

Hinweise zu einer effektiven und transparenten Datenanalyse

Probleme mit der quantitativen Datenanalyse lassen sich vermeiden, wenn Sie erstens die Arbeitsschritte in der **richtigen Reihenfolge** abarbeiten und zweitens (**und vor allem**) mit einer **Programmsyntax** arbeiten. Diese lässt sich entweder erzeugen, indem man die Syntaxbefehle direkt in ein Syntaxfenster schreibt (das bedarf allerdings guter Kenntnisse der Programmiersprache) oder (in SPSS oder im r Commander) im Menü jeweils auf „Einfügen“ (bzw. „Paste“) klickt statt auf „OK“.

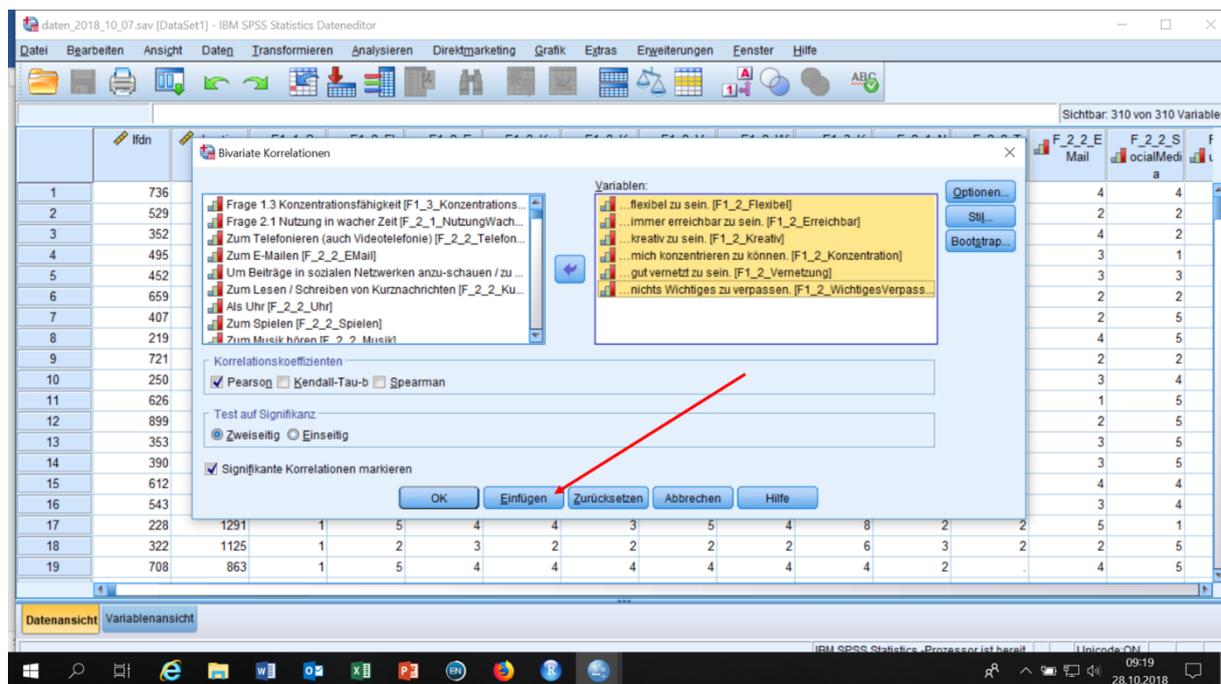


Abb. 1: Erzeugen von Syntaxbefehlen in SPSS durch „Einfügen“

Dann fügt das Programm den gewünschten Befehl am Ende der Syntaxdatei ein. Wenn dann auch noch die Reihenfolge der Befehle sinnvoll ist (siehe dazu unten), führt das idealerweise zu Folgendem: Alle durchgeführten Datenmanipulationen und Auswertungen können jederzeit wiederholt werden (durch Forschende wie Betreuende), indem einfach die Syntax noch einmal laufen gelassen wird (beginnend jeweils mit dem Laden des Originaldatensatzes). Dies ist dann wichtig, wenn Sie bemerken (und das ist der Standardfall!), dass Sie noch etwas an Ihren Daten verändern müssen. Z.B. könnten noch unmögliche Codes in Ihren Daten vorkommen, die die Ergebnisse verzerren (zum Beispiel „0“-Codes unklarer Herkunft, wie sie in Questback gern vorkommen), oder Sie bemerken, dass Sie eine Recodierung nochmals verändern müssen oder dass Sie Ihre Daten doch besser gewichten. Ohne Syntax müssten Sie dann evtl. tage- oder wochenlang Analysen rekonstruieren und wiederholen. Mit Syntax lässt sich dies dann binnen Sekunden erledigen. Ein weiterer Vorteil der Arbeit mit der Syntax besteht darin, dass sich in ihr mit „Copy & Paste“ leicht zahllose Varianten eines Befehls erzeugen lassen, (z.B. wenn der Effekt verschiedener unabhängiger Variablen in Kreuztabellen überprüft werden soll, dann kann immer der gleiche Befehl verwendet werden, nur der Name der unabhängigen Variable muss ausgetauscht werden). Zudem ermöglicht die Ablieferung der Syntax und des Datensatzes im digitalen Anhang den Betreuenden einen problemlosen Nachvollzug der durchgeführten Analysen. Die Ablieferung der Syntax in bereinigter und kommentierter (siehe Empfehlung 1 unten) Form (also nur die Analysen enthaltend, die für die Arbeit auch verwendet wurden, keine Fehlversuche etc.) wird deshalb auch erwartet!

Eine ideale Syntax arbeitet dabei die folgenden Schritte in dieser Reihenfolge ab:

1. Laden eines bereinigten Datensatzes. Die Prüfung unmöglicher oder unlogischer Eingaben sowie die Löschung unbrauchbarer Fälle kann vorher erfolgen (weil per Syntax auch nicht einfach), weil Sie und Ihre Betreuenden den unbereinigten Datensatz normalerweise nicht benötigen. Eine Kopie des unbereinigten Datensatzes sollten Sie aber auch gespeichert haben (für den Fall, dass Sie beim Bereinigen Fehler machen...)
2. Prüfung der Stichprobe auf Gewichtungbedarf. Durch einfache Häufigkeitsauszählung meist soziodemografischer Variablen lässt sich erkennen, inwieweit die Verteilung wichtiger Merkmale in der Stichprobe der Verteilung in der Grundgesamtheit entspricht. Bei überschaubaren Abweichungen lassen sich dann Gewichtungsfaktoren berechnen, mit denen jeder einzelne Fall entweder herauf- oder heruntergewichtet werden kann.
3. ggf. Erstellung von Gewichtungsvariablen. Sie vergleichen die Verteilung der soziodemografischen Variablen mit vorhandenen Daten über Ihre Grundgesamtheit und berechnen daraus für jede Person einen Gewichtungsfaktor. Bsp.: In der Grundgesamtheit gibt es 50% Frauen, 50% Männer. Sie haben in der Stichprobe aber 60% Frauen und 40% Männer (das dritte Geschlecht jetzt einmal beiseite gelassen). Aus der Formel $x=50/40$ ergibt sich ein Gewichtungsfaktor x für die Männer von 1,25, analog für die Frauen ($x=50/60$) = 0,83.
4. Datentransformation. Vor Beginn statistischer Analyse ist zu prüfen, inwieweit Variablen verändert (recodiert), neu erstellt (zum Beispiel als Indizes) und/oder gelabelt werden müssen. Recodierungen sind vor allem nötig
 - bei negativ gepolten Items, die mit positiv gepolten Items zu Indizes zusammengefasst werden sollen.
 - bei feinen Skalen, die zu größeren Klassen zusammengefasst werden sollen.
5. Datenanalyse. Erst jetzt, mit bereinigten, evtl. gewichteten, recodierten und gelabelten Daten sollten Sie mit der statistischen Datenanalyse beginnen. Dies schließt nicht aus, dass Sie im Laufe der Analyse weiteren Bedarf an Datentransformationen erkennen. Diese lassen sich dann im entsprechenden Syntaxblock nachtragen.

Empfehlung 1: Kommentieren Sie die Syntax auch, damit Sie selbst, aber auch Ihre Betreuenden später leicht verstehen, welche Syntaxteile welche Funktion erfüllen sollen. In SPSS beginnen Kommentarzeilen mit „*“, in R mit „##“. Durch diese Kennzeichnung erkennt das Programm, dass es sich nicht um „Befehle“ handelt und ignoriert diese Zeilen. So bekommen Sie keine Fehlermeldungen.

Empfehlung 2: Speichern Sie am Ende einer Sitzung nur die Syntax ab, aber nicht den veränderten Datensatz. Denn bei der nächsten Sitzung mit einem veränderten Datensatz weiterzurechnen, beinhaltet Problemquellen: Einerseits könnten Sie etwas falsch gemacht und dies dann abgespeichert haben. Andererseits kann es passieren, dass bei erneutem Durchlaufen der Syntax dann bestimmte selbstbezügliche Recodierungen noch einmal ausgeführt und dadurch unwirksam werden (Bsp. „1=5, 2=4, 3=3, 4=2, 5=1“). Dies lässt sich vermeiden, wenn Sie immer mit dem Originaldatensatz arbeiten und die nötigen Manipulationen daran jeweils neu durchführen, indem Sie die entsprechenden Syntaxteile markieren und ablaufen lassen.

Nachbemerkung: Im Prinzip gilt alles oben Gesagte für SPSS wie R in gleicher Weise, weshalb auf konkrete Bezüge darauf weitgehend verzichtet wurde.