

Zu Abschnitt 10.1.2: Inspektion von Häufigkeitsverteilungen

Berechnung des Beispiels mit SPSS®

In der Datei „Kap10_explorativ.sav“ sind die im Lehrbuch auf Seite 205 abgebildeten Rohdaten (Abb. 10.1) abgespeichert. Es handelt sich um eine Variable, die den Therapieerfolg einer Maßnahme indizieren soll. Die Datei besteht aus einer einzigen Variablen („Therapie“) mit $n = 500$ Fällen. Wir zeigen Ihnen nun, wie Sie

- die Histogramme in Abbildung 10.1 reproduzieren können,
- ein Box-Whisker-Diagramm erstellen (s. Abb. 10.2) und
- die Häufigkeitsverteilung auf eine signifikante Abweichung von der Normalverteilung hin überprüfen können.
- Ferner zeigen wir Ihnen, wie Sie Fälle innerhalb eines bestimmten Wertebereichs aus der Analyse ausschließen können.

Berechnung des Beispiels mit Hilfe der SPSS®-Syntax

Die SPSS®-Syntax für eine explorative Analyse der Rohwerte lautet:

EXAMINE

VARIABLES=Therapie

/PLOT BOXPLOT HISTOGRAM NPLOT

/COMPARE GROUP

/STATISTICS DESCRIPTIVES

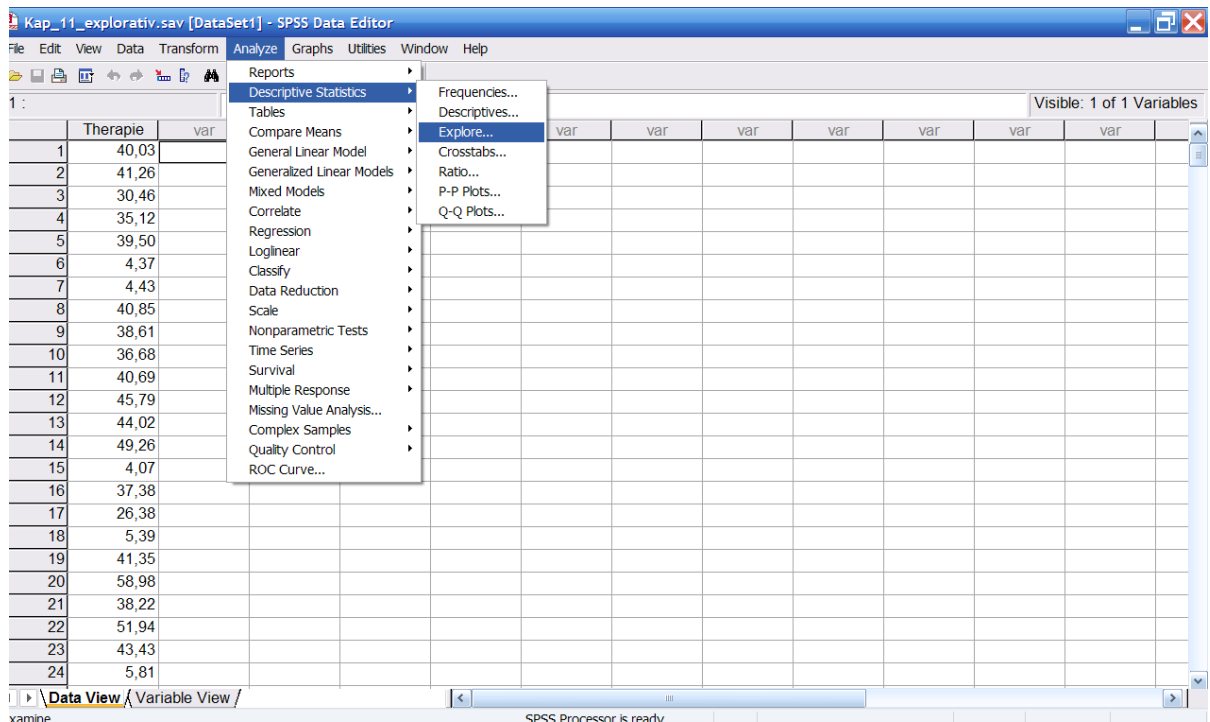
/CINTERVAL 95

/MISSING LISTWISE

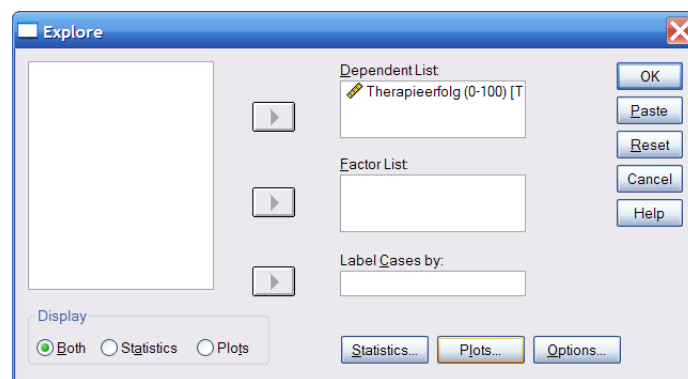
/NOTOTAL.

Berechnung des Beispiels mit Hilfe der SPSS®-Menüführung

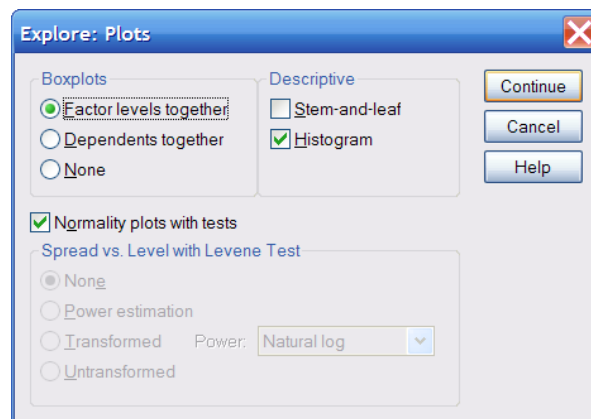
Wenn Sie die Datenanalyse über die Menüführung rechnen wollen, klicken Sie im Menü „Analyze“ auf „Descriptive Statistics“ und anschließend auf „Explore“ (s. Screenshot):



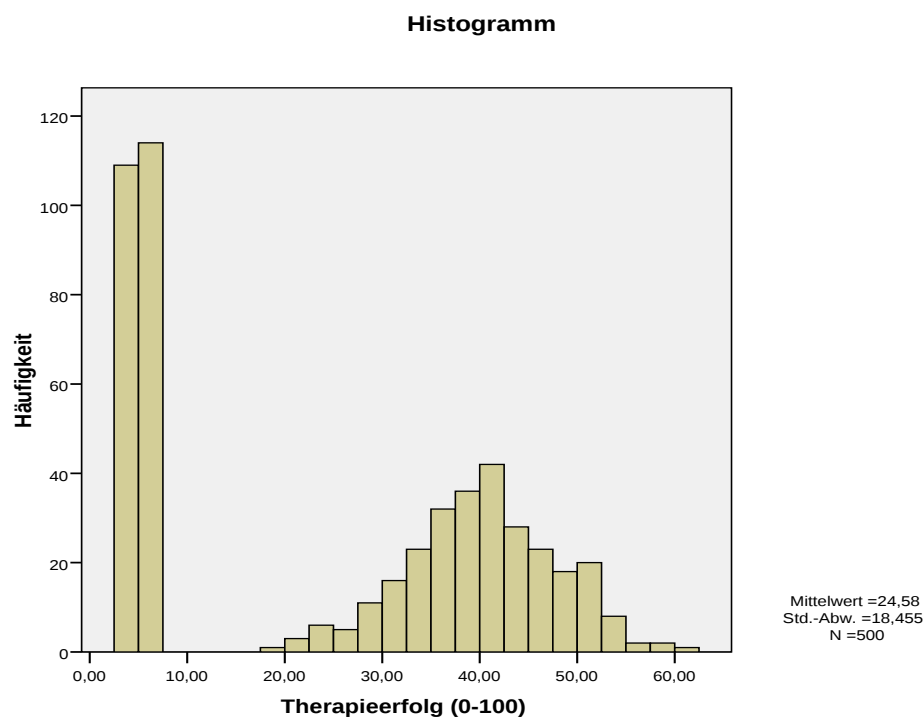
Es öffnet sich das Dialogmenü „Explore“. Schieben Sie (mit Hilfe der Maus oder mit Hilfe der Pfeilspitzen) die Variable „Therapie“ in das Feld „Dependent List“. Dann klicken Sie auf „Plots“.



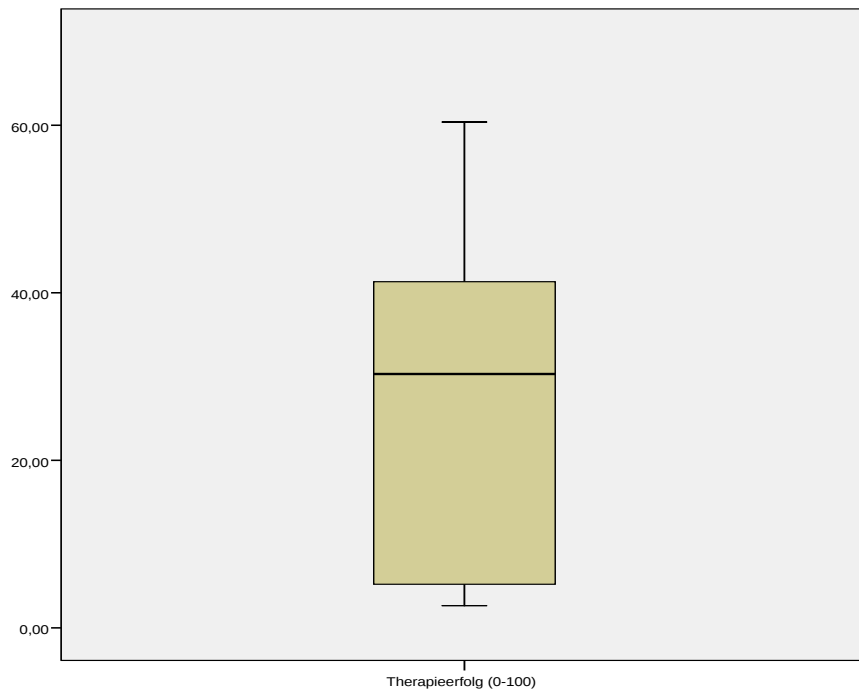
Im Untermenü „Explore: Plots“ aktivieren Sie die Klickkästchen „Histogramm“ und „Normality plots with tests“. Das Kästchen „Stem-and-leaf“ können Sie deaktivieren. Gehen Sie weiter mit „Continue“, im Menü „Explore“ klicken Sie auf „OK“.



In der SPSS-Ausgabe finden Sie unter anderem ein Histogramm, das die Häufigkeitsverteilung der Variable „Therapie“ zeigt. Die Verteilung weicht stark von der Normalverteilung ab, weil die Werte unter 10 einen eigenen Gipfel bilden. Dieses Histogramm ist in Abbildung 10.1 (linke Seite) abgebildet.



Neben dem Histogramm enthält die Ausgabe ein Box-Whisker-Plot, das ebenfalls Auskunft über die Häufigkeitsverteilung der Variable „Therapie“ gibt (s. Abb. 10.2).



Um die Häufigkeitsverteilung der Normalverteilung anzupassen, empfiehlt es sich Werte, die unter 10 liegen, aus der Analyse auszuschließen.

Auswahl von Fällen mit Hilfe der SPSS®-Syntax

Die Syntax zur Auswahl von Fällen lautet:

USE ALL.

COMPUTE filter_\$=(Therapie > 10).

VARIABLE LABEL filter_\$ 'Therapie > 10 (FILTER)'.

VALUE LABELS filter_\$ 0 'Not Selected' 1 'Selected'.

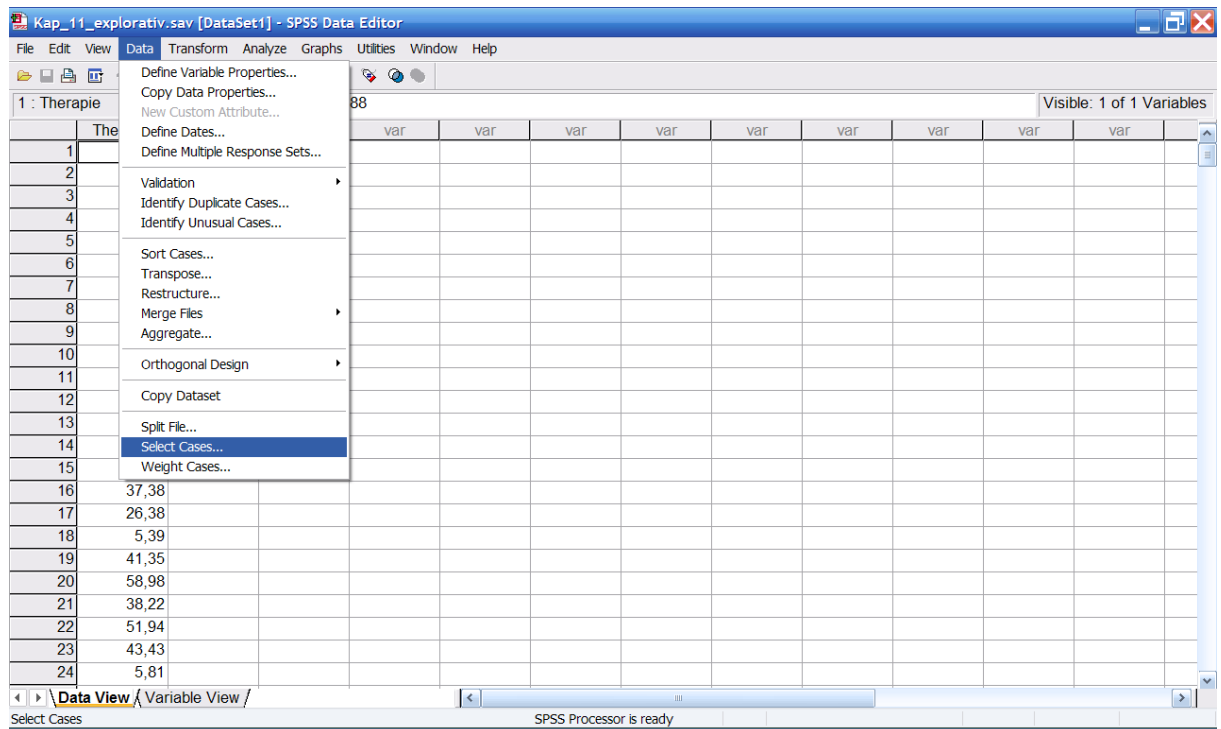
FORMAT filter_\$ (f1.0).

FILTER BY filter_\$.

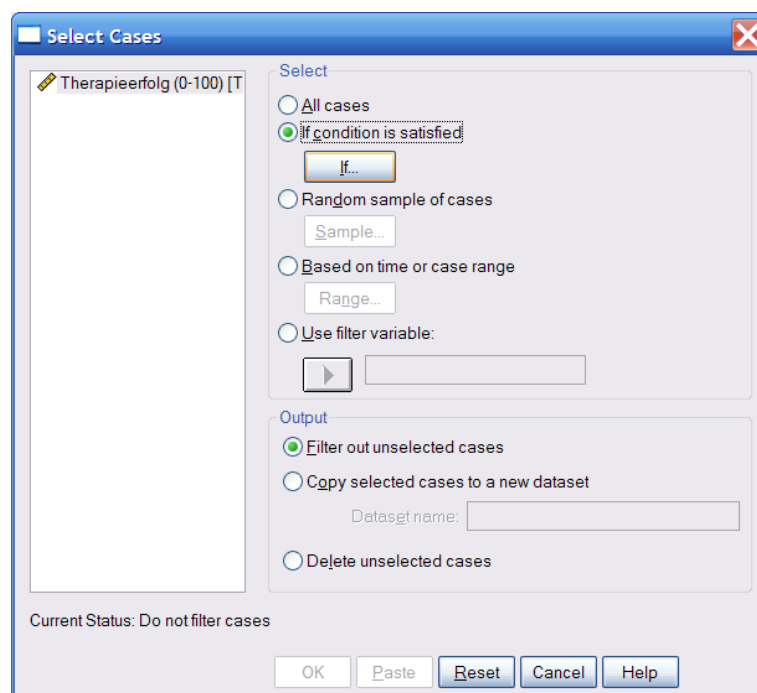
EXECUTE .

Auswahl von Fällen mit Hilfe der SPSS®-Menüführung

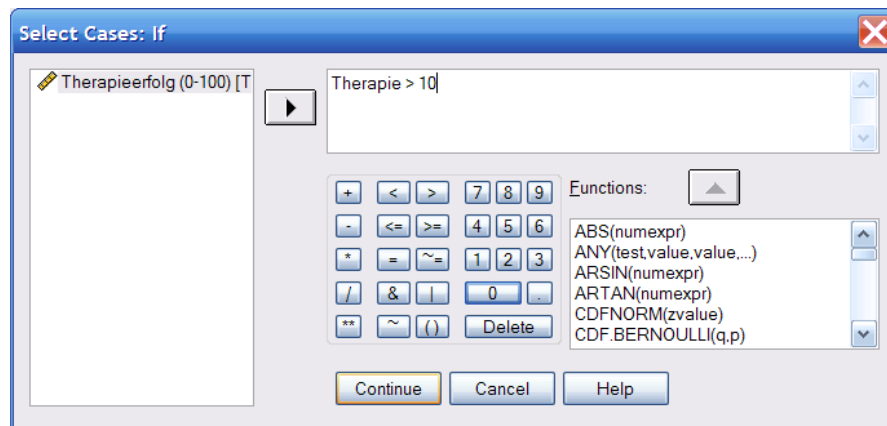
Um die Fälle über die Menüführung auszuwählen, gehen Sie auf „Data“, „Select Cases“ (siehe Screenshot):



Es öffnet sich das Dialogmenü „Select Cases“. Dort aktivieren Sie das Kästchen „If condition is satisfied“. Achten Sie darauf, dass im Bereich „Output“ das Klickkästchen „Filter out unselected cases“ aktiviert ist, sonst werden unter Umständen die übrigen Daten gelöscht. Dann klicken Sie auf „If“.



Im Menü „Select Cases: If“ können Sie die Bedingung definieren, nach der die Fälle ausgewählt werden. Da in die kommende Analyse nur Fälle aufgenommen werden, die Werte über 10 haben, geben Sie folgenden Auswahlbefehl ein: zuerst schieben Sie die Variable „Therapie“ in das leere Feld und dann geben Sie (über die Buttons oder die Tastatur) den Befehl „> 10“ ein. Klicken Sie auf „Continue“ und schließlich auf „OK“.



In der Datenansicht erscheint nun die neue Variable „filter_\$“. In der ersten Spalte, sind die Fälle durchgestrichen, die ausgeschlossen wurden.

*Kap_11_explorativ.sav [DataSet1] - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

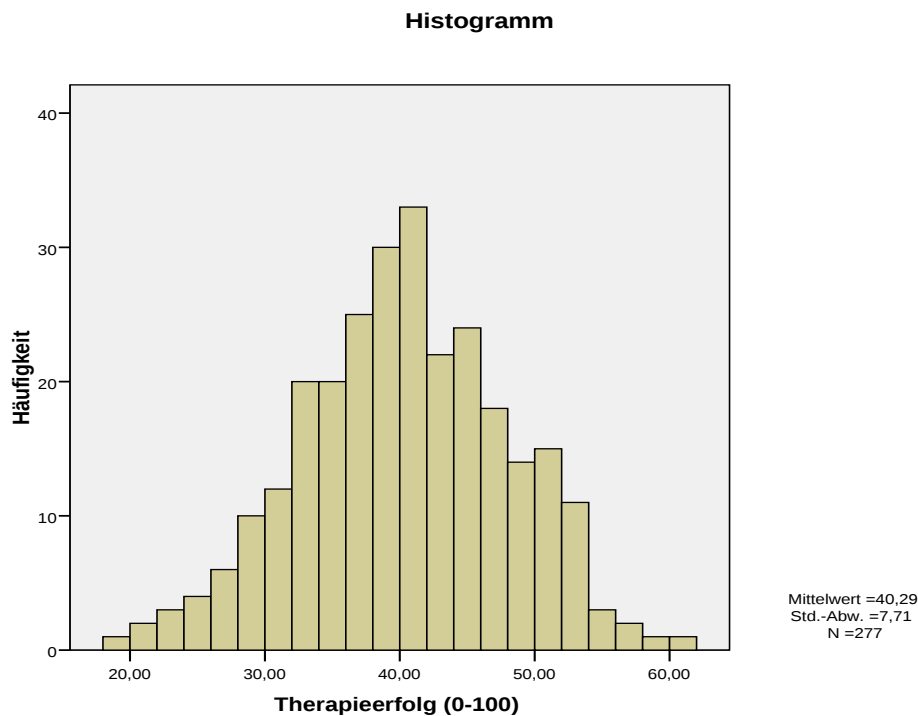
1 : Therapie 40,0253884822388 Visible: 2 of 2 Variables

	Therapie	filter_\$	var	var	var	var	var	var	var	var	var	var
1	40,03	1										
2	41,26	1										
3	30,46	1										
4	35,12	1										
5	39,50	1										
6	4,37	0										
7	4,43	0										
8	40,85	1										
9	38,61	1										
10	36,68	1										
11	40,69	1										
12	45,79	1										
13	44,02	1										
14	49,26	1										
15	4,07	0										
16	37,38	1										
17	26,38	1										
18	5,39	0										
19	41,35	1										
20	58,98	1										
21	38,22	1										
22	51,94	1										
23	43,43	1										
24	5,81	0										

SPSS Processor is ready Filter On

Jetzt soll überprüft werden, wie sich die Häufigkeitsverteilung im Datensatz verändert hat, nachdem die Werte unter 10 ausgeschlossen wurden. Dazu führen Sie erneut eine explorative Datenanalyse durch. Im Menü „Explore“ geben Sie als abhängige Variable wieder die Variable „Therapie“ ein und nicht die Filtervariable.

In der SPSS®-Datenausgabe zeigt sich im Histogramm, dass die Verteilung jetzt viel stärker der Normalverteilung gleicht. Dieses ist in Abbildung 10.1 (rechte Seite) abgebildet.



Daneben lässt sich die Annahme, dass sich die Verteilung durch das Ausschließen der Fälle der Normalverteilung angenähert hat, auch inferenzstatistisch absichern.

In der Datenausgabe sind in der Tabelle „Tests auf Normalverteilung“ die Ergebnisse des Kolmogorov-Smirnov-Tests und des Shapiro-Wilk-Tests aufgeführt (nähere Erläuterungen zu diesen Tests finden Sie im Lehrbuch von Eid et al., 2013, Abschn. 10.6). Beide Tests sind auf einem Fehlniveau von $\alpha = 5\%$ nicht signifikant, das heißt, dass die vorliegende Verteilung nicht signifikant von der Normalverteilung abweicht.

Tests auf Normalverteilung

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Therapieerfolg (0-100)	,023	277	,200*	,996	277	,752

*. Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

Zum Vergleich zeigt sich in den Testergebnissen der ersten Datenanalyse, in die alle Rohwerte einbezogen wurden, dass die Häufigkeitsverteilung signifikant von einer Normalverteilung abweicht ($p < .001$).

Tests auf Normalverteilung

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Therapieerfolg (0-100)	,271	500	,000	,823	500	,000

a. Signifikanzkorrektur nach Lilliefors

Berechnung des Beispiels mit einem anderen Programm

Die Originaldaten für dieses Beispiel sind auch in einem allgemein lesbaren ASCII-Format abgespeichert („Kap10_explorativ.dat“).