

Kapitel 5

FRAGESTELLUNG 1

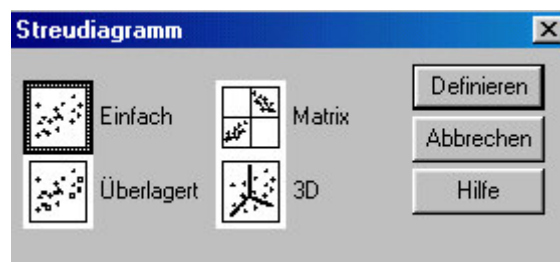
Öffne die Datei „**alctobac.sav**“.

Zuerst werden wir ein Streudiagramm erstellen, um einen grafischen Überblick von diesem Datensatz zu erhalten.

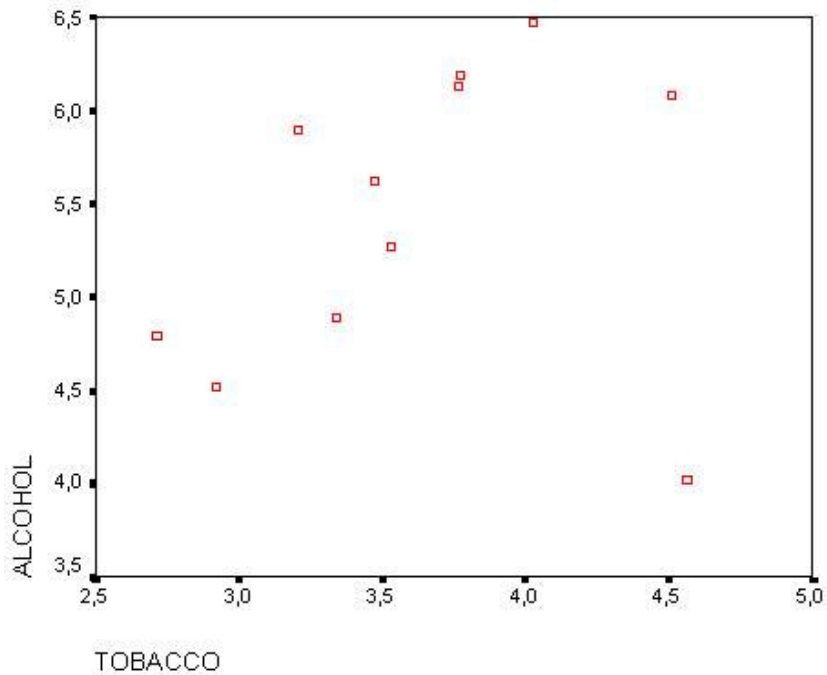
Gehe dazu auf „*Grafiken / Streudiagramm*“



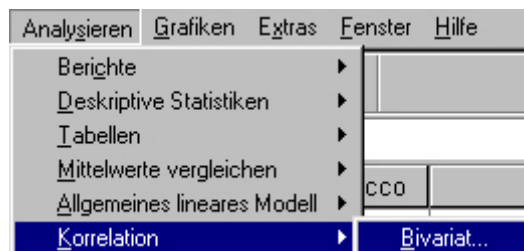
wähle „*Einfach*“ aus.



Mit „*Definieren*“ kommen wir ins nächste Menü, in dem wir als **Y-Variable** „alcohol“ und als **X-Variable** „tobacco“ auswählen. Bestätige mit „*OK*“.



Als nächsten Schritt wollen wir den Korrelationskoeffizienten berechnen. Dazu müssen wir auf „Analysieren / Korrelation / Bivariat“ gehen.



Als Variablen sind „alcohol“ und „tobacco“ anzugeben. Weiters sollte „Pearson“ und bei „Test auf Signifikanz“ der Punkt „zweiseitig“ ausgewählt sein. Bestätige mit „OK“.



Wir erhalten nun den Korrelationskoeffizienten nach Pearson für alle Werte (also **mit** dem „outlier“ NorthernIreland) $r_{xy} = 0.224$ und dem Signifikanzwert (zweiseitig) $p = 0.509$.

Korrelationen

		ALCOHOL	TOBACCO
ALCOHOL	Korrelation nach Pearson	1,000	,224
	Signifikanz (2-seitig)	,	,509
	N	11	11
TOBACCO	Korrelation nach Pearson	,224	1,000
	Signifikanz (2-seitig)	,509	,
	N	11	11

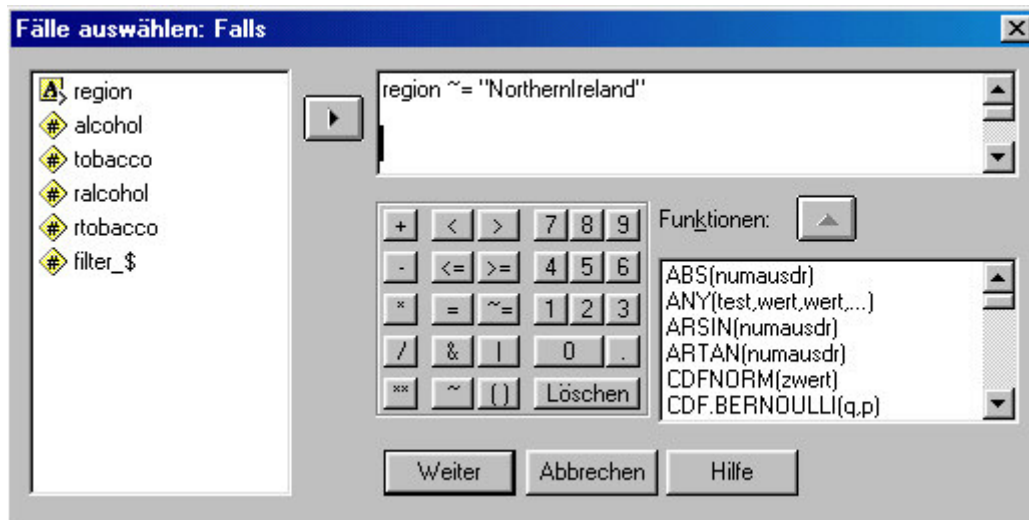
Im nächsten Schritt berechnen wir den Korrelationskoeffizienten **ohne** NorthernIreland. Dazu müssen wir mittels des Menüpunkts „Daten / Fälle auswählen“ den Ausreißer rausfiltern.



Klicke danach rechts im Feld „Auswählen“ auf „Falls Bedingung zutrifft“.



Jetzt müssen wir auf „Falls...“ drücken und als Formel **region~="NorthernIreland"** eingeben. Mit „Weiter“ und „OK“ bestätigen



„region“ ist als Stringvariable definiert, daher müssen wir NorthernIreland unter Hochkomma setzen (Hinweis: es ist auch auf Gross/Kleinschreibung zu achten). Zum Befehl Fälle auswählen siehe Kapitel 3 Fragestellung 3.

	region	alcohol	tobacco	ralcohol	rtobacco	filter_\$
1	North	6,47	4,03	11,000	9,000	Ausgewählt
2	Yorkshire	6,13	3,76	9,000	7,000	Ausgewählt
3	Northeast	6,19	3,77	10,000	8,000	Ausgewählt
4	EastMidlands	4,89	3,34	4,000	4,000	Ausgewählt
5	WestMidlands	5,63	3,47	6,000	5,000	Ausgewählt
6	EastAnglia	4,52	2,92	2,000	2,000	Ausgewählt
7	Southeast	5,89	3,20	7,000	3,000	Ausgewählt
8	Southwest	4,79	2,71	3,000	1,000	Ausgewählt
9	Wales	5,27	3,53	5,000	6,000	Ausgewählt
10	Scotland	6,08	4,51	8,000	10,000	Ausgewählt
11	NorthernIreland	4,02	4,56	1,000	11,000	Nicht ausgewählt

Wenn wir jetzt den Korrelationskoeffizienten berechnen (dabei müssen die gleichen Schritte machen wie eingangs erklärt gemacht werden), sehen wir, dass der Zusammenhang zwischen den zwei Variablen viel stärker ist. **Ohne** den „outlier“ NorthernIreland beträgt $r_{xy} = 0.784$ und der Signifikanzwert (zweiseitig) ist $p = 0.007$.

Korrelationen

		ALCOHOL	TOBACCO
ALCOHOL	Korrelation nach Pearson	1,000	,784**
	Signifikanz (2-seitig)	,	,007
	N	10	10
TOBACCO	Korrelation nach Pearson	,784**	1,000
	Signifikanz (2-seitig)	,007	,
	N	10	10

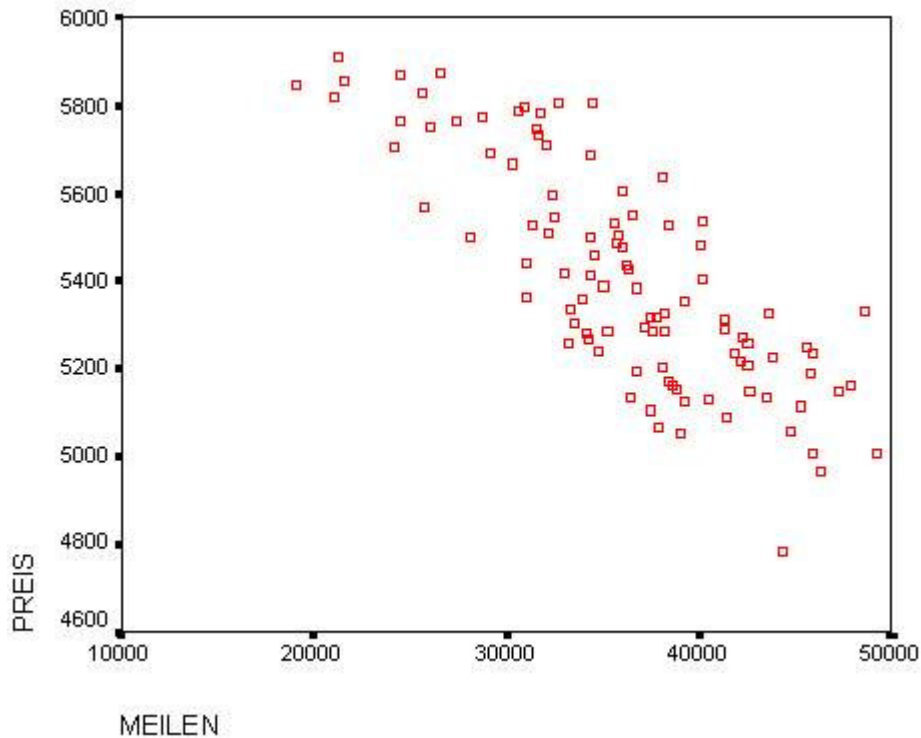
** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

FRAGESTELLUNG 2

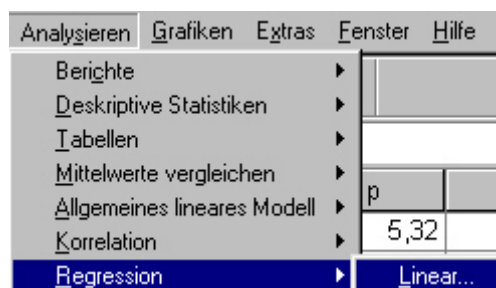
Öffne die Datei „2ndhandcar.sav“.

Für den grafischen Überblick erstellen wir wie in Fragestellung 1 ein **einfaches Streudiagramm**.

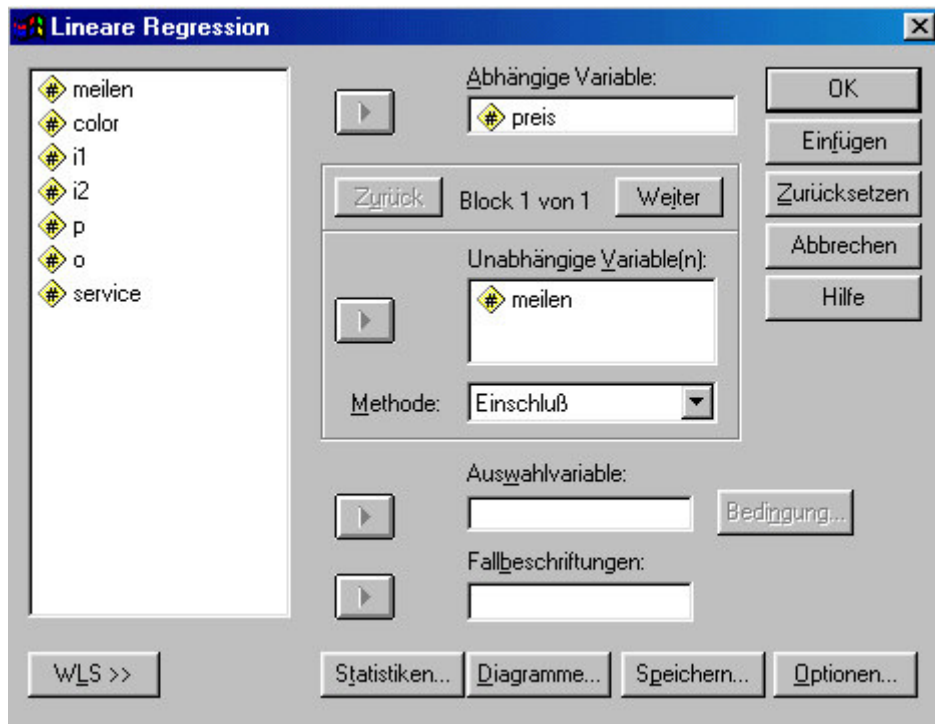
Für die Y-Achse nehmen wir „preis“, für die X-Achse „meilen“. Das Streudiagramm sollte folgendermaßen aussehen:



Zum Berechnen des Regressionsmodells gehen wir auf „*Analysieren / Regression / Linear*“.



Als „abhängige Variable“ (Y-Variable) geben wir „preis“ und als unabhängige Variable (X-Variable) „meilen“ an.



Um die vorhergesagten Werte in die Datenansicht zu bekommen, müssen wir auf „Speichern“ drücken. Folgende Auswahlmöglichkeiten sind anzukreuzen:



Sowohl bei „Vorhergesagte Werte“, als auch bei den „Residuen“ wählen wir „nicht standardisiert“ aus. Mit „Weiter“ und „OK“ bestätigen.

Wir erhalten in der Datenansicht die vorhergesagten Werte (Variable „pre_1“) und die Residuen (Variable „res_1“). Im Ausgabefenster werden wir mit viel Output konfrontiert, von dem für uns aber vorläufig nur die Tabelle **Koeffizienten^a** interessant ist.

Koeffizienten^a

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten	T	Signifikanz
		B	Standardfehler	Beta		
1	(Konstante)	6533,383	84,512		77,307	,000
	MEILEN	-3,12E-02	,002	-,806	-13,495	,000

a. Abhängige Variable: PREIS

Die Konstante liegt bei **6533.383** und der „Anstieg“ beträgt **-0.031**. Daraus ergibt sich die Regressionsgleichung **$Y = 6533.383 - 0.031 * X$**

Zur Überprüfung dieses Modells betrachten wir im Menü „Ausgabe / Modellzusammenfassung“. Hier wird der R^2 – Wert angegeben, mit dem ausgesagt wird, wie viel Prozent der Daten durch das Modell erklärt wird (in unserem Fall also **65%**).

Modellzusammenfassung^b

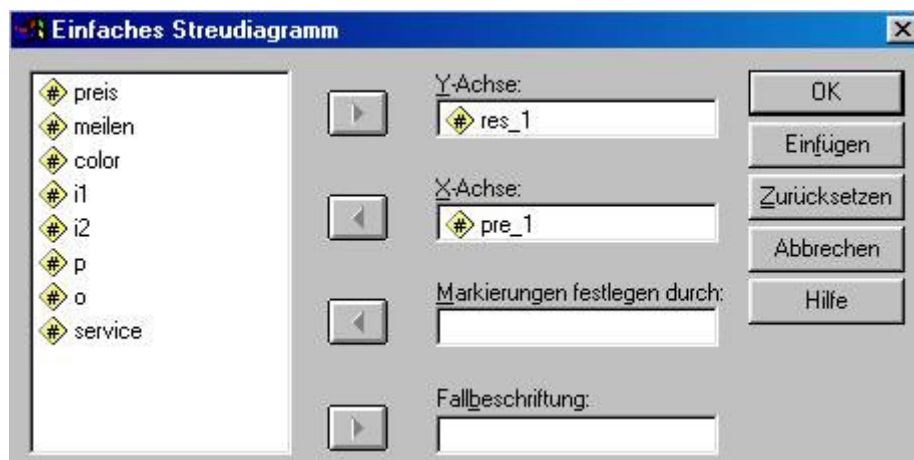
Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,806 ^a	,650	,647	151,57

a. Einflussvariablen : (Konstante), MEILEN

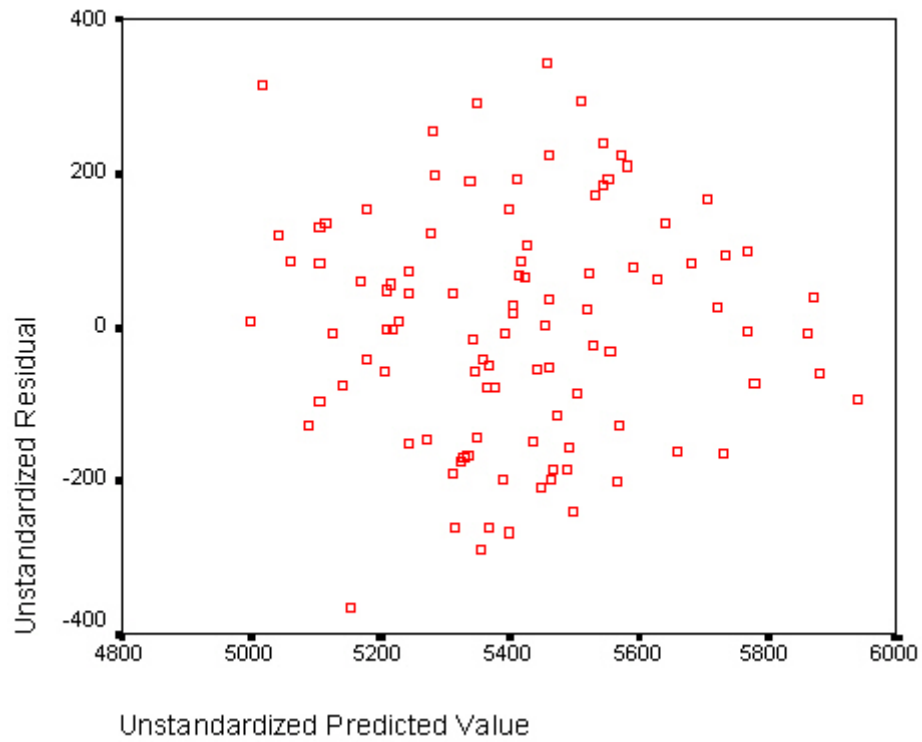
b. Abhängige Variable: PREIS

Grafisch eignet sich ein Residuen – Plot sehr gut, um das Modell zu überprüfen. Dazu gehen wir auf „Grafiken / Streudiagramm“ und wählen „einfach“ aus.

Danach geben wir als Y-Variable „res_1“ und als X-Variable „pre_1“ an. Diese Variablen wurden durch die obige Auswahl von **Residuen** und **Vorhergesagte Werte**, (engl. „predicted“, daher „pre_“) von SPSS erzeugt und in der Datenmatrix hinten angefügt.



Mit „OK“ bestätigen.



Der Residuenplot zeigt kein deutliches Muster, daher dürften die Voraussetzungen zur Berechnung eines Regressionsmodells erfüllt sein.

FRAGESTELLUNG 3a

Öffne die Datei „bicycle.sav“.

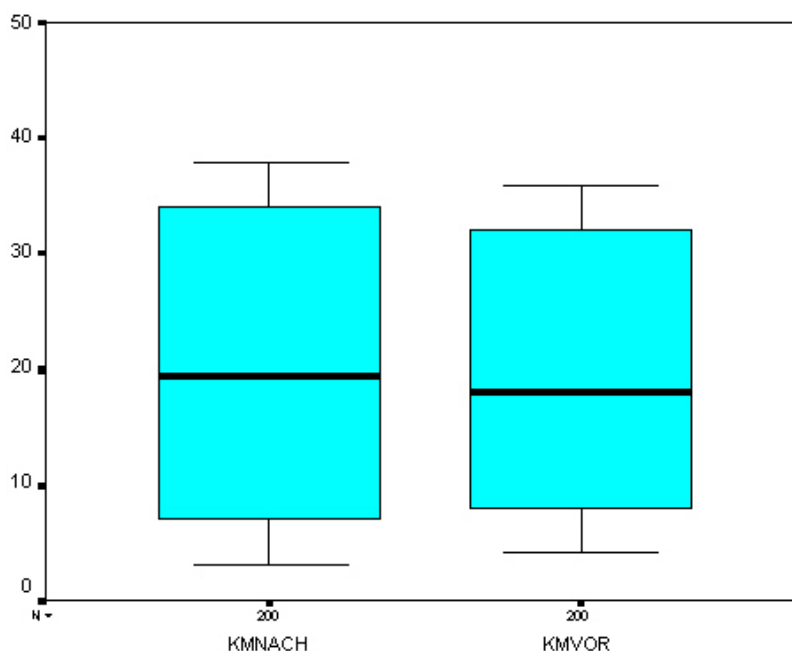
Als deskriptive Methode eignet sich ein Boxplot. Dazu gehen wir auf „Grafiken / Boxplot“. Im anschließenden Menü „Einfach“ und „Auswertung über verschiedene Variablen“ auswählen und mit „Definieren“ bestätigen.



Im folgenden Menü beide Variablen in die rechte Box schieben und mit „OK“ bestätigen.



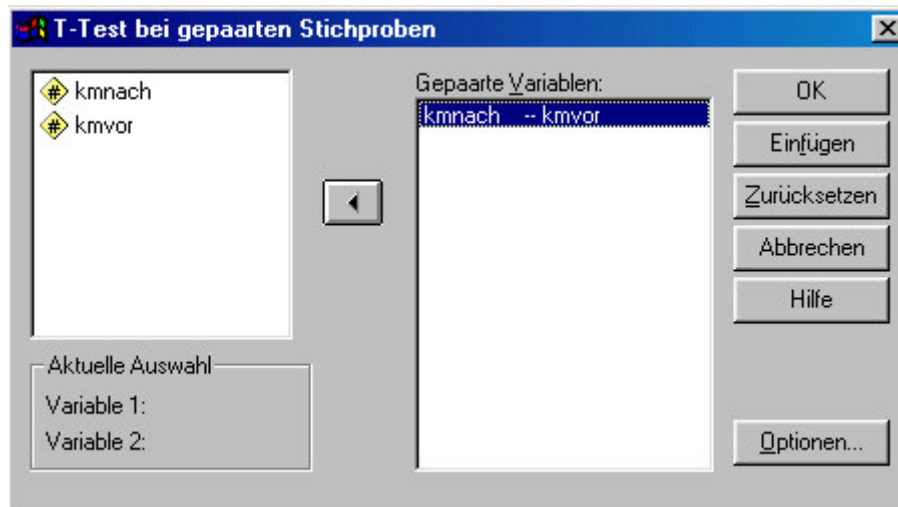
Wir erhalten den Boxplot:



Gehe nun auf „Analysieren / Mittelwerte vergleichen / T-Test bei gepaarten Stichproben“ aus.



Gehe wie folgt vor:



Zuerst „kmnach“ und dann „kmvor“ markieren und die Variablen hinüber schieben.

Mit „OK“ bestätigen.

Test bei gepaarten Stichproben

		Gepaarte Differenzen				T	df	Sig. (2-seitig)	
		Mittelwert	Standardabweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz				
					Untere				Obere
Paaren 1	KMNACH - KMOVOR	,7700	3,0650	,2167	,3426	1,1974	3,553	199	,000

Wir erhalten den **T-Wert 3.553** und ein Signifikanzwert, der mit 0,000 angegeben ist. Dieser Wert ist natürlich nicht wirklich 0 aber so klein, dass bei Rundung auf 3 Stellen hinter dem Komma noch immer 0,000 stehen bleibt.

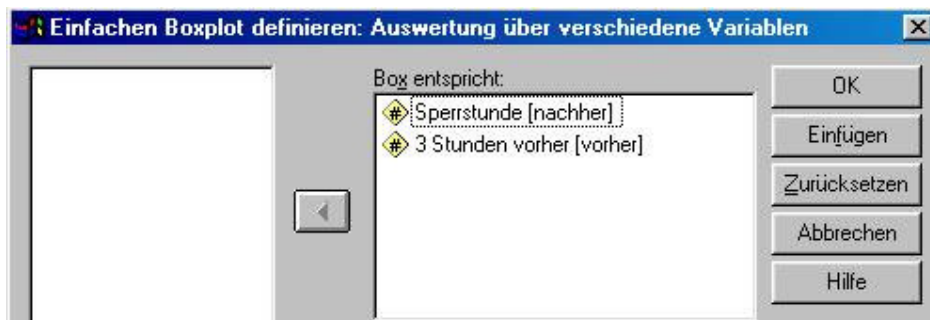
FRAGESTELLUNG 3b

Öffne die Datei „**alcattrneu.sav**“.

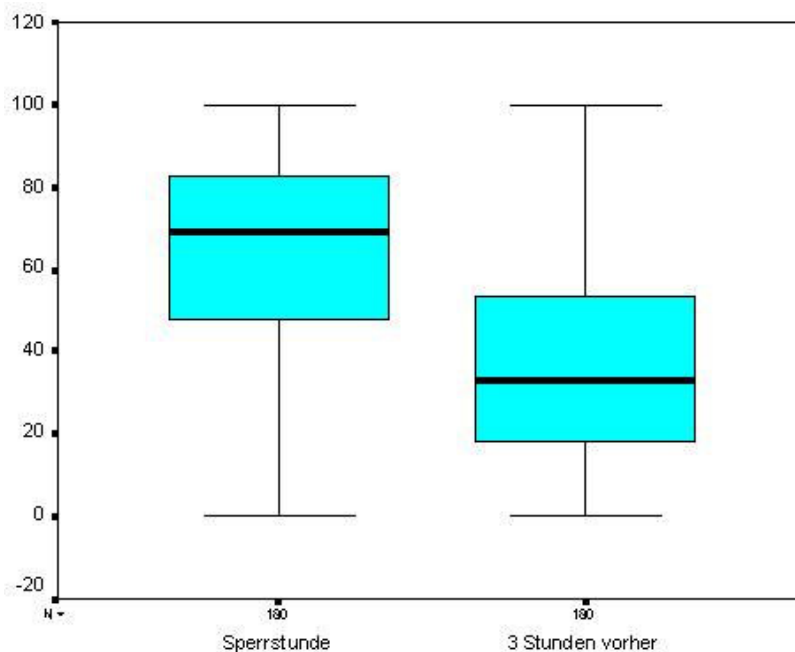
Gehe auf „**Grafiken / Boxplot**“. Im anschließenden Menü wiederum „**Einfach**“ und „**Auswertung über verschiedene Variablen**“ auswählen und mit „**Definieren**“ bestätigen.



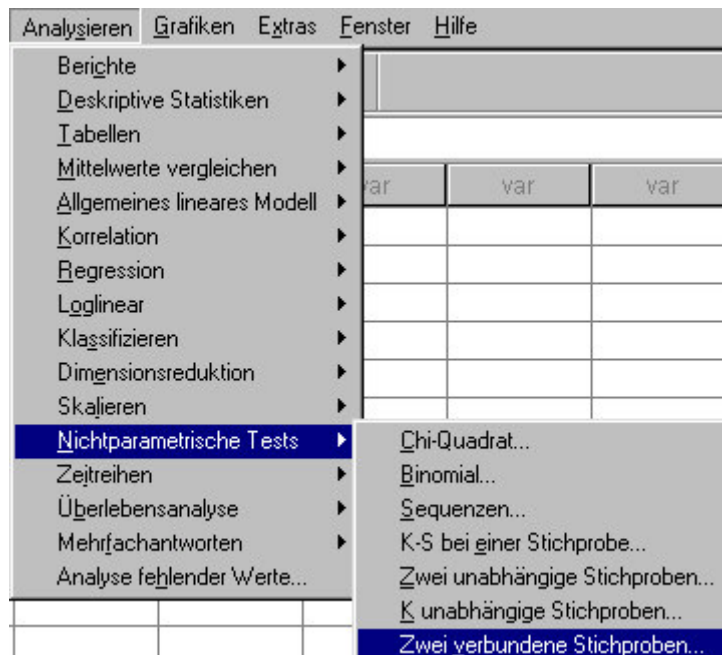
Schieben Sie die zwei variablen wiederum nach rechts und drücken dann auf „**OK**“.



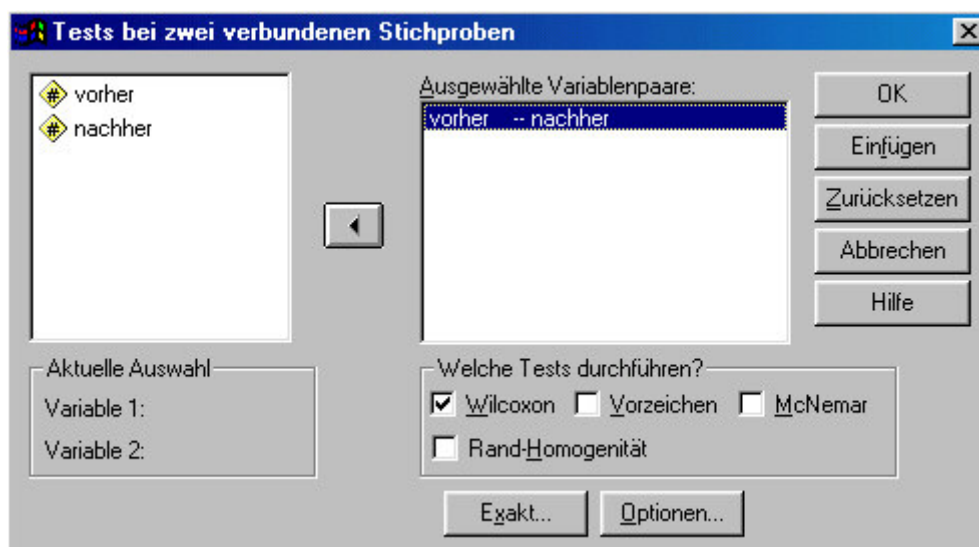
Der Boxplot:



Gehe nun auf „Analysieren / Nichtparametrische Tests / Zwei verbundene Stichproben“.



Als nächstes markiere die Variable „vorher“ und dann „nachher“. Jetzt können die Variablen in das Feld „Ausgewählte Variablenpaare“ verschoben werden.



Drücke auf „OK“.

Wir erhalten nun als **Ausgabe** den **Wilcoxon –Signifikanzwert**:

Ränge

		N	Mittlerer Rang	Rangsumme
Sperrstunde - 3 Stunden vorher	Negative Ränge	39 ^a	61,67	2405,00
	Positive Ränge	138 ^b	96,72	13348,00
	Bindungen	3 ^c		
	Gesamt	180		

a. Sperrstunde < 3 Stunden vorher

b. Sperrstunde > 3 Stunden vorher

c. 3 Stunden vorher = Sperrstunde

Statistik für Test^b

	Sperrstunde - 3 Stunden vorher
Z	-8,015 ^a
Asymptotische Signifikanz (2-seitig)	,000

a. Basiert auf negativen Rängen.

b. Wilcoxon-Test