weiter* ≫.

Auf der anschließenden Karte *Verbindung*, siehe Abbildung 2.21, könnten wir nun die passende ODBC-Schnittstelle auswählen, wenn wir sie denn schon erstellt hätten, vgl. den vorigen Abschnitt 2.5. Da dies jedoch noch nicht geschehen ist, wählt man unter Punkt *1. Verbindungszeichenfolge verwenden* und klickt dann auf *Erstellen...*.

Im folgenden Dialog aus Abbildung 2.22 ist noch keine passende `.dsn`-Datei vorhanden. Also klickt man auf den Knopf *Neu...* hinter dem Feld *DSN-Name*, worauf sich ein Fenster wie in Abbildung 2.23 öffnet.

Dort wählen wir den *Microsoft Access-Treiber (`*.mdb`)* und klicken auf *Weiter >*. Im anschließenden Dialog aus Abbildung 2.24 vergeben wir einen Namen für die Datenbankverknüpfung, im Beispiel *Accessdatenbank*, und klicken erneut auf *Weiter >*. Abschließend müssen wir nochmal mit *Fertig stellen* bestätigen, siehe Abbildung 2.25.

Nun kommen wir zum Dialog aus Abbildung 2.26, in welchem wir jetzt die gewünschte *Datenbank: Auswählen* können. Jetzt ist es endlich soweit: In diesem Dialog, siehe Abbildung 2.27, können wir die Datenbankdatei[4] `Verkauf.mdb` auswählen und mit *OK* bestätigen. Dann gelangen wir zurück in den Dialog aus Abbildung 2.26, klicken nochmals *OK*, und kommen wieder in den Dialog aus Abbildung 2.21.

Im Gegensatz zu früher findet sich jetzt bei *Verbindungszeichenfolge* ein

[4] Im Gegensatz zu größeren Datenbanksystemen wie etwa MySQL sind ACCESS-Datenbanken in einer einzelnen Datei zusammengefasst, die sich auch an einem beliebigen Ort der Festplatte befinden darf. Bei MySQL dagegen liegt die Datenbank in einem Unterordner des Verzeichnisses `C:\Programme\...\MySQL\data` vor, wobei für eine jede Tabelle drei Einzeldateien angelegt sind.

Abb. 2.21: *Verbindungszeichenfolge erstellen.*

Eintrag wie etwa

```
DBQ=D:\EigeneDateien\...\Verkauf.mdb;
DefaultDir=D:\...;
Driver=Microsoft Access-Treiber (*.mdb);
DriverId=281;FIL=MS Access;
FILEDSN=C:\Programme\Gemeinsame Dateien\ODBC\Data
Sources\Accessdatenbank.dsn;
MaxBufferSize=2048;MaxScanRows=8;PageTimeout=5;
SafeTransactions=0;Threads=3;
UID=admin;UserCommitSync=Yes;
```

Ein weiteres Mal wählt man *OK* und verfährt nun völlig analog wie in Abschnitt 2.4 ab Abbildung 2.8.

Das in den Abschnitten 2.4 bis 2.6 beschriebene Vorgehen zum Andocken an Datenbanken ist zugegebenermaßen nicht ganz einfach. Es sei aber bemerkt, dass der Großteil der Schritte pro Datenbank nur ein einziges Mal zu durchlaufen ist. Ist nämlich erst einmal eine Verknüpfung zu einer Datenbank erstellt worden, so kann man bei zukünftigen Zugriffen, wie bereits erwähnt, bei Abbildung 2.10 aus Abschnitt 2.4 einsteigen.

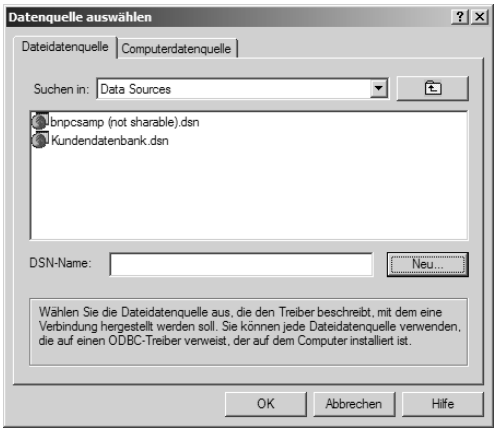


Abb. 2.22: *Neue Dateidatenquelle erstellen.*

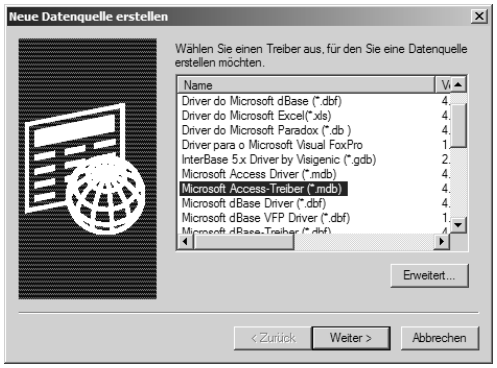


Abb. 2.23: *Microsoft Access-Treiber auswählen.*

Abb. 2.24: *Benennen der Dateidatenquelle ...*