

Dr. Matthias Rudolf
Modul M3: Multivariate Statistik

Aufgaben und Lösungshinweise zum Computerseminar ERA:

Einfache lineare Regressionsanalyse

Laden Sie Datei mit dem Datensatz, den Sie im ersten Computerseminar eingegeben und gespeichert haben (**Datei** → **Öffnen** → **Daten...**).

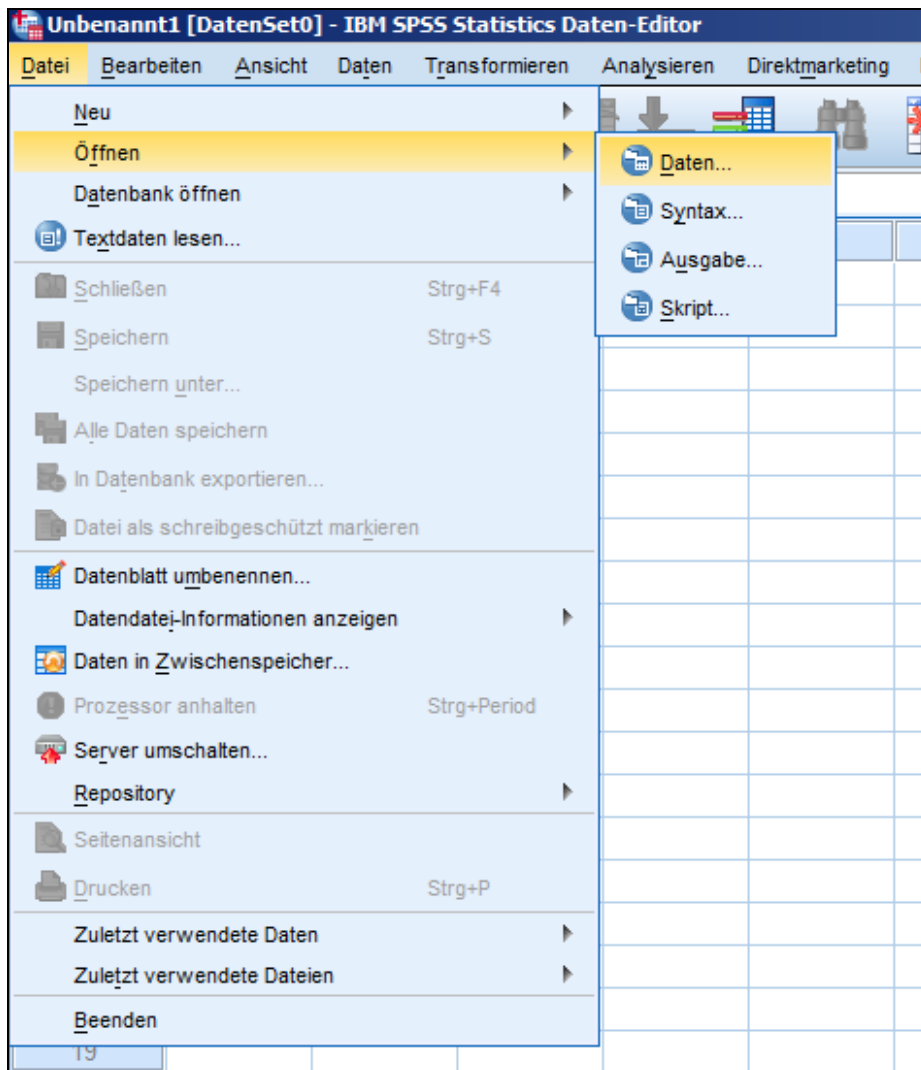


Bild 1

Im danach erscheinenden Fenster (Bild 2) ist der Dateiname einschließlich des Pfades der Datendatei einzugeben, die Sie während des letzten Seminars gespeichert haben.

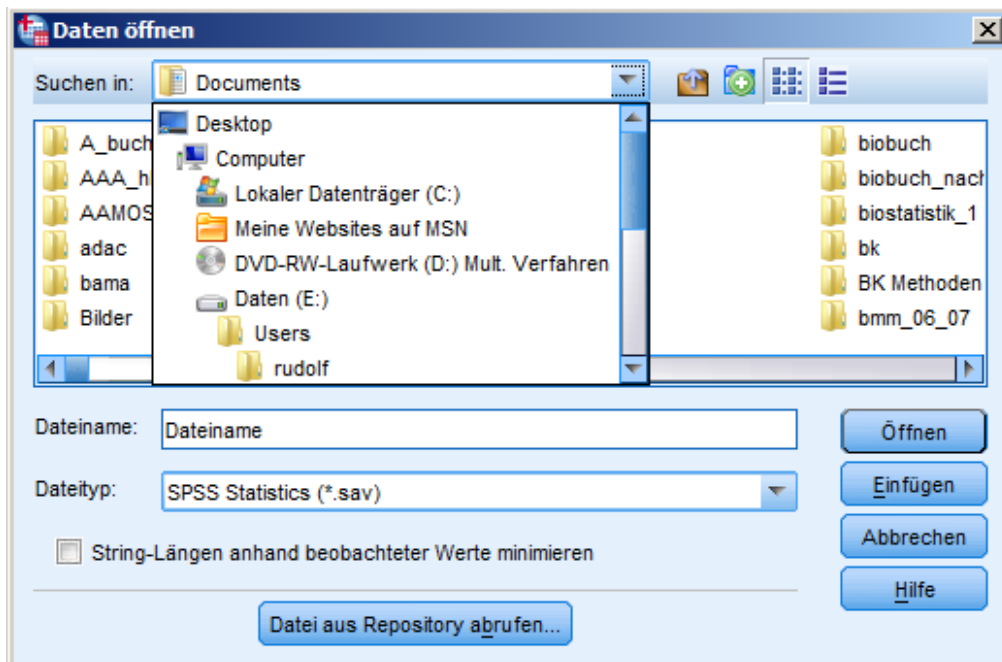


Bild 2

Die Daten sind in der folgenden Tabelle noch einmal zusammengefasst.

vp	selbst	swe	intell	aengst	geschl
1	2	23	100	24	1 (für ♂)
2	8	26	102	30	0 (für ♀)
3	5	25	94	46	0
4	8	32	88	5	1
5	3	26	104	32	1
6	17	31	105	2	0
7	18	43	102	5	1
8	3	21	95	27	1
9	10	31	98	21	0
10	12	32	112	15	1
11	10	26	97	8	0
12	12	35	98	12	0
13	11	29	105	30	0
14	8	25	105	22	1
15	15	29	94	35	0
16	10	33	109	16	1

Aufgaben:

- (1) Es soll mittels Korrelationsanalyse betrachtet werden, welche der Prädiktoren einen signifikanten Zusammenhang mit dem Kriterium aufweisen. Die Zusammenhänge sind grafisch zu veranschaulichen.
- (2) Anschließend sind für alle Prädiktoren einfache lineare Regressionen zu berechnen. Die Ergebnisse dieser Regressionsanalysen sind zu interpretieren, die Möglichkeiten der Vorhersage der Selbstsicherheit aus den Werten der Prädiktoren sind zu beurteilen.
- (3) Weiter soll ermittelt werden, ob in dieser Stichprobe Unterschiede zwischen Männern und Frauen auftreten, d.h. welche Prädiktoren bei Männern einen anderen Einfluss auf das Kriterium haben als bei Frauen. Dazu wurde die Variable Geschlecht (*geschl*) erhoben.

Mit Ausnahme der Variable Geschlecht sind alle Variablen intervallskaliert.

Korrelationsanalyse

Zu Beginn soll der Zusammenhang zwischen Selbstsicherheit und den einzelnen Prädiktoren **grafisch** veranschaulicht werden. Aus den Grafiken lassen sich erste Schätzungen über die Höhe der Korrelationskoeffizienten ableiten.

Von der Datenansicht ausgehend, klicken Sie auf **Grafik (Menüleiste)** → **Diagrammerstellung**. Es öffnet sich ein gleichnamiges Dialogfeld (Bild 3)¹.

Wählen sie hier **Streu-/Punktdiagramm** [1], danach **Einfaches Streudiagramm** [2] und ziehen sie dieses mit gedrückter linker Maustaste in den Vorschaubereich [3].

Als y-Variable ist jeweils Selbstsicherheit einzutragen, als x-Variable für die erste Grafik der Prädiktor Selbstwirksamkeitserwartung (swe). Dazu müssen diese Variablen ebenfalls aus dem Variablenauswahlbereich [4] durch Ziehen in die Diagrammvorschau übertragen werden.

Mit **OK** wird die Erstellung der Grafik gestartet.

¹ Das eventuell erscheinende kleinere Fenster Diagrammerstellung können Sie mittels OK schließen.

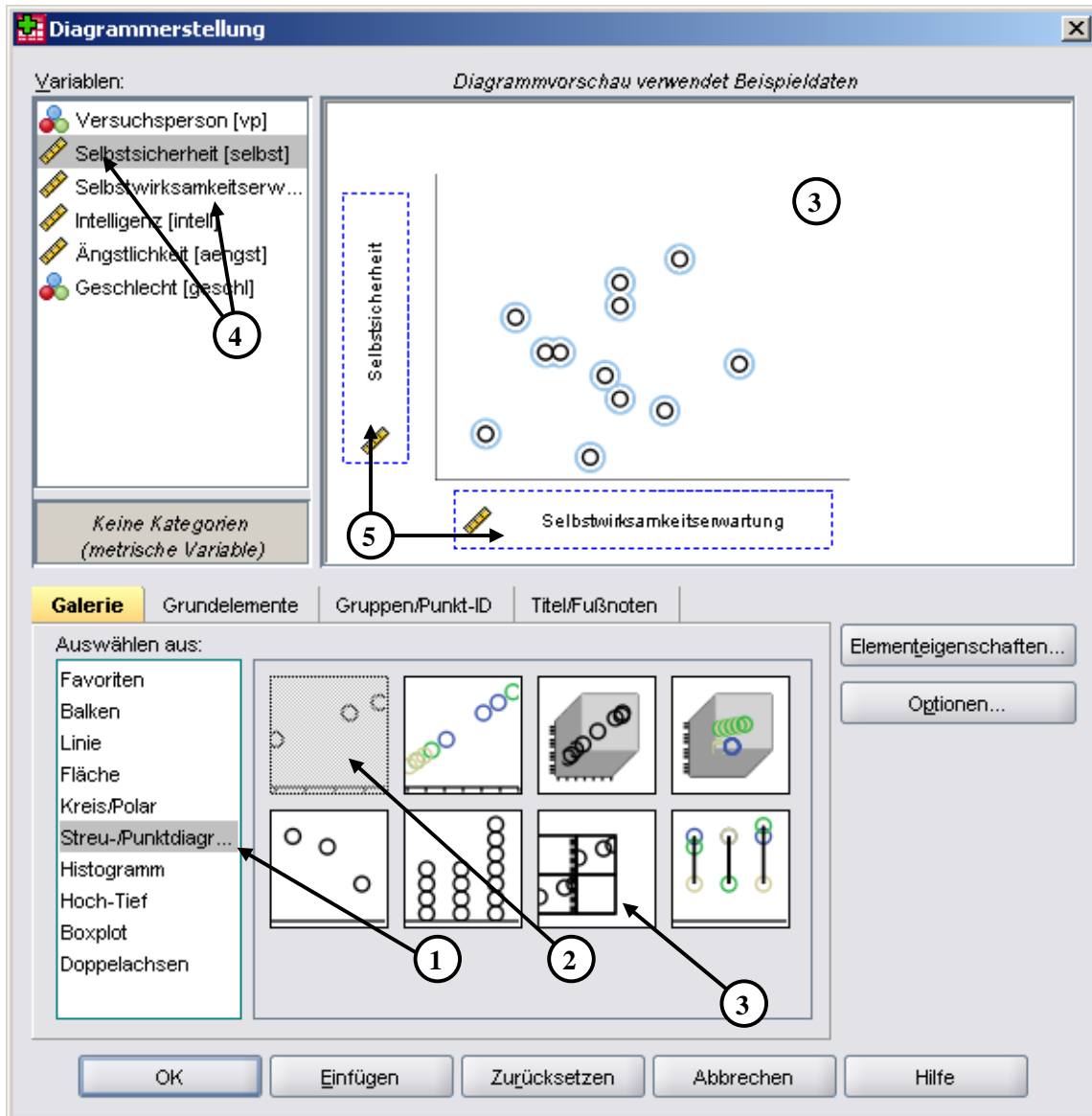


Bild 3

Das folgende Bild (Bild 4) zeigt das Diagramm mit der Variablen Selbstwirksamkeitserwartung (swe) auf der Abszisse. Jeder Teilnehmer ist durch einen Punkt im Diagramm repräsentiert [1].

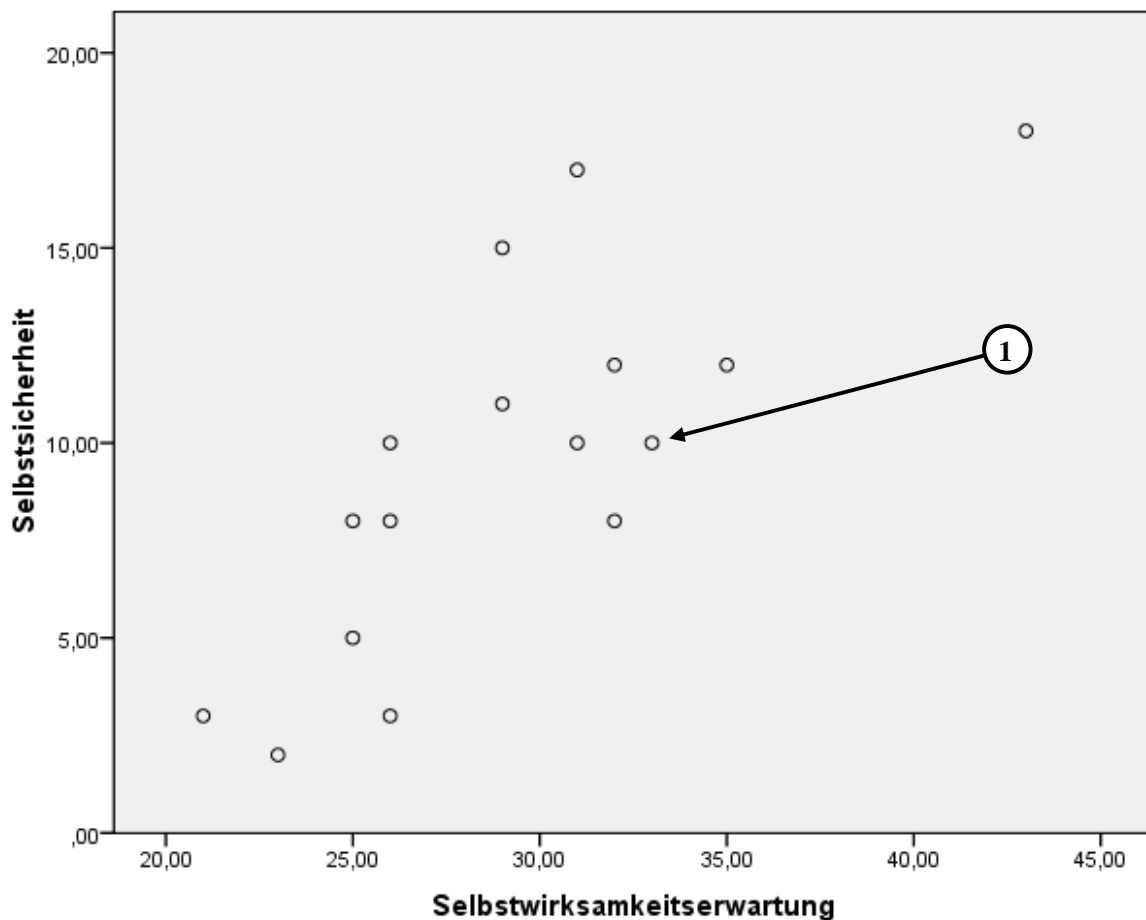


Bild 4

Aufgabe: Notieren Sie, welchen Wert des Produkt-Moment-Korrelationskoeffizienten Sie für diese Daten erwarten und vergleichen Sie Ihren geschätzten Wert später mit dem Ergebnis der Korrelationsanalyse.

Zusatzaufgabe: Verbessern Sie das Aussehen der Grafik unter Verwendung des Diagrammeditors.

Mittels linearer Korrelationsanalyse soll nun ermittelt werden, ob ein signifikanter linearer Zusammenhang zwischen Selbstsicherheit und Selbstwirksamkeitserwartung existiert.

Dazu klicken Sie ausgehend von der Datenansicht auf **Analysieren (Menüleiste)** → **Korrelation** → **bivariat**. Es öffnet sich das Fenster zur bivariaten Korrelation (Bild 5).

Hinweis: Weitere Möglichkeiten an dieser Stelle sind die Berechnung von partiellen Korrelationen, sowie von Distanzen. Multiple Korrelationen können im Zusammenhang der Regressionsanalyse berechnet werden.

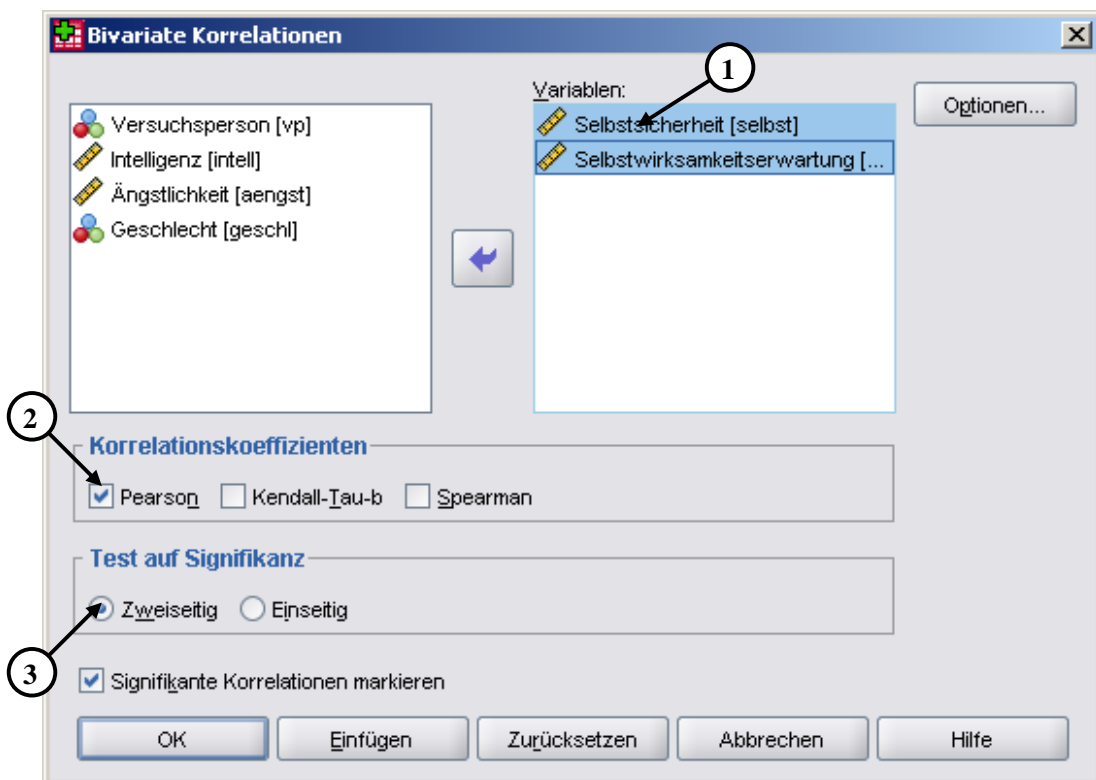


Bild 5

Die in die Analyse einzubeziehenden Variablen sind in das *Variablenfenster* [1] zu übernehmen. In unserem Beispiel stehen im Variablenfenster das Kriterium Selbstsicherheit und der Prädiktor Selbstwirksamkeitserwartung.

Es soll der *Produkt-Moment-Korrelationskoeffizient nach Pearson* [2] mit *zweiseitigem Signifikanztest* [3] berechnet werden.

Die *Optionen* können - nach vorgenommener Überprüfung - unverändert bleiben.

OK startet die Berechnung.

Hinweis: Neben dem Produkt-Moment-Korrelationskoeffizienten (Pearson) können auch die Rangkorrelationskoeffizienten nach Spearman bzw. Kendall (Tau-b) berechnet werden.

Ergebnisausgabe

Die berechnete Korrelationsmatrix (Bild 6) schließt sich nun an unsere bisherigen Ergebnisse im Ausgabefenster an.

		Selbtsicherheit	Selbstwirksamkeitserwartung
Selbtsicherheit	Korrelation nach Pearson	1,000	,776**
	Signifikanz (2-seitig)		,000
	N	16	16
Selbstwirksamkeitserwartung	Korrelation nach Pearson	,776**	1,000
	Signifikanz (2-seitig)	,000	
	N	16	16

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Bild 6

In der ersten Zeile jeder Zelle [1] ist der Korrelationskoeffizient angegeben. Ein Sternchen hinter dem Koeffizienten würde eine signifikante Korrelation auf dem 5% Niveau anzeigen, zwei Sternchen weisen auf eine signifikante Korrelation auf dem 1% Niveau hin.

In der zweiten Zeile ist der berechnete p-Wert angegeben [2]. Durch einen zweifachen Doppelklick darauf erhält man den genauen Wert von $p=0.000408$, welcher trotz Darstellung nicht genau 0.000 beträgt.

In der dritten Zeile findet man die Stichprobengröße.

In unserem Beispiel ergibt sich zwischen Selbstsicherheit und Selbstwirksamkeitserwartung ein Korrelationskoeffizient von $r=0.776$, der signifikant ($p<0.01$) von Null verschieden ist.

Aufgabe: Erstellen Sie grafische Darstellungen für die Zusammenhänge zwischen Selbstsicherheit und Ängstlichkeit sowie zwischen Selbstsicherheit und Intelligenz. Schätzen Sie auf Grund der jeweiligen Grafik, welche Korrelation zwischen den Variablen besteht. Berechnen Sie danach die Korrelationen und vergleichen Sie Ihre Schätzung mit dem Ergebnis.

Zusatzaufgabe: Stellen Sie die Beziehungen der Variablen Selbstsicherheit, Selbstwirksamkeitserwartung, Ängstlichkeit und Intelligenz in einer Streudiagramm-Matrix (Bild 3 [3]) dar, die Sie mit dem Diagramm-Editor ansprechend gestalten.

Regressionsanalyse

Nachdem wir uns einen umfassenden Überblick über die Zusammenhänge der einzelnen Prädiktoren mit dem Kriterium verschafft haben, sind einfache lineare Regressionen zu berechnen.

Ausgehend von der Datenansicht klicken Sie unter **Analysieren** → **Regression** → **Linear**. Es erscheint folgendes Dialogfenster (Bild 7):

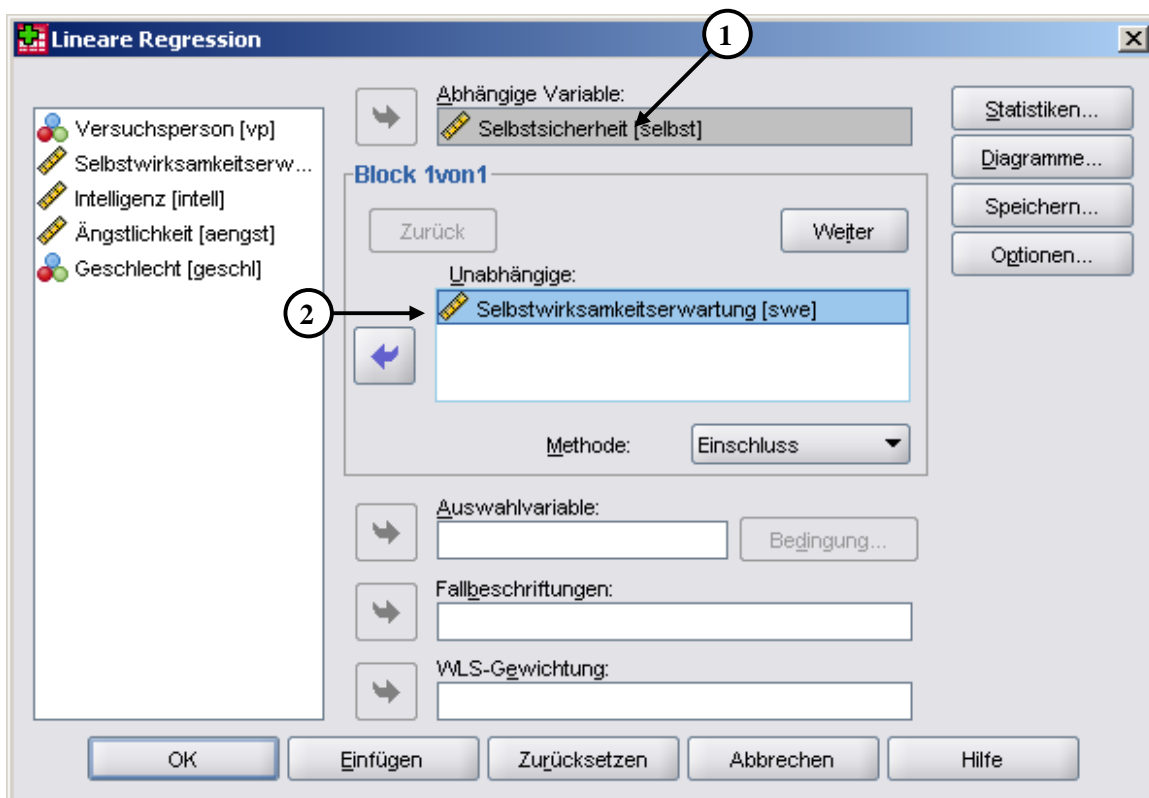


Bild 7

In unserem Beispiel ist die Kriteriumsvariable Selbstsicherheit in das *Feld für die Abhängige Variable* [1] zu verschieben und der Prädiktor Selbstwirksamkeitserwartung in das *Feld für die Unabhängige Variable* [2].

Die Berechnung wird mit **OK** gestartet.

Ergebnisausgabe

Die nächsten drei Abbildungen beinhalten die Ergebnisse der Analyse. Zunächst beschäftigen wir uns mit der **Tabelle Koeffizienten** (Bild 8).

Modell		Koeffizienten ^a				Signifikanz
		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten		
		B	Standardfehler	Beta	T	
1	(Konstante)	-10,647	4,443		-2,397	,031
	Selbstwirksamkeitserwartung	,690	,150	,776	4,606	,000

a. Abhängige Variable: Selbstsicherheit

Bild 8

Der Koeffizient b_0 [1] gibt den Nulldurchgang der Regressionsgeraden an, der Koeffizient b_1 [2] den Anstieg der Geraden.

Für die Regression zwischen Selbstsicherheit und Selbstwirksamkeitserwartung ergibt sich demnach folgende Regressionsgleichung:

$$y = -10.6 + .69 \cdot x \text{ (bzw. Selbstsicherheit} = -10.6 + .69 \cdot \text{Selbstwirksamkeit).}$$

Mit der Regressionsgeraden können anhand bekannter x-Werte die dazugehörigen y-Werte geschätzt werden.

Dividiert man die Regressionskoeffizienten b_0 und b_1 durch ihre Standardfehler, erhält man die t-Werte [3] für die Nullhypothesen $b_0=0$ und $b_1=0$ (zweiseitige Tests). Die letzte Spalte gibt die jeweiligen p-Werte an.

Werden die Werte vor der Berechnung der Gleichung z- standardisiert, erhält man anstelle von b_1 das standardisierte β -Gewicht (Beta [4]). Für den Nulldurchgang ergibt sich dann der Wert 0.

Bei der einfachen linearen Regression ist das β -Gewicht von 0.776 identisch mit dem Korrelationskoeffizienten $r=0.776$ (vgl. Bild 6).

Die folgende Abbildung zeigt die Tabelle **Modellzusammenfassung** (Bild 9).

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,776 ^a	,602	,574	3,11624

a. Einflußvariablen : (Konstante), Selbstwirksamkeitserwartung

Bild 9

R ist der multiple Korrelationskoeffizient [1].

Da bei der einfachen Regression lediglich ein Prädiktor in die Berechnung eingeht, ist R identisch mit dem Produkt-Moment-Korrelationskoeffizienten r (im Beispiel $r=0.776$ zwischen Selbstsicherheit und Selbstwirksamkeitserwartung, vgl. Bild 6).

Quadriert man r bzw. R, erhält man das multiple Bestimmtheitsmaß R^2 [2]. Dieses gibt den Anteil der Varianz des Kriteriums Selbstsicherheit an, der durch die Variable Selbstwirksamkeitserwartung aufgeklärt wird. In unserem Beispiel werden 60% der Varianz der Variable Selbstsicherheit durch die Variable Selbstwirksamkeitserwartung aufgeklärt.

Hinweis: Auf das korrigierte Bestimmtheitsmaß und auf den Standardfehler des Schätzers wird im kommenden Praktikum eingegangen (vgl. Rudolf & Müller (2012) S. 47 und S.55).

Die Beurteilung der globalen Güte der Regression erfolgt mithilfe der Tabelle **ANOVA** (Bild 10).

ANOVA^b

Modell		Quadratsumme	df	Mittel der Quadratrate	F	Signifikanz
1	Regression	206,047	1	206,047	21,218	,000 ^a
	Residuen	135,953	14	9,711		
	Gesamt	342,000	15			

a. Einflußvariablen : (Konstante), Selbstwirksamkeitserwartung

b. Abhängige Variable: Selbstsicherheit

Bild 10

Der F-Test (vgl. Rudolf & Müller (2012) S.46/47) prüft das Bestimmtheitsmaß, indem er angibt, ob der Anteil der erklärten Varianz signifikant von Null verschieden ist.

Die Quadratsumme der erklärten Varianz [1] beträgt in unserem Beispiel 206,05, die Quadratsumme der nicht erklärten Varianz [2] oder auch Quadratsumme der Residuen beträgt 135,96. Die Gesamtvarianz ist die Summe beider Varianzanteile.

Der F-Test berechnet nun: $F = (206,047/df) / (135,953/df) = 21,218$ [3]. Der Zusammenhang zwischen Selbstsicherheit und Selbstwirksamkeitserwartung ist signifikant [4], wie auch schon die Korrelationsanalyse ergab.

Aufgabe: Berechnen Sie erneut eine einfache Regressionsanalyse, jedoch mit Variablen, die Sie vorher z- standardisieren (**Analysieren** → **Deskriptive Statistiken** → **Deskriptive Statistik** → **Standardisierte Werte als Variable speichern**). Vergleichen Sie diese mit den zuvor berechneten Ergebnissen und interpretieren Sie das Ergebnis.

Zusatzaufgabe: Erstellen Sie erneut ein Streudiagramm der Variablen Selbstsicherheit und Selbstwirksamkeitserwartung. Finden Sie heraus, wie Sie im Diagrammeditor eine Regressionsgerade grafisch in das Diagramm einbringen können und erzeugen Sie eine insgesamt optisch ansprechende Grafik.

Arbeit mit der SPSS-Syntax

Mit der beschriebenen Menü-gestützten Arbeitsweise ist ein effektiver und schneller Einstieg in SPSS möglich. Die in SPSS realisierten Möglichkeiten auch komplexer Verfahren können komfortabel erfasst und ohne Probleme angewendet werden.

Gelegentlich ist es jedoch in umfangreichen Datenanalysen notwendig, die gleichen, oft sehr komplexen Analysen wiederholt mit verschiedenen Variablen oder mit unterschiedlichen Probanden durchzuführen. Ein Beispiel kann die sehr umfangreiche Auswertung (beschreibende Statistiken, Faktorenanalysen, varianz- und regressionsanalytische Tests) eines Fragebogens sein, bei dem über mehrere Jahre zu unterschiedlichen Zeiten Daten verschiedener Probanden nach exakt dem gleichen Vorgehen auszuwerten sind. Spätestens hier stößt die Effektivität der bisher beschriebenen Vorgehensweise aus wenigstens zwei Gründen an Grenzen: Einerseits ist es sehr zeitaufwändig, die gleichen Einstellungen in den unterschiedlichen Dialogfenstern immer wieder vornehmen zu müssen, um die analogen Analysen durchführen zu können. Andererseits sind damit Fehlermöglichkeiten gegeben, weil natürlich bei wiederholt notwendigen Einstellungen in den Dialogfenstern die Gefahr fehlerhafter Einstellungen nicht auszuschließen ist.

Aus diesen Gründen ist es in Anwendungen mit wiederholt durchzuführenden Analysen oft sinnvoll, mit der SPSS-Syntax zu arbeiten. Der Syntax-Editor von SPSS ist ein Textfenster, in dem man Befehle direkt eingeben oder in dem man die in den Dialogfenstern getroffenen Einstellungen in Befehlsform speichern kann. Die erzeugten Befehlsfolgen kann man effektiv mit unterschiedlichen Probandengruppen wiederholt abarbeiten, wobei auch zum Beispiel die verwendeten Variablen sehr einfach ausgetauscht werden können („Suchen“ und „Ersetzen“).

Das grundsätzliche Prinzip der Arbeit mit der SPSS-Syntax soll an dem einfachen Beispiel der Korrelationsanalyse erläutert werden. Dabei soll davon ausgegangen werden, dass die Daten im Hauptspeicher vorliegen. Für die Korrelationsanalyse soll eine Syntax-Datei erstellt werden, die die wiederholte Durchführung exakt der gleichen Befehlsfolge für unterschiedliche Probanden oder für andere Variablen effektiv ermöglicht.

Zunächst sind die in Bild 5 beschriebenen Einstellungen über die Dialogfenster vorzunehmen. Danach ist aber nicht die Analyse mit OK (Bild 14 [1]) zu starten, sondern es ist Einfügen [2] anzuklicken. Daraufhin öffnet sich das Fenster des Syntax-Editors, das in Bild 15 dargestellt ist.

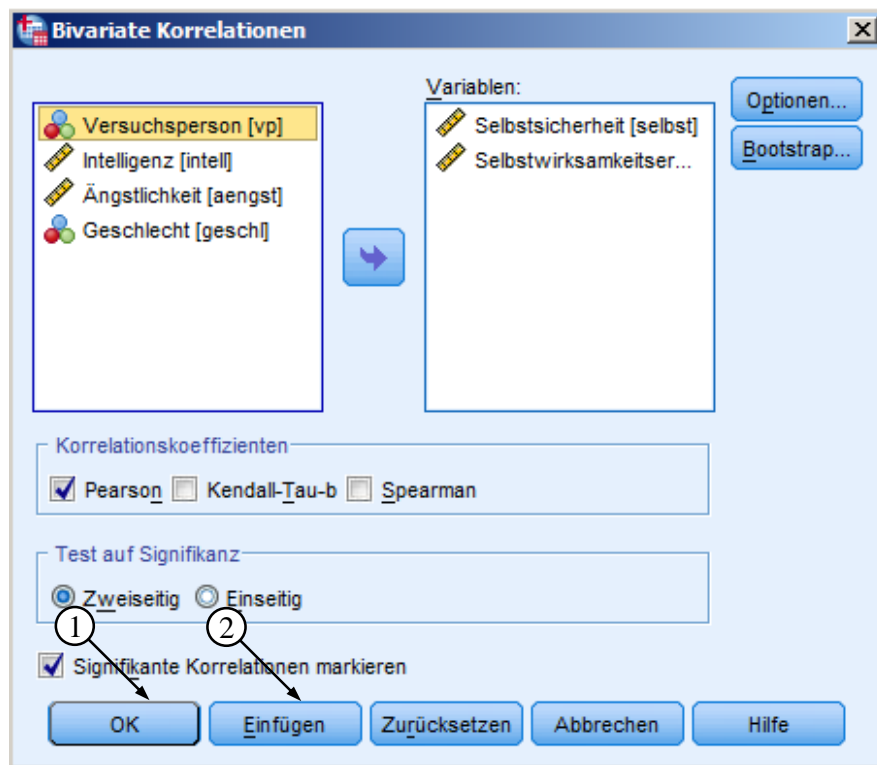


Bild 14

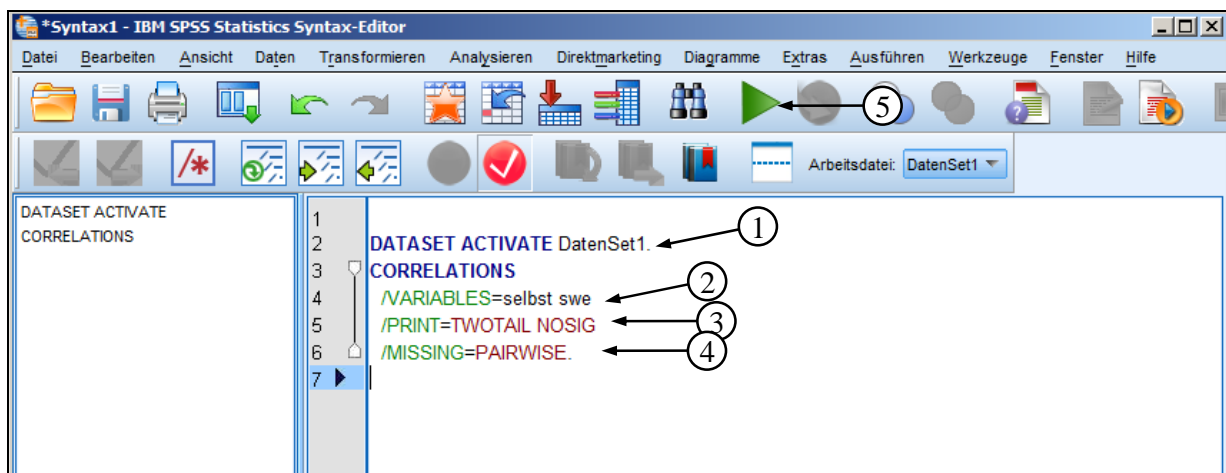


Bild 15

In dem in Bild dargestellten Fenster sind die bisherigen Eingaben in Befehlsform enthalten. Das aktive DatenSet1 soll verwendet werden [1]. In die Korrelationsanalyse (CORRELATIONS) sollen die Variablen selbst und swe einbezogen werden [2]. In der folgenden Zeile [3] ist dargestellt, dass ein zweiseitiger (TWOTAIL) Signifikanztest durchgeführt werden soll, wobei die signifikanten Korrelationen zu markieren sind (NOSIG). Bei fehlenden Werten soll die entsprechende Versuchsperson bei der Berechnung der

paarweisen Korrelationen ausgeschlossen werden, in denen die betreffende Variable vorkommt (PAIRWISE). Die Analyse kann nun durch Anklicken des Pfeils [5] gestartet werden. Die Ergebnisse entsprechen natürlich exakt den Ergebnissen aus Bild 6.

Nun sollen in Bild 7 (Regressionsanalyse) abgebildeten Einstellungen erzeugt und ebenfalls in das Syntax-Fenster übernommen werden. Die Einstellungen sind wie in den Erläuterungen zu Bild 7 beschrieben vorzunehmen, anschließend ist in Bild 7 statt OK Einfügen anzuklicken. Danach ergibt sich das in Bild 16 dargestellte Syntaxfenster.

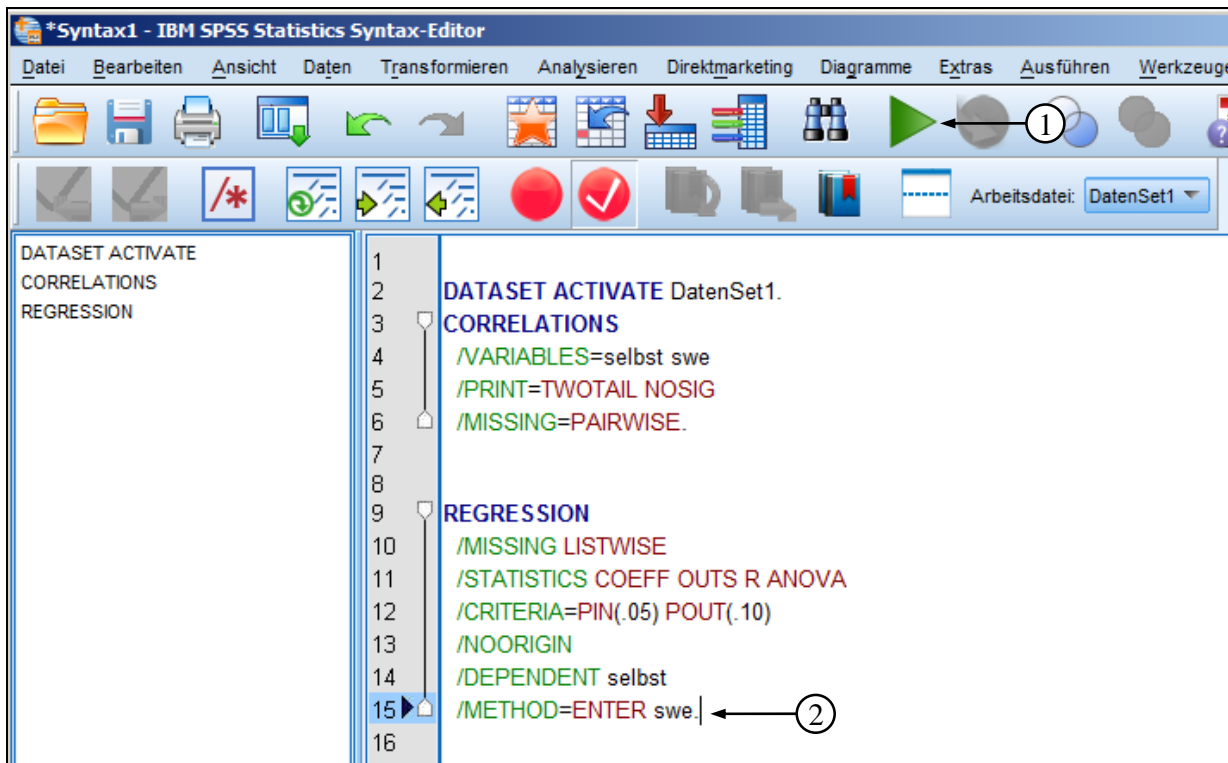


Bild 16

Allgemein kann durch das Anklicken des grünen Dreiecks [1] der Befehl gestartet werden, in dem der Mauszeiger gerade steht. In Bild 16 befindet sich der Zeiger in einer Zeile des Befehls REGRESSION [2], der die konkrete Umsetzung der vorher eingegebenen Anweisungen veranlasst. Wenn mehrere Befehle nacheinander ausgeführt werden sollen, sind diese vor dem Anklicken des grünen Dreiecks [1] zu markieren.

Im Syntax-Fenster sind nun alle eingegebenen Befehle zusammengefasst. Die Syntax-Datei kann abgespeichert und jederzeit erneut geladen werden, wodurch die wiederholte Durchführung der Befehlsfolge zum Beispiel mit neuen Probanden möglich ist. Mit Suchen und Ersetzen können Variablenamen in der gesamten Syntax-Befehlsfolge ausgetauscht werden, wodurch die Berechnungen mit anderen Variablen problemlos ermöglicht werden. Die Vorteile dieser Vorgehensweise bei komplexen Befehlsfolgen, die wiederholt abgearbeitet werden müssen, dürften unmittelbar klar sein.

Die Erzeugung der Syntax-Befehle aus den Dialogfenstern ist der einfachste Weg zur Erstellung einer Syntax-Datei. Alternativ können Syntax-Befehle direkt formuliert werden. Eine wichtige Unterstützung dafür bietet die Datei IBM SPSS Statistics Command Syntax Reference (Bild 17), die man durch Anklicken von Hilfe und danach Befehlssyntax-Referenz (Command Syntax Reference) im Hauptmenü aufrufen kann. In dieser sehr umfangreichen Datei ist die Struktur von allen möglichen Syntax-Befehlen angegeben.

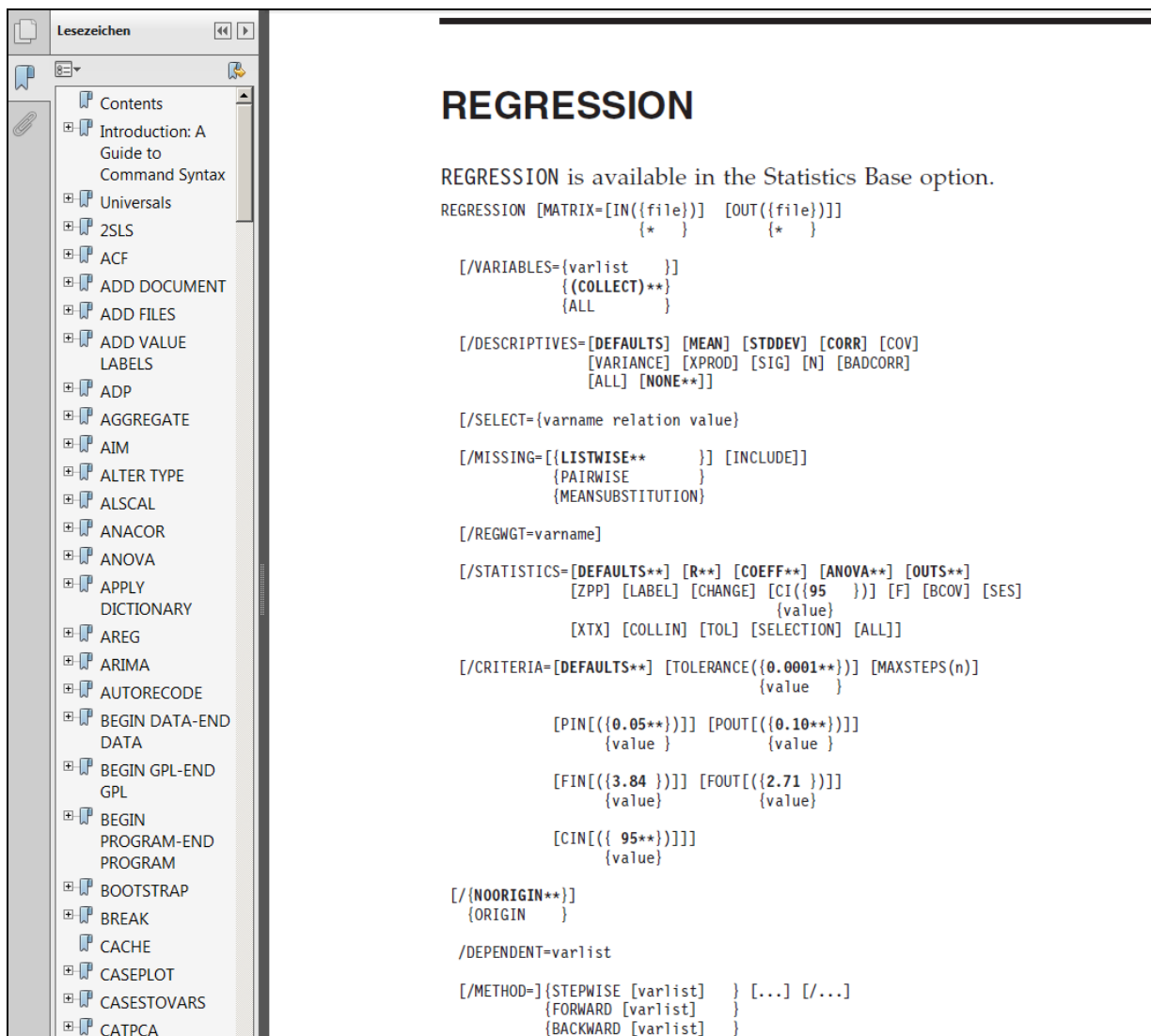


Bild 17

Ausführliche Erläuterungen zur Arbeit mit der SPSS-Syntax geben die Einführungen von Zöfel (2002) und Brosius (2005).

Die flexibelsten Möglichkeiten zur Arbeit mit SPSS bietet für fortgeschrittene Anwender die Skriptprogrammierung. Skriptdateien sind Programme, die auch Berechnungen ermöglichen, die in SPSS routinemäßig nicht enthalten sind. Sie ermöglichen die Gestal-

tung von Dialogfenstern und von Output-Bestandteilen. Eine nachvollziehbare Einführung in die Möglichkeiten und die grundsätzliche Vorgehensweise der Programmierung von Skripten gibt Akremi (2008).

Zusatzaufgabe: Erzeugen Sie analoge Syntax-Befehle zur Analyse der linearen Zusammenhänge zwischen Intelligenz und Ähnlichkeit (Kopieren der vorhandenen Befehle, Ersetzen der Variablennamen) und starten Sie die Analyse mittels der Syntax-Befehle.