

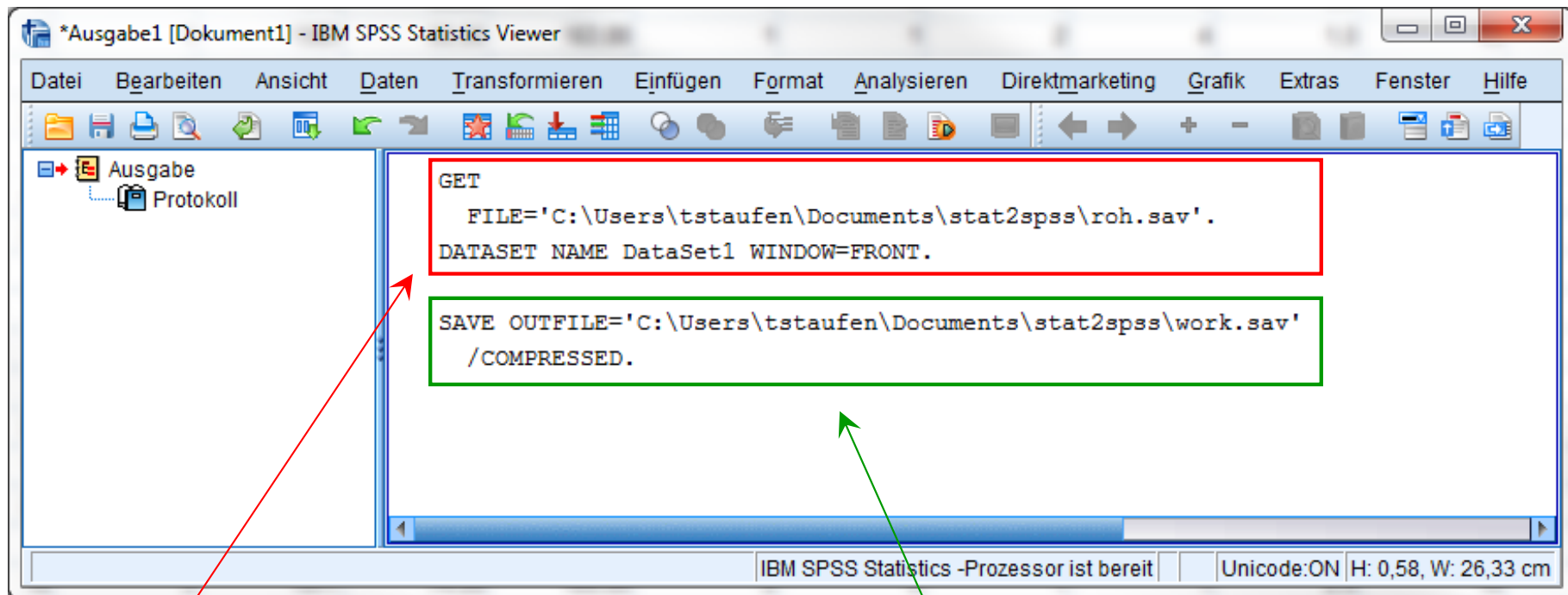
- 1 Workflow bei (größeren) SPSS-Projekten
- 2 Selektives Speichern von Daten
- 3 Datenimport
- 4 Datenexport
- 5 Transponieren von Datendateien

Workflow bei (größeren) SPSS-Projekten

- Bei umfassenderen „SPSS-Projekten“ ist die Verwendung der Syntax aus den genannten Gründen unbedingt anzuraten.
- Im Folgenden werden noch ein paar Hinweise gegeben, wie man solche Projekte aufbauen kann, um eine entsprechende Übersichtlichkeit und Datensicherheit zu gewährleisten.
- Sinnvoll ist, die **Transformationen** und die **Analysen** in separate Steuerdateien auszugliedern und einer Rohdaten- und einer Analysedatei zu unterscheiden.
- In diesem Zusammenhang ist es nützlich, Datendateien auch innerhalb der Syntax öffnen und speichern zu können, was wir im Folgenden zuerst behandeln.

Öffnen und Speichern von Datendateien per Syntax

- Das Öffnen und Speichern der SPSS-Datendateien kann auch in der Syntax erfolgen. Die Befehle kann man z.B. im Ergebnis-Viewer sehen, wenn man vorab die entsprechenden Aktionen per Dialog aufgerufen hat. Sie lassen sich auch von dort in die Syntax kopieren.



The screenshot shows the IBM SPSS Statistics Viewer interface. The main window displays the following syntax commands:

```
GET  
  FILE='C:\Users\tstausen\Documents\stat2spss\roh.sav'.  
DATASET NAME DataSet1 WINDOW=FRONT.  
  
SAVE OUTFILE='C:\Users\tstausen\Documents\stat2spss\work.sav'  
  /COMPRESSED.
```

Two red boxes highlight the `GET` command and the `SAVE OUTFILE` command. A red arrow points from the `GET` command to a text box below, and a green arrow points from the `SAVE OUTFILE` command to another text box below.

Der **GET FILE**-Befehl zum Öffnen einer SPSS-Datendatei

Der **SAVE OUTFILE**-Befehl zum Speichern einer SPSS-Datendatei

Öffnen und Speichern von Datendateien per Syntax

- Nützlich sind diese beiden Befehle z.B., wenn man den **Öffnen**-Befehl an den Beginn und den **Speichern**-Befehl an das Ende einer (**Transformations-**) **Steuerdatei** setzt, die alle in dem Projekt benötigten Berechnungen und Umcodierungen von Variablen enthält:

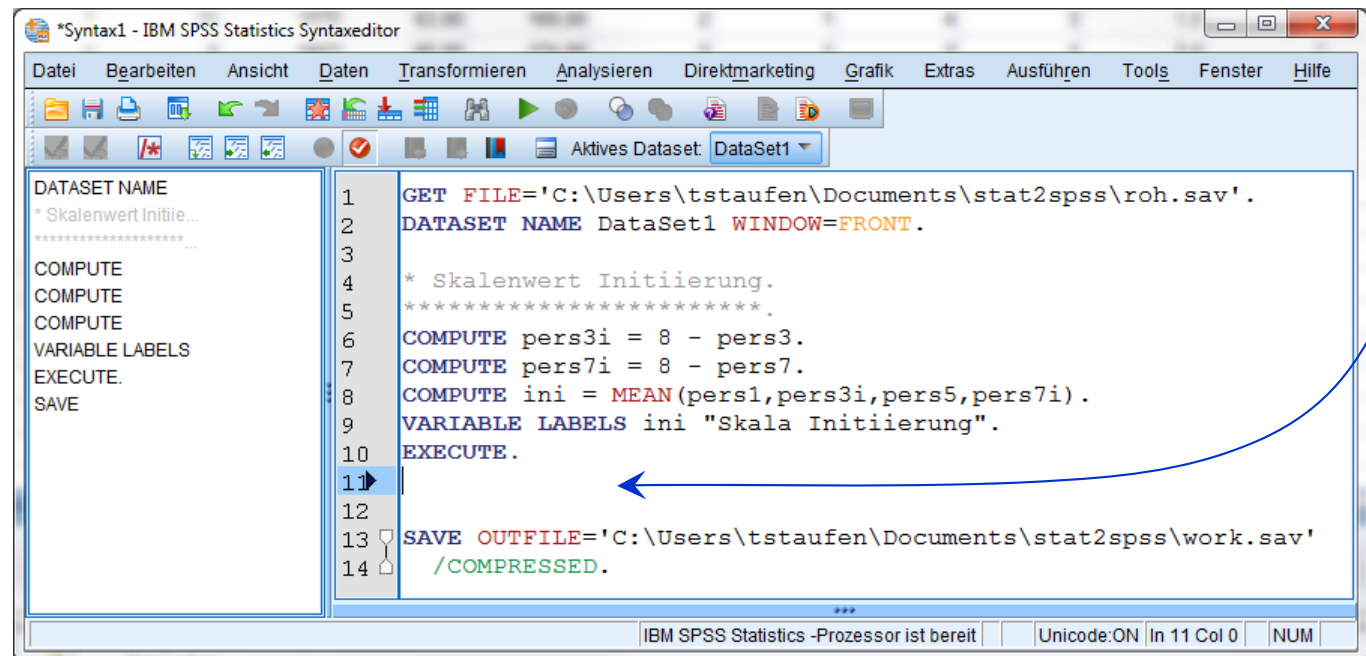
```
GET FILE='d:\stat2spss\roh.sav'  
DATASET NAME Rohdaten WINDOW=FRONT.  
  
* Skalenwert Initiierung.  
*****.  
COMPUTE pers3i = 8 - pers3.  
COMPUTE pers7i = 8 - pers7.  
COMPUTE ini = MEAN.3(pers1,pers3i,pers5,pers7i).  
VARIABLE LABELS ini "Skala Initiierung".  
EXECUTE.  
  
SAVE OUTFILE='d:\stat2spss\work.sav'  
/COMPRESSED.
```

Transformationen

- Wenn man nun die gesamte Steuerdatei markiert (z.B. per (Strg) (A)) und ausführt (z.B. per (Strg) (R)) wird automatisch die Rohdatendatei **roh.sav** geöffnet, alle Transformationen werden durchgeführt und anschließend die komplette Datendatei incl. der Transformationen unter einem anderen Namen (hier: **work.sav**) gespeichert.

Workflow bei (größeren) SPSS-Projekten

- Die Vorteile dieser Automatisierung sind (u.a.):
 - Die Rohdatendatei (hier `roh.sav`) bleibt, nachdem man sie einmal kontrolliert und korrigiert hat, immer unangetastet. Alle möglichen - ggf. auch fehlerhaften - Berechnungen werden nie in der Rohdatendatei gespeichert und können dort auch kein Unheil anrichten.
 - Man kann die Berechnungen jederzeit korrigieren oder auch [hier](#) ergänzen und anschließend die Daten (hier in der Datei `work.sav`) leicht auf den neuesten Stand bringen, in dem man die gesamte Steuerdatei markiert und ausführt.
 - Die Namen der Datendateien sind dokumentiert.



The screenshot shows the IBM SPSS Statistics Syntax Editor window. The title bar reads '*Syntax1 - IBM SPSS Statistics Syntaxeditor'. The menu bar includes 'Datei', 'Bearbeiten', 'Ansicht', 'Daten', 'Transformieren', 'Analysieren', 'Direktmarketing', 'Grafik', 'Extras', 'Ausführen', 'Tools', 'Fenster', and 'Hilfe'. The toolbar contains various icons for file operations and execution. The main text area contains the following syntax script:

```
1 GET FILE='C:\Users\tstausen\Documents\stat2spss\roh.sav'.
2 DATASET NAME DataSet1 WINDOW=FRONT.
3
4 * Skalenwert Initiierung.
5 *****
6 COMPUTE pers3i = 8 - pers3.
7 COMPUTE pers7i = 8 - pers7.
8 COMPUTE ini = MEAN(pers1,pers3i,pers5,pers7i).
9 VARIABLE LABELS ini "Skala Initiierung".
10 EXECUTE.
11
12
13 SAVE OUTFILE='C:\Users\tstausen\Documents\stat2spss\work.sav'
14 /COMPRESSED.
```

A blue arrow points from the word 'hier' in the text above to line 11 of the syntax script. The status bar at the bottom indicates 'IBM SPSS Statistics -Prozessor ist bereit', 'Unicode:ON', 'In 11 Col 0', and 'NUM'.

Workflow bei (größeren) SPSS-Projekten

- Es ist also sinnvoll, die Berechnungen (also z.B. COMPUTE und IF-Anweisungen sowie VARIABLE und VALUE LABELS) in eine Transformations-Steuerdatei zu schreiben und die Analysen mittels Prozeduren in einer andere Analyse-Steuerdatei.
- Der **Transformations-Steuerdatei** enthält alle Berechnungen, Erzeugungen neuer Variablen, Umcodierungen, usw. Die am Ende gespeicherte Datei (hier `work.sav`) ist die Basis für die weiteren Analysen (Prozeduren).
- Die **Analyse-Steuerdatei(en)** arbeiten mit der um alle Transformationen ergänzten Datendatei (hier `work.sav`). Wenn man will, kann man ihr (ihnen) den Öffnen-Befehl der betreffenden Datei voranstellen:

```
GET FILE='d:\stat2spss\work.sav'.  
DATASET NAME ArbeitsDaten WINDOW=FRONT.  
  
* Häufigkeitsverteilung von Körpergröße & -gewicht.  
*****.  
FREQUENCIES  
  VARIABLES=groesse gewicht  
  /ORDER=ANALYSIS.
```

Workflow bei (größeren) SPSS-Projekten

phantas	sex	gebtag	gebmon	gebjahr	gewicht	groesse	religion	raucher	schulh	schu
1	adelerolf	2	24	10	1983	56.00	160.00	2	0	4
2	aida	2	19	11	1976	68.00	168.00	1	0	3
3	Angela	2	14	1	1983	50.00	163.00	1	1	2
4	Apfel	2	8	8	1974	60.00	160.00	1	0	2
5	Apollina	2	27	6	1978	50.00	153.00	1	1	1
6	atom	2	11	11	1982	65.00	171.00	2	1	3
7	Bacardi	1	20	7	1982	62.00	177.00	2	1	1
8	bachus	2	29	11	1962	168.00	174.00	2	1	1
9	Bianunda	2	7	10	1976	63.00	169.00	2	1	4
10	beltatik	2	1	3	1977	60.00	174.00	2	1	2
11	Bienemaja	2	7	12	1973	63.00	165.00	1	1	3
12	Budnikovna	2	3	3	1975	49.00	163.00	1	1	1
13	candia	2	25	7	1981	53.00	168.00	1	0	1
14	Christmas	2	19	9	1983	47.00	164.00	2	1	4
15	cookies	2	22	2	1984	57.00	173.00	2	0	4

roh.sav
(Rohdaten)

phantas	sex	gebtag	gebmon	gebjahr	gewicht	groesse	religion	raucher	schulh	schu
1	adelerolf	2	24	10	1983	56.00	160.00	2	0	4
2	aida	2	19	11	1976	68.00	168.00	1	0	3
3	Angela	2	14	1	1983	50.00	163.00	1	1	2
4	Apfel	2	8	8	1974	60.00	160.00	1	0	2
5	Apollina	2	27	6	1978	50.00	153.00	1	1	1
6	atom	2	11	11	1982	65.00	171.00	2	1	3
7	Bacardi	1	20	7	1982	62.00	177.00	2	1	1
8	bachus	2	29	11	1962	168.00	174.00	2	1	1
9	Bianunda	2	7	10	1976	63.00	169.00	2	1	4
10	beltatik	2	1	3	1977	60.00	174.00	2	1	2
11	Bienemaja	2	7	12	1973	63.00	165.00	1	1	3
12	Budnikovna	2	3	3	1975	49.00	163.00	1	1	1
13	candia	2	25	7	1981	53.00	168.00	1	0	1
14	Christmas	2	19	9	1983	47.00	164.00	2	1	4
15	cookies	2	22	2	1984	57.00	173.00	2	0	4

work.sav
(Rohdaten plus transformierte Variablen)

```
GET FILE='d:\stat2spss\roh.sav'.
DATASET NAME Rohdaten WINDOW=FRONT.
```

```
* Skalenwert Initiierung.
*****.
COMPUTE pers3i=8-pers3.
COMPUTE pers7i=8-pers7.
COMPUTE ini=MEAN.3(pers1,pers3i,pers5,pers7i).
VARIABLE LABELS ini "Skala Initiierung".
EXECUTE.
```

```
SAVE OUTFILE='d:\stat2spss\work.sav'
/COMPRESSED.
```

```
GET FILE='d:\stat2spss\work.sav'.
DATASET NAME Arbeitsdaten WINDOW=FRONT.
```

```
* Häufigkeitsverteilung von Körpergröße.
*****.
FREQUENCIES
  VARIABLES=groesse
  /ORDER=ANALYSIS.
```

...

tra.sps

(Transformationen)

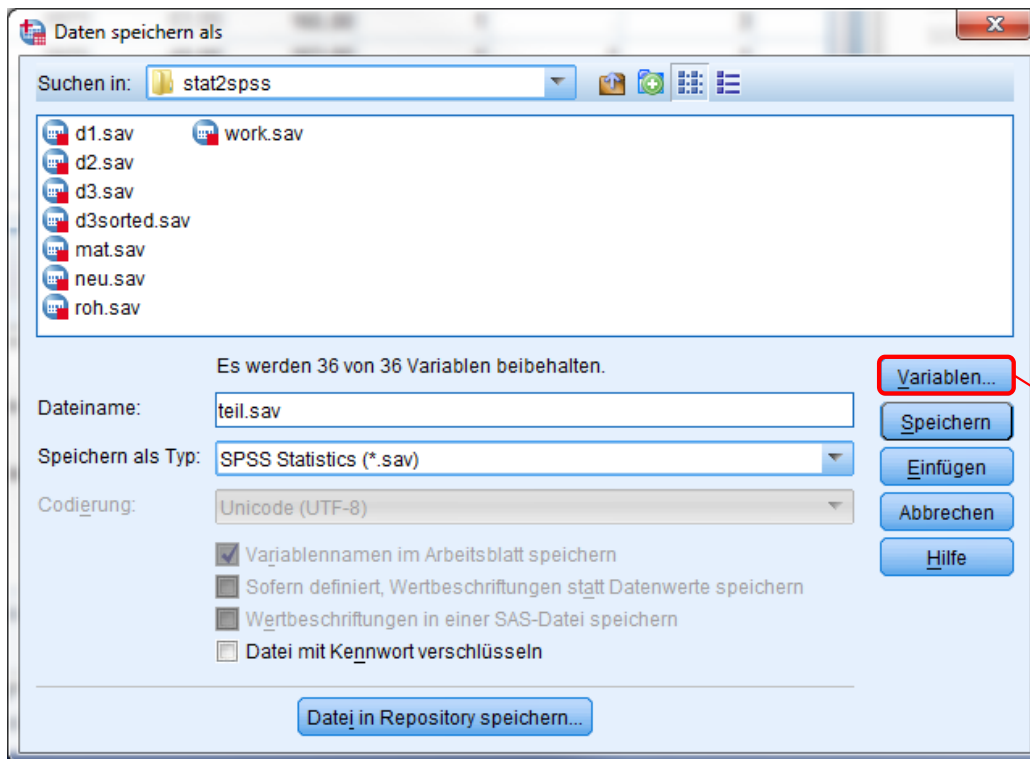
deskr.sps

(Analysen)

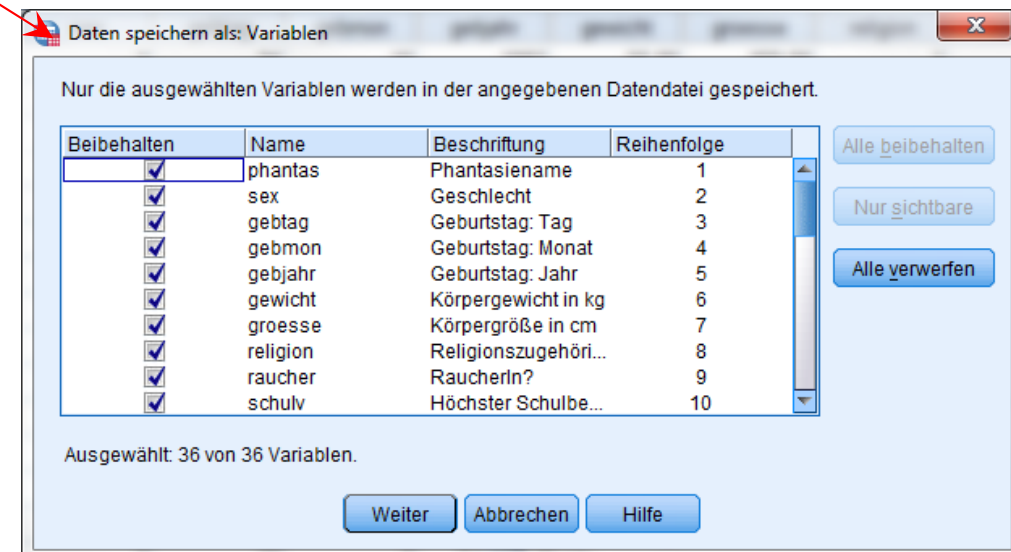
- 1 Workflow bei (größeren) SPSS-Projekten
- 2 **Selektives Speichern von Daten**
- 3 Datenimport
- 4 Datenexport
- 5 Transponieren von Datendateien

SPSS: Selektiv Daten speichern

Will man nicht alle Variablen einer Datendatei in SPSS speichern, kann man unter Datei / Speichern unter... unter der Option (Variablen) auch einzelne Variablen aus der aktuellen Datendatei auswählen, die dann nicht gespeichert werden.



Dazu deaktiviert man bei den betreffenden Variablen in der Spalte „Beibehalten“ das Häkchen. Möglich ist auch, zunächst mittels (Alle verwerfen) alle Variablen zu deaktivieren und danach nur die wieder zu aktivieren, die man speichern will.



SPSS: Selektiv Daten speichern

- Einfacher geht es häufiger über die Syntax mittels **/DROP** (=die folgenden Variablen werden nicht gespeichert) bzw. **/KEEP** (=nur die folgenden Variablen werden gespeichert):
- **Beispiel:** Abspeichern der Daten in einer neuen SPSS-Datei `neu.sav`, die alle Variablen der aktuellen Datei außer `phantas` und `sfach` enthalten soll:

```
SAVE OUTFILE='d:\stat2spss\neu.sav'  
  /DROP = phantas sfach  
  /COMPRESSED.
```

- **Beispiel:** Abspeichern der Daten in einer neuen SPSS-Datei `mat.sav`, nur die Variablen `mat1` bis `mat9` enthalten soll:

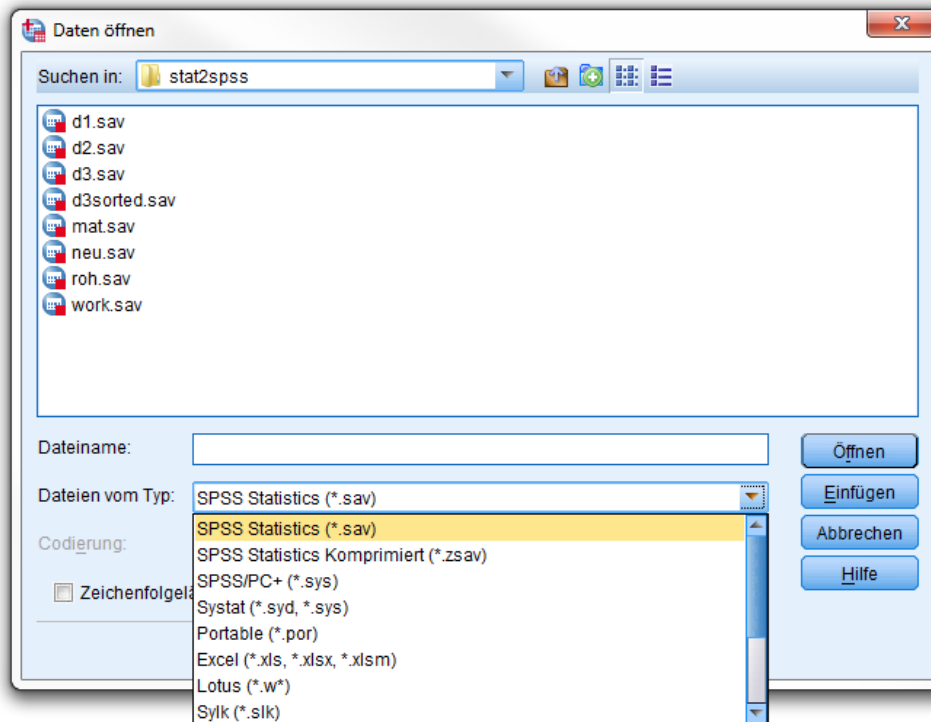
```
SAVE OUTFILE='d:\stat2spss\mat.sav'  
  /KEEP = mat1 mat2 mat3 mat4 mat5 mat6 mat7 mat8 mat9  
  /COMPRESSED.
```

- 1 Workflow bei (größeren) SPSS-Projekten
- 2 Selektives Speichern von Daten
- 3 Datenimport
- 4 Datenexport
- 5 Transponieren von Datendateien

- SPSS bietet die Möglichkeit, Daten, die in bestimmten anderen, gebräuchlichen Formaten vorliegen, in SPSS zu importieren oder SPSS-Datendateien in diese Formate zu exportieren.
 - Dies kann einerseits dem Daten-Austausch zwischen verschiedenen Programmen dienen.
 - Als einfache Alternative dazu besteht bei kleinen Datenmengen oft auch die Möglichkeit, die Daten mittels Copy-und-Paste z.B. zwischen SPSS und Excel hin oder her zu kopieren. Dabei können aber keine Variablennamen mit übernommen werden.
 - Prinzipiell gibt es vor allem zwei Probleme, die immer wieder beim Datenaustausch auftauchen: (1) Ein Programm verwendet bei Zahlen das **Dezimalkomma**, das andere den Dezimalpunkt; (2) **Missing data** werden (intern) von beiden Programmen unterschiedlich codiert.
 - Werden neben den Daten weitere Informationen wie Variablennamen mit übertragen, ergeben sich dabei leicht weitere Probleme bei den Konventionen bzgl. der Variablennamen (z.B. der maximalen Länge, Umlaute etc.).
 - Im Prinzip sollten diese Problem durch die Import- und Export-Filter gelöst sein; eine sorgfältige Kontrolle nach dem Import/Export ist aber **unbedingt erforderlich!**

SPSS: Andere Dateiformate importieren

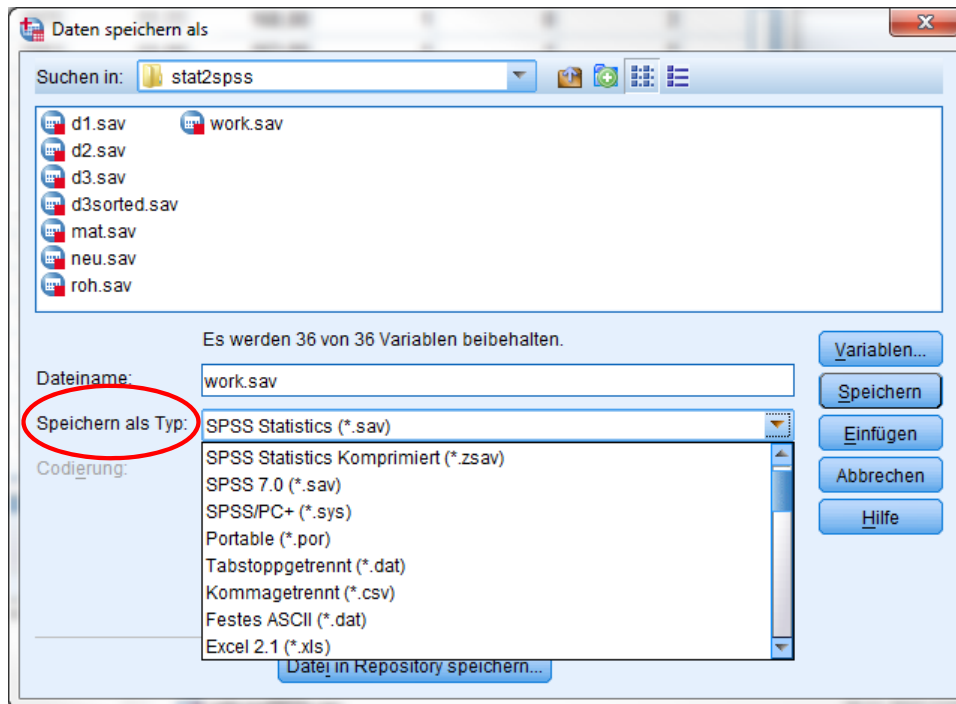
- Unter Datei / öffnen / Daten_... bietet SPSS unter „Dateityp“ folgende **Import-Filter** an:



- Statistikprogramme:
- SPSS incl. Vorgänger (.sav, .sys, .por)
 - Systat (.sys, .syd)
 - SAS (verschiedene)
 - Stata (.dta)
- Tabellenkalkulation:
- Excel (.xls, .xlsx, .xlsm)
- Datenbanken:
- dBase (.dbf)
- Text:
- Festes ASCII (.dat), tabstopgetrennt (.dat), kommasetrennt (.csv) ...

SPSS: Andere Dateiformate exportieren

➤ Unter Datei / Speichern unter... bietet SPSS unter „Typ“ folgende **Export-Filter** an:



➤ In folgende Formate kann u.a. exportiert werden:

- Statistikprogramme: Systat, SAS, Stata
- Tabellenkalkulation: Excel, Lotus
- Datenbanken: dBase
- Text: festes ASCII, kommagetrennt, tabulatorgetrennt

➤ Beim Exportieren ist auch immer auf die Version zu achten (z.B. Excel 2.1, 97-2003, 2007).

➤ Beim Exportieren können auch die zu exportierenden Variablen unter (Variablen) eingeschränkt werden (s.o.)

- Viele andere Programme (z.B. LISREL, EQS oder MPlus zur Strukturgleichungsmodellierung, NORM zum Missing-data handling, BILOG zur Probabilistischen Testtheorie) speichern ihre Daten in speziellen Systemdateien, die SPSS nicht importieren kann.
- Ein Teil dieser Programme kann aber SPSS-Datendateien importieren.
- Es ist also meistens leichter, in SPSS eingegebene Daten in andere Formate zu exportieren als anders herum.

- Eine weitere Möglichkeit besteht darin, die Daten als sog. **fest formatierte ASCII-Daten** einzugeben.
 - Vorteil I: Diese können von (fast) allen Programmen importiert werden (auch von SPSS).
 - Vorteil II: Insbesondere bei Eingabe von numerischen (einstelligen) Datenwerten (z.B. Ratings) ist die Dateneingabe deutlich schneller, weil nicht jede Dateneingabe durch die (Enter)- oder eine andere Taste quittiert werden muss.
 - Vorteil III: Die Daten lassen sich zum Korrekturlesen leichter ausdrucken.
 - Nachteil: Die Daten müssen danach korrekt importiert werden.

- Vorgehen beim Erstellen einer **fest formatierte ASCII-Datendatei**:
 - Man öffnet eine leere Datei in einem **Editor**, z.B. dem Windows standardmäßig beigefügten Microsoft-Editor. (Prinzipiell geht auch ein Textverarbeitungssystem wie Word, aber dann muss man aufpassen, dass man die Daten als ASCII-Daten mit der Option „Nur Text“, d.h. ohne Formatierungsinformationen speichert).
 - Im einfachsten Fall werden jetzt die Daten für eine Person in eine Zeile hintereinander weg geschrieben, so dass in allen Zeilen in jeder Spalte immer die gleichen Informationen stehen. (Es geht auch über mehrere Zeilen.)
 - Wenn nötig, sollte das Dezimalkomma verwendet werden. (Später kann man durch die Ersetzungsfunktion des Editors alle Kommata auch durch Punkte ersetzen.)
 - Missing data werden mittels Leertaste codiert.
 - Falls Variablen unterschiedlich lang sein können, z.B. Körpergewicht, wird immer die maximal mögliche Länge vorgesehen und die Daten dann rechtsbündig eingegeben.
 - Leerzeichen zwischen Blöcken von Daten erleichtern die Orientierung
 - Die Datei sollte mit der Endung `.dat` gespeichert werden.
 - Weitere Informationen z.B. in SPSS-Buch von Diehl & Staufenbiel (2007), Kap. 85 & 86

SPSS: ASCII-Dateneingabe

Missing data

Vp 1 →
Vp 2 →

Vp-Nr. Geschlecht Alter Gewicht 15 Persönl. Items 17 Interessen-Items

Variable	Spalten	Anmerkungen
Vp-Nr	1-3	$n = 230$ benötigt 3 Spalten
Geschlecht	5	1 = „weiblich“, 2 = „männlich“
Alter	7-8	in Jahren
Körpergewicht	10-14	in kg mit einer Nachkommastelle
15 Persönlichkeits-Items	16-30	
17 Interessen-Items	32-48	

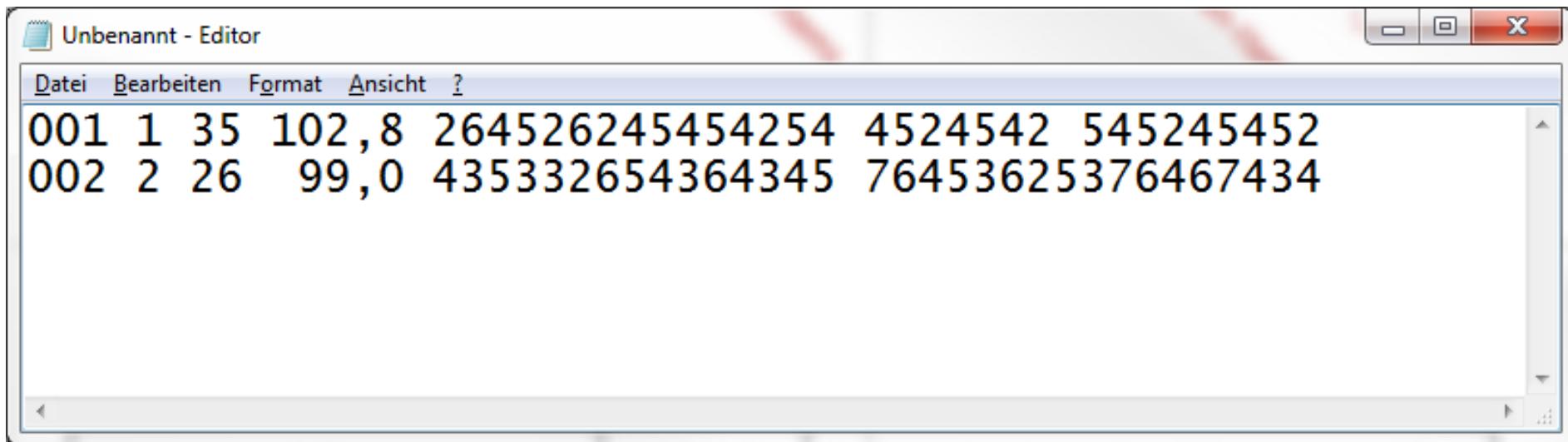
SPSS: ASCII-Dateneingabe

Variable	Spalten	Anmerkungen	SPSS-Variablenname
Vpn-Nr	1-3	N=230	vpnr
Geschlecht	5	1=„weiblich“, 2=„männlich“	sex
Alter	7-8		alter
Körpergewicht	10-14	in kg mit einer Nachkommastelle	gewicht
15 Persönlichkeits-Items	16-30		pers1 bis pers15
17 Interessen-Items	32-48		inter1 bis inter17

Die ASCII-Daten können unter Datei /Öffnen/Daten... und der Angabe des Datei-Typs „Text (*.txt, *.dat)“ menügesteuert importiert werden oder (einfacher) per Syntax wie folgt (**RECORDS=1** bedeutet, dass alle Daten in einer Zeile stehen):

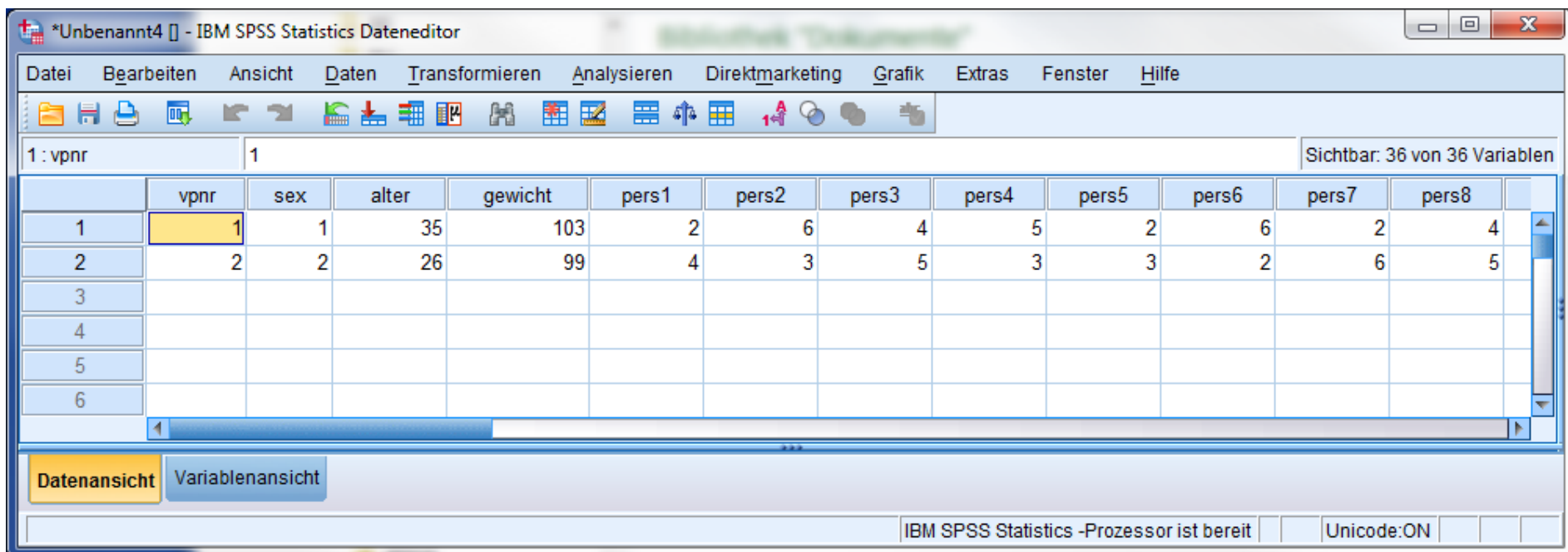
```
SET BLANKS=SYSMIS UNDEFINED=WARN.  
DATA LIST FILE='d:\stat2spss\x.dat' FIXED RECORDS=1  
  /1 vpng 1-3 sex 5 alter 7-8 gewicht 10-14  
    pers1 TO pers15 16-30 inter1 TO inter17 32-48.  
EXECUTE.
```

SPSS: ASCII-Dateneingabe



Unbenannt - Editor

```
Datei Bearbeiten Format Ansicht ?
001 1 35 102,8 264526245454254 4524542 545245452
002 2 26 99,0 435332654364345 76453625376467434
```



*Unbenannt4 [] - IBM SPSS Statistics Dateneditor

Datei Bearbeiten Ansicht Daten Transformieren Analysieren Direktmarketing Grafik Extras Fenster Hilfe

1 : vpnr 1 Sichtbar: 36 von 36 Variablen

	vpnr	sex	alter	gewicht	pers1	pers2	pers3	pers4	pers5	pers6	pers7	pers8
1	1	1	35	103	2	6	4	5	2	6	2	4
2	2	2	26	99	4	3	5	3	3	2	6	5
3												
4												
5												
6												

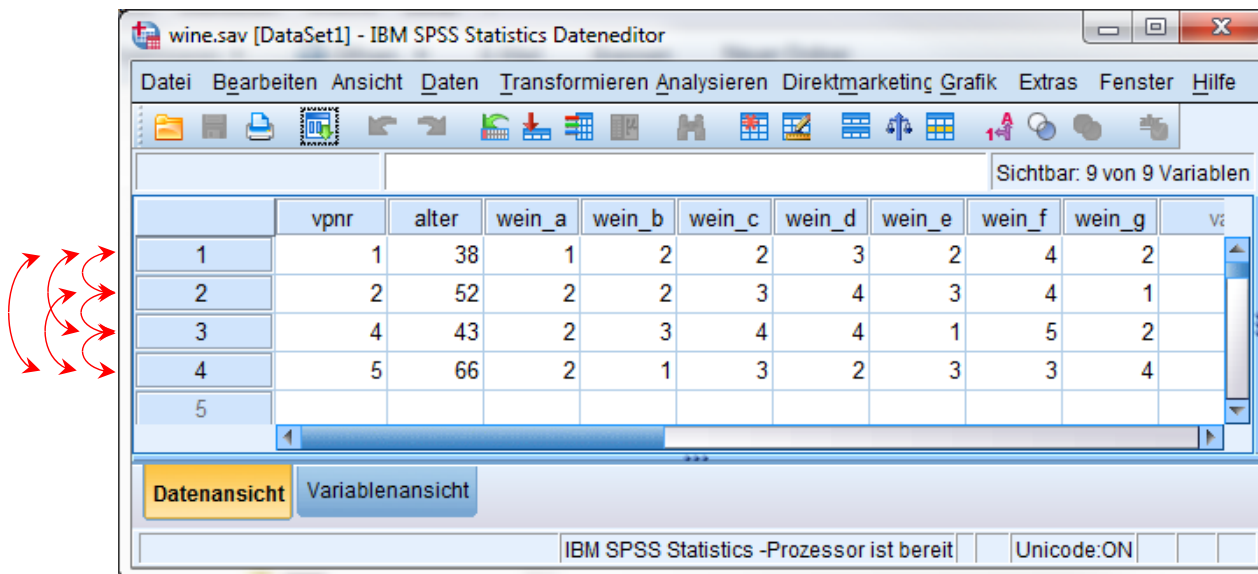
Datenansicht Variablenansicht

IBM SPSS Statistics -Prozessor ist bereit Unicode:ON

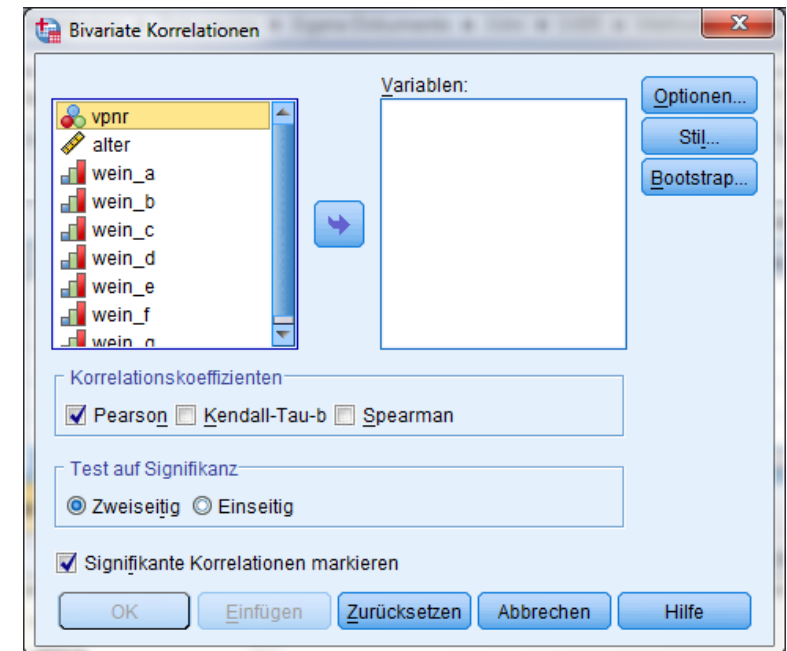
- 1 Workflow bei (größeren) SPSS-Projekten
- 2 Selektives Speichern von Daten
- 3 Datenimport
- 4 Datenexport
- 5 Transponieren von Datendateien

Transponieren einer Datendatei

- **Beispiel:** In einer Blindverkostung haben vier Weinkenner sieben Weine „A“ bis „G“ bezüglich ihrer Qualität auf einer Notenskala von 1 bis 6 beurteilt. Die Daten sind in SPSS wie unten gezeigt eingegeben. Wir wollen prüfen, ob die Weinkenner übereinstimmend die Qualität der Weine einstufen und zwischen ihren Urteilen Rangkorrelationen (zwischen den Personen über die Weine hinweg) bestimmen.
- Dabei besteht das Problem, dass die Anordnung der Daten eine solche Analyse nicht direkt ermöglicht, weil wir ja Korrelationen zwischen Personen und nicht zwischen Variablen bestimmen wollen.



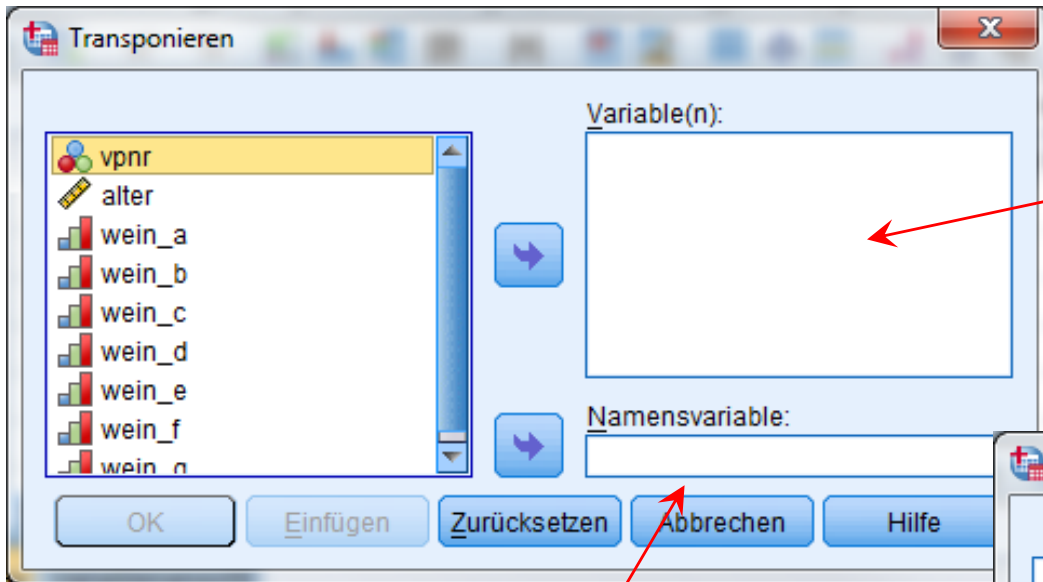
	vpnr	alter	wein_a	wein_b	wein_c	wein_d	wein_e	wein_f	wein_g	vi
1	1	38	1	2	2	3	2	4	2	
2	2	52	2	2	3	4	3	4	1	
3	4	43	2	3	4	4	1	5	2	
4	5	66	2	1	3	2	3	3	4	
5										



- Wir müssen daher zunächst Zeilen und Spalten (Personen und Variablen) in der Datendatei vertauschen.

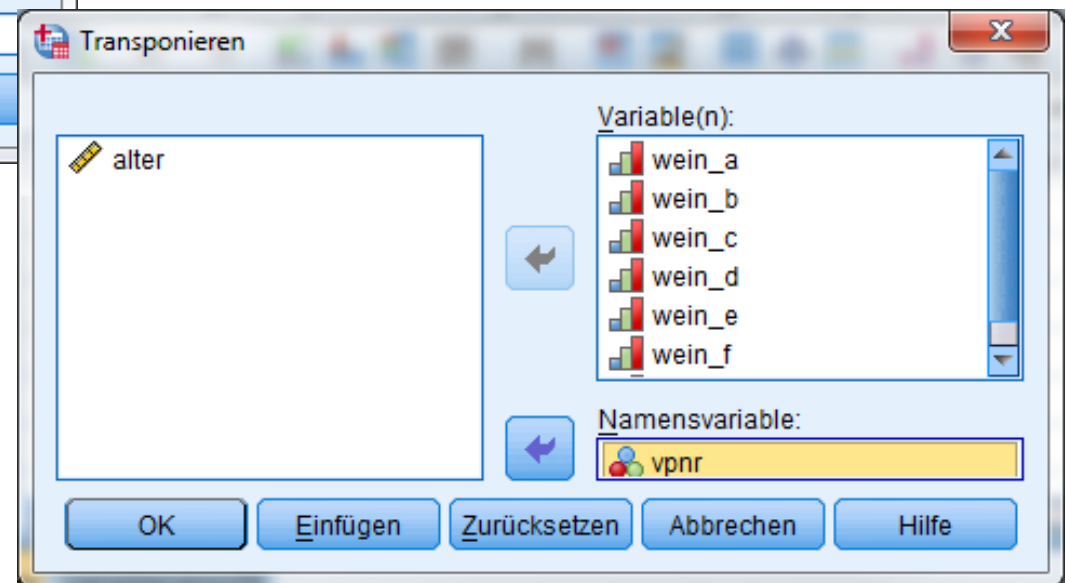
Transponieren einer Datendatei

Das Vertauschen von Zeilen und Spalten einer Datendatei in SPSS (transponieren, kippen) in kann man unter Daten/Transponieren... vornehmen.



Angabe aller Variablen, die transponiert werden sollen. (Die nicht angegebenen Variablen werden nicht mit in die neu erstellte Datendatei aufgenommen).

Hier kann eine Namensvariable angegeben werden, deren Inhalte dann als Variablennamen bei der Transponierung verwendet werden.



Transponieren einer Datendatei

	vpnr	alter	wein_a	wein_b	wein_c	wein_d	wein_e	wein_f	wein_g	va
1	1	38	1	2	2	3	2	4	2	
2	2	52	2	2	3	4	3	4	1	
3	4	43	2	3	4	4	1	5	2	
4	5	66	2	1	3	2	3	3	4	
5										

Durch das Transponieren wird eine neue Datendatei erzeugt.



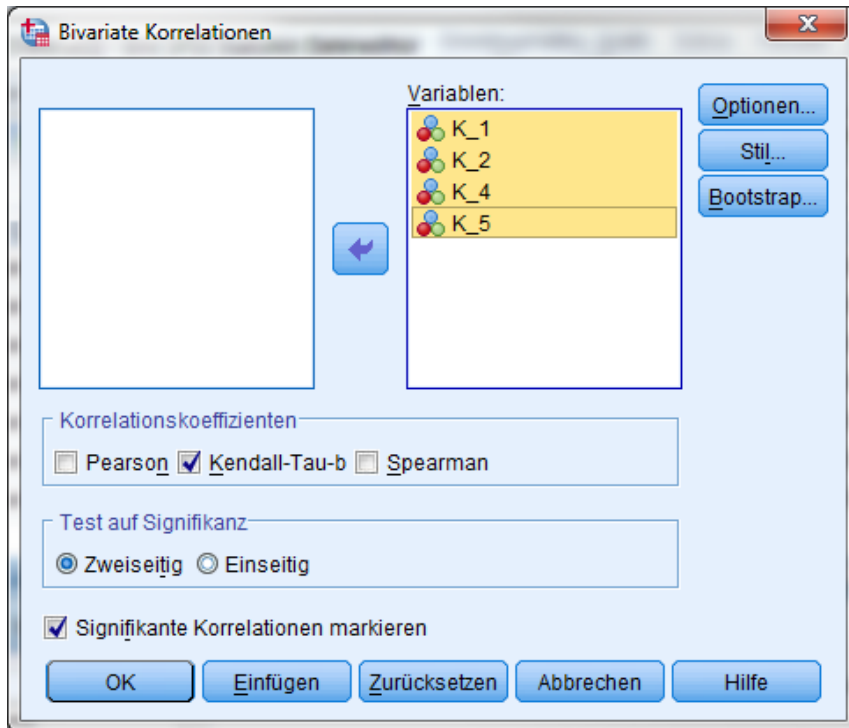
Da die Namensvariable in diesem Fall nur Zahlen enthält, stellt SPSS diesen den String **K_** voran.

Die „alten“ Variablennamen werden immer in einer String-Variablen mit dem Namen **CASE_LBL** gespeichert.

	CASE_LBL	K_1	K_2	K_4	K_5	var
1	wein_a	1,00	2,00	2,00	2,00	
2	wein_b	2,00	2,00	3,00	1,00	
3	wein_c	2,00	3,00	4,00	3,00	
4	wein_d	3,00	4,00	4,00	2,00	
5	wein_e	2,00	3,00	1,00	3,00	
6	wein_f	4,00	4,00	5,00	3,00	
7	wein_g	2,00	1,00	2,00	4,00	
8						

Transponieren einer Datendatei

Nun lässt sich die gewünschte Analyse durchführen (ggf. wird man die Variablen noch eingängiger umbenennen wollen, was hier nicht geschehen ist), in dem man die Rangkorrelation – hier Kendall τ_b – unter Analysieren/Korrelation/Bivariat... aufruft (besser wäre γ):



Man sieht, dass die Beurteiler 1, 2 und 4 relativ übereinstimmend urteilen (alle $\tau_b > .54$), die Urteile von 5 hingegen mit denen der Kollegen in einem hohen Zusammenhang stehen.

Korrelationen

			K_1	K_2	K_4	K_5
Kendall-Tau-b	K_1	Korrelationskoeffizient	1,000	,669	,652	,125
		Sig. (2-seitig)	.	,057	,060	,725
		N	7	7	7	7
K_2	K_2	Korrelationskoeffizient	,669	1,000	,541	-,057
		Sig. (2-seitig)	,057	.	,112	,870
		N	7	7	7	7
K_4	K_4	Korrelationskoeffizient	,652	,541	1,000	-,056
		Sig. (2-seitig)	,060	,112	.	,871
		N	7	7	7	7
K_5	K_5	Korrelationskoeffizient	,125	-,057	-,056	1,000
		Sig. (2-seitig)	,725	,870	,871	.
		N	7	7	7	7

Aufgrund der geringen Stichprobenumfänge (hier $n = 7$ Weine) wird keine der Korrelationen bei $\alpha = 0.05$ statistisch signifikant.