

5 Einfache Datentransformationen

Mit SPSS können Sie Datentransformationen durchführen, die von einfachen Aufgaben wie der Zusammenfassung von Kategorien für eine Analyse bis zur Erzeugung neuer Variablen auf der Grundlage komplexer Berechnungen und bedingter Anweisungen reichen. Personen, die bereits mit Tabellenkalkulationen gearbeitet haben und von dort die Möglichkeit der Eingabe von Formeln in Tabellenzellen kennen, werden diese Möglichkeit bei SPSS vermissen. In SPSS gibt es für solche Transformationen und Berechnungen eigene Menüpunkte. In der Datendatei selbst stehen nur Datenwerte, aber keine Formeln.

Fortgeschrittene Anwender setzen für diese Aufgaben die Syntaxsprache von SPSS ein. Diese Befehlssprache von SPSS stellt alle zur Programmierung notwendigen Konstrukte wie Schleifenbefehle, Sprungbefehle und bedingte Anweisungen sowie eine große Zahl an Funktionen zur Verfügung.

5.1 Umkodierung von Variablen

Sie können Datenwerte durch Umkodierung modifizieren. Dies ist vor allem nützlich bei der Zusammenfassung oder dem Kombinieren von Kategorien. Sie können die Werte innerhalb vorhandener Variablen umkodieren oder das Ergebnis des Umkodierens als neue Variable dem Daten-Editor hinzufügen. Die Dialogfelder, in denen Sie die Umkodierung vornehmen können, finden Sie unter **Transformieren/Umkodieren in dieselben Variablen** bzw. unter **Transformieren/Umkodieren in andere Variablen**.

Ein Beispiel aus unserer Datendatei: Die befragten Studenten sollten für verschiedene Softwaretypen auf einer fünfstufigen Rating-Skala angeben, wie intensiv sie Programme des jeweiligen Typs nutzen. Nehmen wir an, für eine Präsentation der Ergebnisse reiche es aus, nur ganz grob anzugeben, ob die Befragten Programme des jeweiligen Anwendungstyps überhaupt kennen/nutzen. Dazu fassen wir die ersten vier Stufen der Skala zusammen. Da wir die detaillierteren Informationen der fünfstufigen Skala nicht verlieren wollen, wählen wir eine Umkodierung in eine neue Variable.

Im Dialogfeld **Umkodieren in andere Variablen** (siehe Abbildung 57) wird aus der Quellvariablenliste die umzukodierende Variable ausgewählt und unter „Ausgabevariable“ der Namen der neuen Variable eingegeben (in unserem Beispiel `dtncbankn` in die Variable `dtncbank2`). Unter „Beschriftung“ kann ein Variablenlabel für die neue Variable vergeben werden. Durch Klicken auf die Schaltfläche „Ändern“ wird die neue Variable in das mittlere Feld übertragen, und wir können mit der Eingabe der umzukodierenden Werte beginnen. Dazu klicken wir auf die Schaltfläche „Alte und neue Werte“, wodurch wir in das in Abbildung 58 gezeigte Dialogfeld **Umkodieren in andere Variablen: Alte und neue Werte** gelangen. Hier können wir angeben, welche Werte(bereiche) wir durch welchen Wert ersetzen wollen. Jede Angabe wird durch

Klicken auf die Schaltfläche „Hinzufügen“ in die Liste der durchzuführenden Transformationen aufgenommen. Nachdem alle Transformationen festgelegt sind, gelangen wir durch Klicken auf die Schaltfläche „Weiter“ wieder in das Dialogfeld **Umkodieren in andere Variablen**, das wir nun durch Klicken auf die Schaltfläche „OK“ verlassen können. Wenn in den Systemeinstellungen festgelegt ist, dass Transformationen direkt ausgeführt werden (vgl. Abschnitt 9 „Verändern der SPSS-Systemeinstellungen“), werden die Werte der neuen Variable sofort berechnet, und die neue Variable wird im Eingabefenster wie eine selbst eingegebene Variable angezeigt.

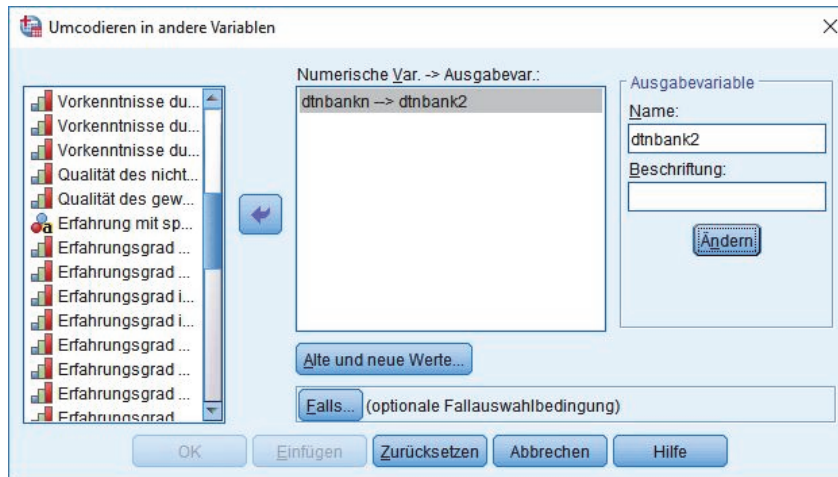


Abbildung 57: Dialogfeld zum Umkodieren der Werte einer Variablen in eine andere Variable

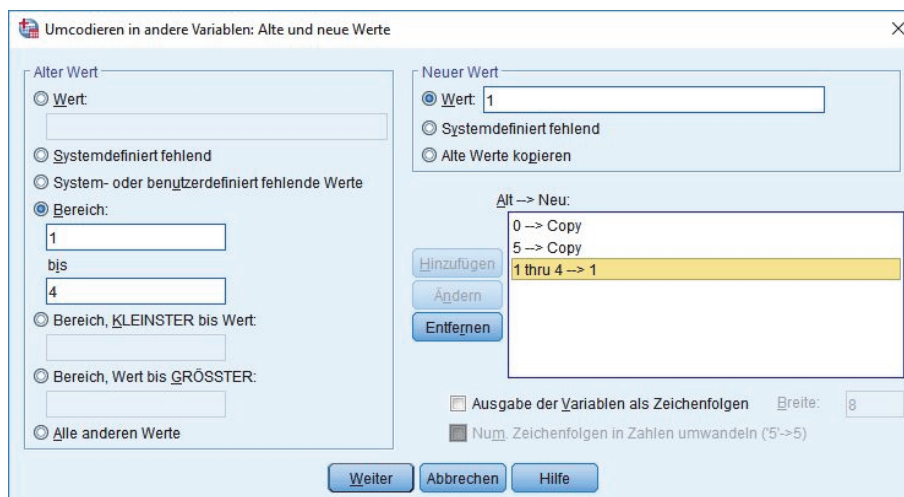


Abbildung 58: Unterdiallogfeld zur Eingabe der Wertetransformationen („Alte und neue Werte“)

Falls Sie eine solche Umkodierung nicht, wie in unserem Beispiel geschehen, für alle Fälle durchführen wollen, können Sie durch Eingabe einer Bedingung nur die Fälle auswählen, die für Sie relevant sind, bevor Sie die Berechnung mit „OK“ starten. Dazu wählen Sie die Schaltfläche „Falls“ und formulieren in der Dialogbox **Umkodieren in andere Variablen: Falls Bedingung erfüllt ist** Ihre Bedingung in Form eines logischen Ausdrucks und wählen die Option „Fall einschließen, wenn Bedingung erfüllt ist“. Die Dialogbox stellt zur Formulierung der Bedingung die

gleichen Elemente zur Verfügung wie die im nächsten Abschnitt behandelte Dialogbox zur Berechnung einer neuen Variablen und kann analog bedient werden.

Den doch recht umfangreichen Klickoperationen unseres Beispiels entspricht in der SPSS-Steuersprache das nachfolgend wiedergegebene kleine Programmsegment.

```
RECODE
  dtnbankn
  (0=Copy) (5=Copy) (1 thru 4=1) INTO dtnbank2 .
```

Dies mag Ihnen einen Eindruck vermitteln, wie solche umfangreichen Auswahlen in der SPSS-Syntaxsprache aussehen. Viele erfahrene SPSS-Anwender finden es einfacher und eleganter, die SPSS-Steuersprache zu benutzen, um ihre „SPSS-Jobs“ durchzuführen. Es ist natürlich auch möglich, beide Vorgehensweisen zu kombinieren.

5.2 Berechnung neuer Variablen

Um Werte für eine Variable zu berechnen, die auf komplexeren Formeln bestehend aus Zahlen, Variablen und Operationen beruhen, wählen Sie aus dem Menü **Transformieren** den Befehl **Variable berechnen**. In der daraufhin erscheinenden Dialogbox **Variable berechnen** können Sie einen Ausdruck zur Berechnung einer neuen Variablen erzeugen, indem Sie die benötigten Variablen aus der Variablenliste, die benötigten arithmetischen, relationalen und logischen Operatoren aus der „Taschenrechnertastatur“ sowie die gewünschten Funktionen aus dem Listenfeld mit mittlerweile rund 200 Funktionen auswählen, wobei Sie die Funktionen nach Funktionsgruppen filtern können.

Nehmen wir an, wir wollen bei unserem Beispieldatensatz einen Wert für die mittlere Nutzung des Computers über die verschiedenen Anwendungstypen hinweg berechnen. Dann können wir das, wie in der nachfolgenden Abbildung 59 gezeigt, tun, indem wir einen Namen für die neue Variable eingeben (hier: *xnutzung*) und die gewünschte(n) Funktion(en) sowie deren Argumente auswählen. In unserem Beispiel berechnen wir einfach für jede Person den Mittelwert aus den Ratingwerten der verschiedenen Anwendungstypen. Diese Berechnungen könnten wir auch auf die Personen beschränken, die überhaupt Erfahrungen mit irgendwelchen Anwendungen haben, also auf die entsprechende Globalfrage mit „Ja“ geantwortet haben. Diese Begrenzung würde man in der Dialogbox **Variable berechnen: Falls Bedingung erfüllt ist**, die man durch Anklicken der Schaltfläche „Falls“ erreicht, vornehmen.

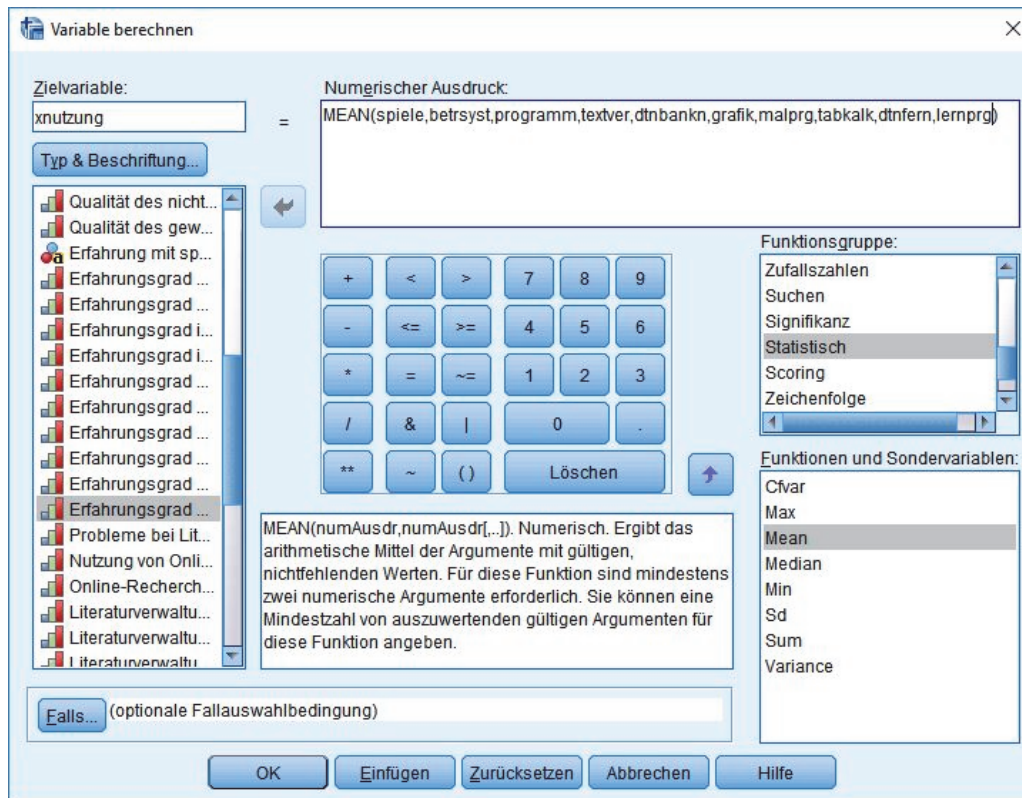


Abbildung 59: Dialogfeld zur Berechnung neuer Variablen mittels numerischer Ausdrücke

Eine genaue Beschreibung der zur Verfügung stehenden Funktionen erhalten Sie beispielsweise in der Online-Hilfe von SPSS, wenn Sie nach dem Stichwort „Funktionen“ suchen.

Auch bei diesem Beispiel sei zur Veranschaulichung das entsprechende Programmsegment gezeigt.

```
COMPUTE
  xnutzung =
  MEAN(spiele,betrsyst,programm,textver,dtnbankn,grafik
  ,malprg,tabkalk,dtnfern,lernprg) .
```

Die durch Umkodierung (RECODE) oder Berechnung (COMPUTE) erzeugten neuen Variablen gehören genauso zu Ihrer Arbeitsdatei wie die von Ihnen selbst eingegebenen. Wenn Sie Ihre Arbeitsdatei abspeichern und somit eine aktualisierte Version der Datendatei erzeugen, werden die neuen Variablen mit gespeichert und stehen Ihnen dann auch in weiteren Sitzungen zur Verfügung. Sie können natürlich auch die Variablendefinition nutzen, um den neuen Variablen Typ, Label usw. zuzuweisen.

Da es bei neuen Anwendern immer wieder zu Missverständnissen kommt, sei noch einmal betont: Anders als beispielsweise bei Formeln in Excel werden in SPSS auch die Ergebnisse der Umkodierungen und Berechnungen nicht aktualisiert, wenn Sie an den Werten von Variablen,

die für Berechnungen herangezogen wurden, Änderungen vornehmen. Sie selbst müssen die Umkodierungen bzw. Berechnungen erneut durchführen, damit die Werte der Zielvariablen neu berechnet werden.

Wenn Sie also wie oben beschrieben die Variable `xNUTZUNG` berechnet haben und anschließend feststellen, dass bei einem Fall ein falscher Wert für die Variable `TEXTVER` in die Datendatei eingetragen wurde, reicht es nicht aus den Wert der Variable `TEXTVER` zu korrigieren. Da der Wert der Variablen `xNUTZUNG` mit dem falschen Wert der Variablen `TEXTVER` berechnet wurde und somit auch nicht stimmt, müssen Sie die Variable `xNUTZUNG` erneut berechnen.

Ebenso ist es selbstverständlich, wenn Sie nach Berechnung der Variable `xNUTZUNG` weitere Fälle eingeben. Auch für diese neuen Fälle führt SPSS keine Berechnung für `xNUTZUNG` durch, sondern setzt einfach die Kodierung für den systembedingt fehlenden Wert ein, wenn Sie keinen gültigen Wert eingeben bzw. berechnen lassen.

Dazu ein Tipp: Nutzen Sie den Syntax-Editor für die Berechnungen. Dazu nehmen Sie alle Einstellungen wie oben gezeigt in dem entsprechenden Dialogfenster (z.B. **Variable berechnen**) vor, bestätigen die Eingabe aber nicht mit „OK“, sondern mit „Einfügen“. Nun schreibt SPSS die Syntax-Befehle für Sie in den Syntax-Editor und zeigt Ihnen diesen an. Anschließend markieren Sie den ganzen Befehl mit der Maus und starten über **Ausführen/Auswahl** die Berechnung. Das wird Ergebnis im Viewer ausgegeben.

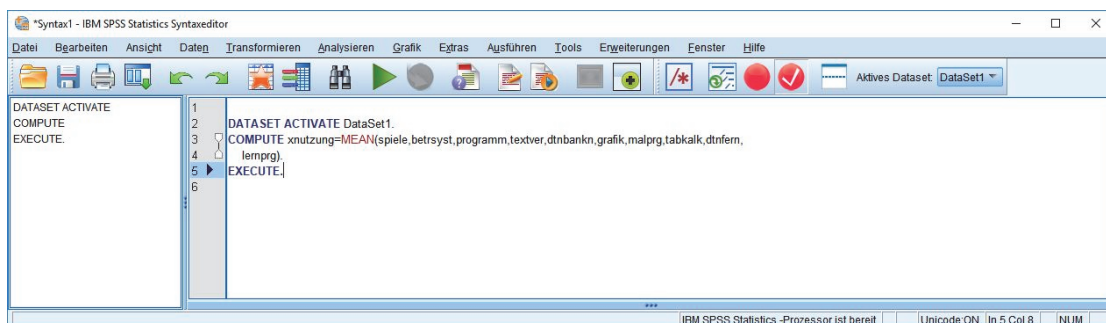


Abbildung 60: Durchführung der Berechnung im Syntax-Editor

Wenn Sie nun einzelne Werte in den Variablen ändern oder Fälle hinzufügen, können Sie die Berechnungen einfach im Syntax-Editor aktualisieren. Speichern Sie die Syntax, bevor Sie den Editor beenden, können Sie die Syntax sogar in späteren wieder öffnen und zur Aktualisierung verwenden!