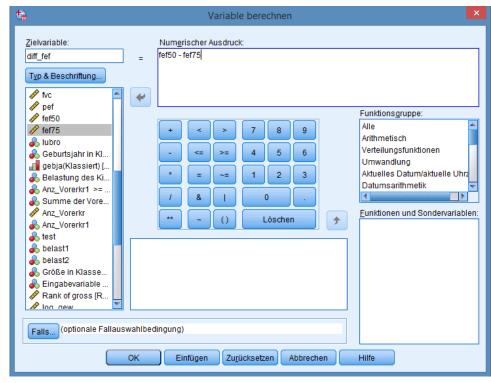
SPSS – Kurs online Lösungen

3) Erzeugen Sie eine Variable, die die Differenz zwischen fef50 und fef75 angibt, Name: diff_fef, Berechnung: fef50 – fef75

Transformieren → Variable berechnen



5) Variable geb_jz (Jahreszeit Geburt) ausgehend von der Variable gebmo erzeugen:

Transformieren -> Umcodieren in andere Variablen



5ff) Schaltfläche "Alte und neue Werte"

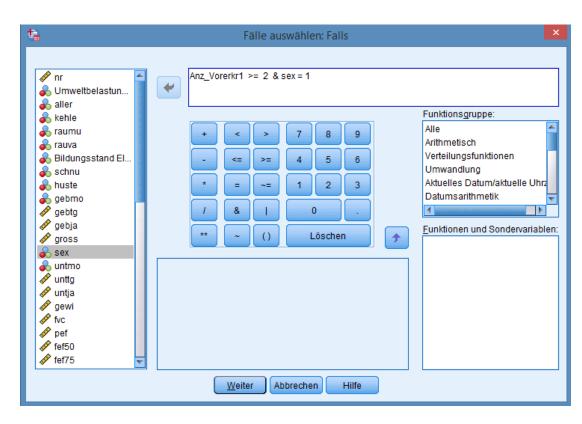


<u>Beachten</u>: Häkchen setzen bei "Ausgabe der Variablen als Zeichenfolgen" oder anstatt "Winter, etc." eine Codierung (1,2,...) nehmen.

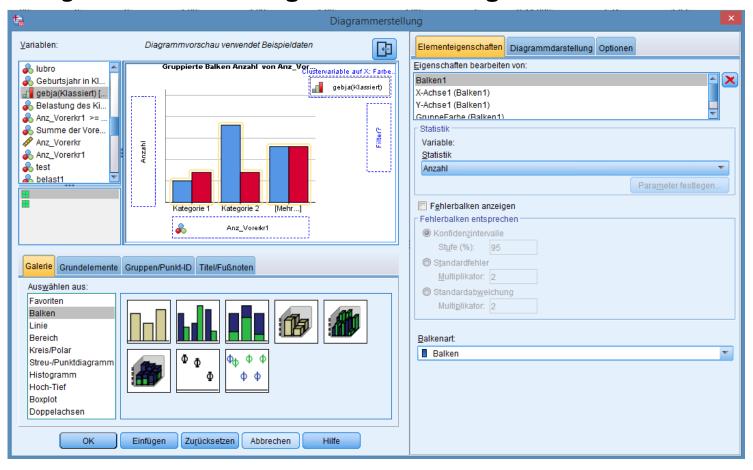
6) Filter setzen: männliche Kinder mit mindestens 2 Vorerkrankungen (Anz_Vorerkr1) auswählen

Daten → Fälle auswählen, Schaltfläche "Falls Bedingung zutrifft",

Formel wie gezeigt eintragen.



 Unterschied zwischen alten und jungen Kindern hinsichtlich der Anzahl an Vorerkrankungen grafisch und per Kreuztabelle zeigen. Grafik → Diagrammerstellung



1ff) Kreuztabelle: Analysieren → Deskriptive Statistiken → Kreuztabellen, Schaltfläche "Statistiken" Häkchen bei Chi-Quadrat; Schaltfläche Zellen, Häkchen bei Häufigkeiten Beobachtet und Erwartet.

Anz_Vorerkr1 * Geburtsjahr in Klassen Kreuztabelle

			Geburtsjahr	in Klassen	
			alt	jung	Gesamt
Anz_Vorerkr1	0	Anzahl	375	631	1006
		Erwartete Anzahl	363,0	643,0	1006,0
	1	Anzahl	127	226	353
		Erwartete Anzahl	127,4	225,6	353,0
	2	Anzahl	43	91	134
		Erwartete Anzahl	48,4	85,6	134,0
	3	Anzahl	13	35	48
		Erwartete Anzahl	17,3	30,7	48 48,0
	4	Anzahl	1	7	8
		Erwartete Anzahl	2,9	5,1	8,0
Gesamt		Anzahl	559	990	1549
		Erwartete Anzahl	559,0	990,0	1549,0

Chi-Quadrat-Tests

	Wert	df	Asymptotisch e Signifikanz (zweiseitig)
Chi-Quadrat nach Pearson	5,164ª	4	,271
Likelihood-Quotient	5,601	4	,231
Zusammenhang linear- mit-linear	4,170	1	,041
Anzahl der gültigen Fälle	1549		

a. 1 Zellen (10,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 2,89.

p-Wert >0.05, also keine sign. Unterschiede zwischen Altersgruppen

2) Häufigkeit allergische Atemwegserkrankungen, Kehlkopfentzündung und wie häufig beides zusammen

Analysieren → Deskriptive Statistiken → Häufigkeiten



Hiermit erhält man die einzelnen Häufigkeiten. Um zu erfahren, bei welchen Probanden beides zusammen auftritt: neue Variable erstellen, z.B. über Transformieren

Variable berechnen, Numerischer Ausdruck = aller+kehle, und sich dann auf dem gleichen Weg wie oben die Häufigkeit angeben lassen. Die "2er" haben beide Vorerkrankungen.

Ubung 2

3) Kinder in drei Größengruppen nach Geschlecht einteilen

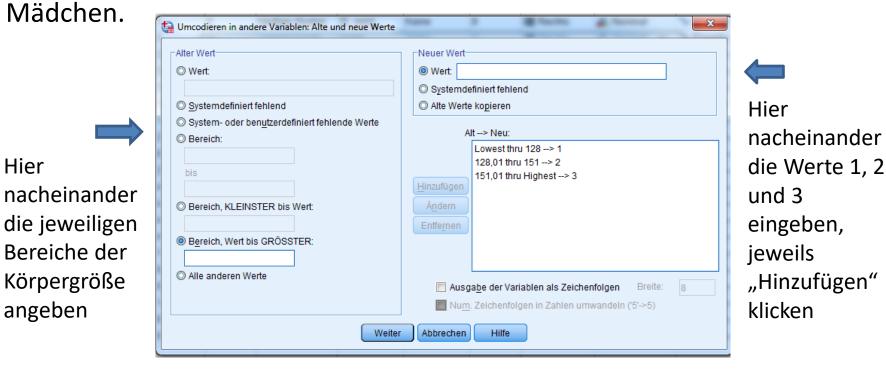
Transformieren → Umcodieren in andere Variablen

Schaltfläche "Alte und neue Werte", zuerst Eingabe der Werte für die

Mädchen.

Hier

angeben

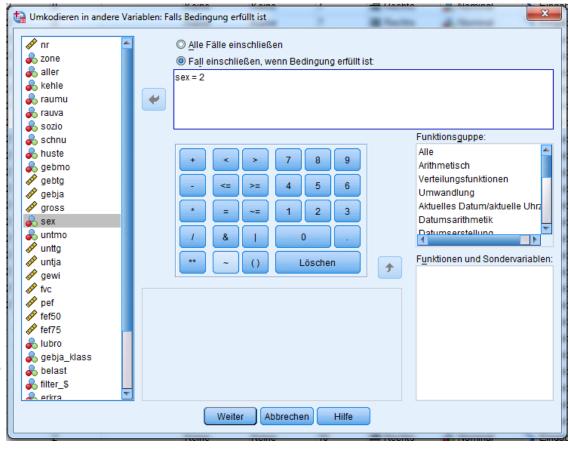


Durch Klicken von "Weiter" kehrt man wieder zu dem vorherigen Fenster zurück.

3ff) Schaltfläche "Falls…" und Eintragen der Bedingung "sex=2", damit die vorher festgelegten Werte nur für die Mädchen

angewendet werden.

Für die Jungen wird der Vorgang wiederholt, die Grenzen der Körpergröße dementsprechend anpassen (vorherige Folie) und bei "Falls" als Bedingung sex=1 angeben. Der Variablenname bleibt gleich!!!



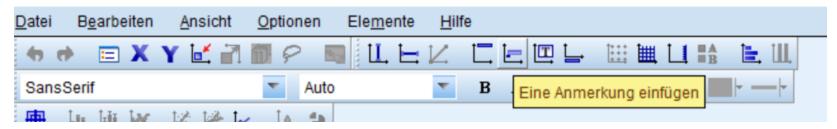
4) Bildungsstand der Eltern als Balkendiagramm Muster, Dicke der Balken, y-Achsenabschnitt verändern. Anmerkung einfügen; "rauva" als Unterteilungsvariable

Grafik → Diagrammerstellung, einfache Balken in das Vorschaufenster ziehen, Variable "sozio" auf x-Achse ziehen, OK klicken.

Dann Grafik doppelt klicken, Balken doppelt klicken, im Eigenschaften-Fenster auf *Füllung und Rahmen*, Muster verändern; dann zu *Optionen für Balken*, mit Schieber Dicke verändern; Eigenschaften Fenster schließen.

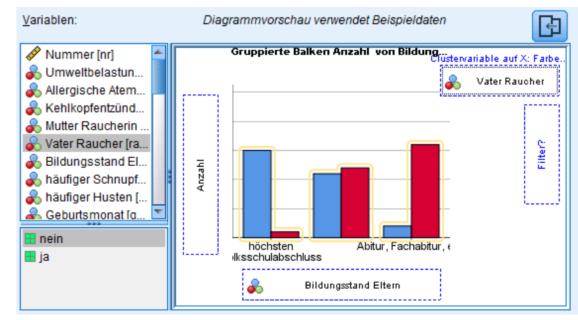
Dann y-Achse doppelt klicken, im Eigenschaften-Fenster auf Skala, dort die Werte verändern. Fenster schließen.

4ff) Dann die Anmerkung einfügen:

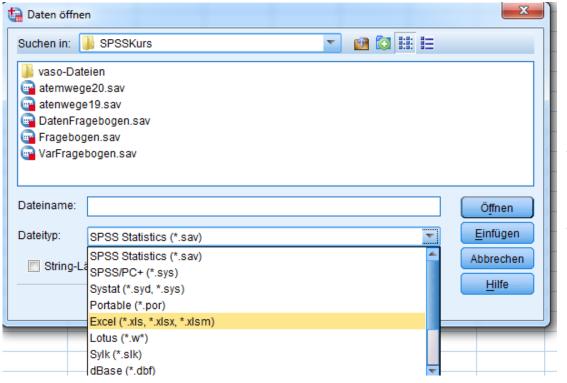


Für die Unterteilung nach "rauva" zurück in Grafik →

Diagrammerstellung, die Gruppierten Balken oben ins Vorschaufenster ziehen, "rauva" als Clustervariable angeben.



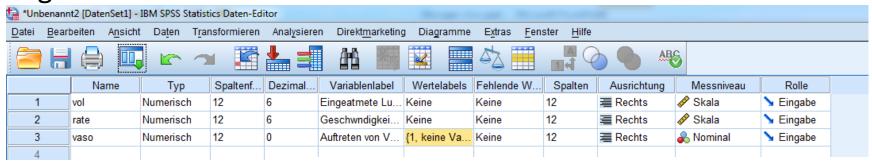
1) Im Menü **Datei** → **Öffnen** → **Daten**, es öffnet sich folgendes Fenster:



Damit die Excel-Dateien angezeigt werden, unten die Auswahl des Dateityps setzen (gelb).

Abfrage, ob erste Zeile die Variablennamen sind: ja!

2) Überprüfen und vervollständigen Sie die Variablenübersicht Eintragen der Variablenlabels und bei "vaso" der Wertelabels gemäß Infoblatt der Daten



3) Menüpunkt
Analysieren →
Deskriptive
Statistiken →
Häufigkeiten

Auftreten von Vasokonstriktion N Gültig 39 Fehlend 0

Statistiken

	Auftreten von Vasokonstriktion							
			Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente		
١	Gültig	keine Vasokonstriktion	20	51,3	51,3		n Aktivier	
		Vasokonstriktion beobachtet	19	48,7	48,7	100 <mark>, ^{dop}</mark>	pelklicke	
		Gesamt	39	100,0	100,0			

Häufigkeiten

3) Führen Sie deskriptive Analyseverfahren durch: wie häufig treten bei der Variable "vaso" 1 und 2 auf. In welchem Bereich liegen die Werte von "vol" und "rate" (Spannweite)? Sind die Werte plausibel?

Menüpunkt Analysieren → Deskriptive Statistiken → Deskriptive Statistik, Variablenauswahl vol und rate.

Eventuell hier, spätestens aber beim Punkt 4 (Grafikerstellung) müsste auffallen, dass

Deskriptive Statistik

	N	Minimum	Maximum	Mittelwert	Standardabw eichung
Eingeatmete Luftmenge	39	-,916291	1,308333	,15959000	,537651922
Geschwndigkeit, mit der die Luft eingeatmet wurde	39	-3,506558	604316,0000	15495,54000	96767,96301
Gültige Werte (Listenweise)	39				

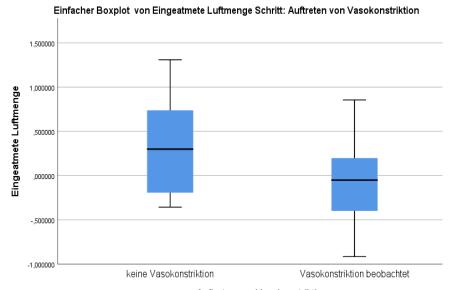
der Maximalwert von "rate" extrem hoch ist. Bei Betrachtung der Werte wird klar, dass das Komma vergessen wurde. Also: Wert 33 in

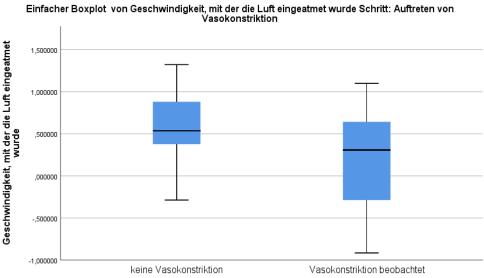
0,604316 ändern, dann sieht die gleiche Abfrage so aus:

	N	Minimum	Maximum	Mittelwert	Standardabw eichung
Eingeatmete Luftmenge	39	-,916291	1,308333	,15959000	,537651922
Geschwndigkeit, mit der die Luft eingeatmet wurde	39	-3,506558	1,321756	,27344592	,899923407
Gültige Werte (Listenweise)	39				

4) Erzeugen Sie Grafiken zu "vol" und "rate", um die Unterschiede in den Werten bei den beiden "vaso"-Gruppen zu zeigen.

Statistisch aussagekräftig wären hier z.B. Boxplots. Menüpunkt **Diagramme → Diagrammerstellung**, "Boxplots" auswählen, Bild "einfacher Boxplot" nach oben ziehen, x-Achse: "vaso", y-Achse: nacheinander "vol" und "rate".





Auftreten von Vasokonstriktion

Auftreten von Vasokonstriktion

Ubung 3

Was sagen die Boxplots aus?

Die Werte bei "vol" (Eingeatmete Luftmenge) scheinen bei beiden Ausprägungen ("vaso" 1 und 2) relativ normalverteilt zu sein (schwarzer Balken (Median) in der Mitte der "Box"), allerdings ist der eine "whisker" bei "vaso" 1 viel kürzer als der andere, es besteht hier also eine leichte Schiefe in der Verteilung. Die Varianzen scheinen relativ homogen zu sein (d.h. "Größe der

Boxen beider Gruppen ungefähr gleich").

Bei der "Geschwindigkeit mit der die Luft eingeatmet wurde" kann auch eine leichte Schiefe in der Verteilung beobachtet werden (schwarzer Balken nicht mittig), insgesamt dürfte dies aber noch als normalverteilt gelten. Varianzhomogenität würde man hier nicht mehr annehmen, da die Boxen doch recht unterschiedlich groß sind.

5) Kann statistisch ein Unterschied zwischen den Werten der Variablen "vol" und "rate" abgesichert werden hinsichtlich des Auftretens von Vasokonstriktionen?

Zuerst muss man einen geeigneten Test auswählen. Die bei Punkt 4 gemachten Aussagen zu Verteilungsannahme und Varianzhomogenität sind hierzu entscheidend. Zusätzlich gibt es Tests, die diese objektiven Aussagen absichern können.

Der Shapiro-Wilk und der KS-Test testet die Normalverteilungsannahme:

Analysieren → Deskriptive Statistik → Explorative Datenanalyse unter

Punkt "Diagramme" Häkchen bei "Normalverteilungsdiagramm mit

Tests" setzen.

Tests auf Normalverteilung

	Auftreten von	Koli	mogorov-Sn	nirnov ^a		Shapiro-Wil	k
	Vasokonstriktion	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Eingeatmete Luftmenge	keine Vasokonstriktion	,137	20	,200	,932	20	,171
	Vasokonstriktion beobachtet	,122	19	,200*	,980,	19	,945
Geschwndigkeit, mit der	keine Vasokonstriktion	,112	20	,200*	,961	20	,573
die Luft eingeatmet wurde	Vasokonstriktion beobachtet	,191	19	,068	,934	19	,204

^{*.} Dies ist eine untere Grenze der echten Signifikanz.

a. Signifikanzkorrektur nach Lilliefors

In unserem Fall haben wir eine Fallzahl <50, würden also eher auf den Shapiro-Wilk Test zurück greifen. Der p-Wert (Werte, die unter "Signifikanz" stehen) ist in allen Fällen größer als 0,05, also kann die Nullhypothese "die Daten folgen der Normalverteilung" nicht verworfen werden. Wir dürfen also parametrische Testverfahren verwenden (Normalverteilung wird vorausgesetzt). Der Test auf Varianzhomogenität (Levene Test) wird über die Auswahl von "nicht transformiert" bei "Streubreite vs. Mittleres Niveau mit Levene Test" in der Explorativen Datenanalyse durch geführt (aber auch automatisch beim Durchführen eines T-Tests). Analysieren → Mittelwerte vergleichen → T-Test bei unabhängigen Stichproben. Vaso ist die Gruppenvariable, vol und rate die Testvariablen.

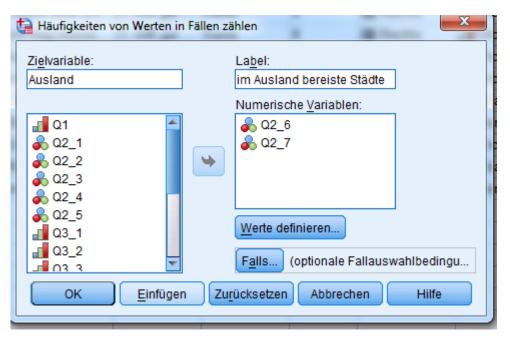
Man erhält folgende Ausgabe:

	Test bei unabhängigen Stichproben										
			Levene- Varianzg				T-	Test für die Mittelv	wertgleichheit		
								Mittlere	Standardfehle	95% Konfider Diffe	
∡l			F	Signifikanz	Т	df	Sig. (2-seitig)	Differenz	r der Differenz	Untere	Obere
7	Eingeatmete Luftmenge	Varianzen sind gleich	1,565	,219	2,704	37	,010	,431361142	,159500823	,108181777	,754540507
		Varianzen sind nicht gleich			2,718	36,262	,010	,431361142	,158686252	,109611155	,753111129
	Geschwndigkeit, mit der	Varianzen sind gleich	1,818	,186	2,140	37	,039	,362977463	,169593864	,019347655	,706607271
	die Luft eingeatmet wurde	Varianzen sind nicht gleich			2,127	34,221	,041	,362977463	,170641216	,016275211	,709679715

Zuerst besagt der Levene-Test der Varianzgleichheit, dass bei "Eingeatmete Luftmenge" Varianzgleichheit angenommen werden kann (p-Wert > 0,2), wir nehmen also die Student T-Test Statistik (Wert obere Zeile), also 0,01, das heißt, es bestehen signifikante Unterschiede zwischen den Vasokonstriktions-Gruppen. Bei der "Geschwindigkeit,…" ist der p-Wert des Levene's Test <0,2, wir gehen also nicht von homogenen Varianzen aus, nehmen daher die Welch T-Test Statistik (untere Zeile). Hier ist der p-Wert 0,041, also auch hier bestehen signifikante Unterschiede zwischen den beiden Gruppen auf dem 5% Niveau.

2) Erstellen Sie eine neue Variable "Ausland", diese soll die Summe der im Ausland bereisten Städte (Paris, Madrid) wider spiegeln.

z.B. über Transformieren -> Werte in Fällen zählen



dann noch bei der Schaltfläche "Werte definieren" die "1" angeben.

Alternativ kann man auch

→ Variable berechnen

wählen und dann die Summe

von Q2 6 und Q2 7 wählen.

3) Wie häufig wurden die einzelnen Städte bereist?

Unelegante Lösung: Analysieren → Deskriptive Statistiken →

Häufigkeiten, Auswahl der Variablen Q2_1 bis Q2_7

Häufigkeitstabelle

Kassel

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	nein	12	60,0	60,0	60,0
	ja	8	40,0	40,0	100,0
	Gesamt	20	100,0	100,0	

Hannover

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	nein	5	25,0	25,0	25,0
	ja	15	75,0	75,0	100,0
	Gesamt	20	100,0	100,0	

Hamburg

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	nein	1	5,0	5,0	5,0
	ja	19	95,0	95,0	100,0
	Gesamt	20	100,0	100,0	

München

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	nein	9	45,0	45,0	45,0
	ja	11	55,0	55,0	100,0
	Gesamt	20	100,0	100,0	

Berlin

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	nein	2	10,0	10,0	10,0
	ja	18	90,0	90,0	100,0
	Gesamt	20	100,0	100,0	

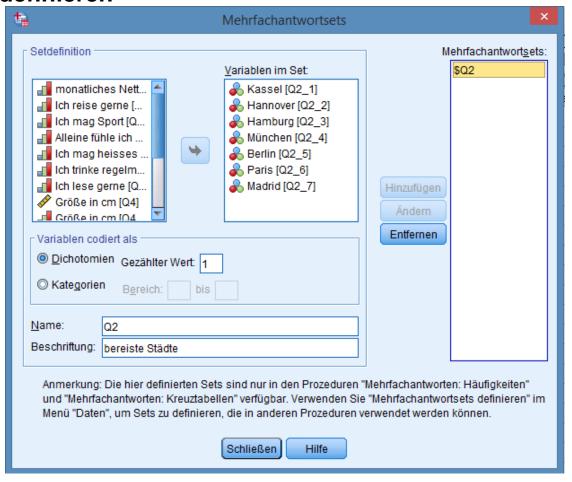
Paris

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	nein	8	40,0	40,0	40,0
	ja	12	60,0	60,0	100,0
	Gesamt	20	100,0	100,0	

Madrid

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	nein	12	60,0	60,0	60,0
	ja	8	40,0	40,0	100,0
	Gesamt	20	100,0	100,0	

3ff) elegante Lösung: **Analysieren** → **Mehrfachantworten** → **Variablensets definieren**



3ff) Dann Analysieren → Mehrfachantworten → Häufigkeiten

Das Set Q2, auswählen, also in den rechten Kasten schieben. Ergebnis:

Häufigkeiten von \$Q2

		Antwo	orten	Prozent der Fälle	
		Z	Prozent		
bereiste Städte ^a	Kassel	19	10,6%	47,5%	
	Hannover	30	16,7%	75,0%	
	Hamburg	33	18,3%	82,5%	
	München	25	13,9%	62,5%	
	Berlin	30	16,7%	75,0%	
	Paris	24	13,3%	60,0%	
	Madrid	19	10,6%	47,5%	
Gesamt		180	100,0%	450,0%	

a. Dichotomie-Gruppe tabellarisch dargestellt bei Wert 1.

4) Erstellen Sie eine neue Variable "fitness". Diese soll bei allen vorab "2" betragen. Wenn Q3_2 mit 1 oder 2 beantwortet wurde erfolgt eine Aufwertung auf "3", wenn Q3_5 mit 1 oder 2 beantwortet wurde erfolgt eine Abwertung um 1 des vorherigen Wertes

Dies geht über **Umkodieren in andere Variablen**. Ausgangsvariable ist Q3_2, die neue Variable soll "fitness" heißen. Die alten und neuen Werte definiert man folgendermaßen (man setzt hier quasi in einem

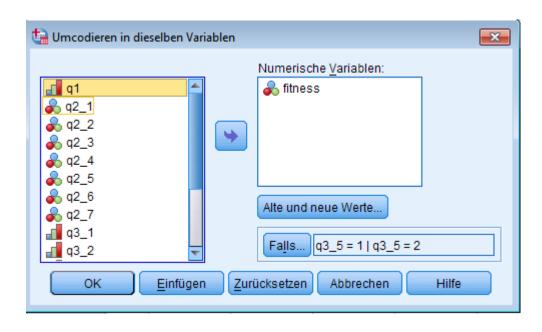


Schritt die Werte auf 2 und führt die Aufwertung auf 3 durch, wenn Q3_2 mit oder 2 beantwortet wurde).

4ff) Nach Erzeugung der Variable fitness führt man die Abwertung durch: **Umkodieren in dieselben Variablen**, Ausgangsvariable ist fitness. Unter "Alte und neue Werte" gibt man an,

dass aus einer 2 eine 1 und aus einer 3 eine 2 werden soll, "Falls…" Q3_5 mit 1 oder 2 beantwortet wurde.

Man könnte dies auch über die Syntax lösen, s. nächste Seite



4ff) *Erzeugen der Variable fitness, alle Werte erhalten eine 2. COMPUTE fitness=2.

*Aufwertung durch Frage Q3_2.

IF $(Q3_2 = 1 | Q3_2 = 2)$ fitness= fitness+1.

*Abwertung durch Frage Q3_5.

IF (Q3_5=1|Q3_5=2) fitness=fitness-1.

EXECUTE.

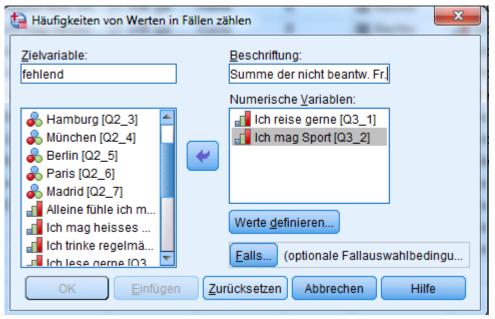
5) Einfügen der Antwortmöglichkeit "weiß nicht" (6): In der Variablenansicht auf die Spalte "Werte" auf die Zeile "Q3_1" klicken, hier den entsprechenden Wert und die Beschriftung einfügen.

Errechnen des Durchschnittswertes: Den Durchschnittswert ("Mittelwert") errechnet man z.B. über Analysieren → Deskriptive Statistik

Zu beachten ist allerdings, dass die Antworten mit einer "6" (weiß nicht) in diese Berechnung mit eingehen, was nicht sinnvoll ist, da ja nur die Skala 1-5 eine Bewertung an sich ist. Um das zu vermeiden, muss man in der Variablenansicht unter "Fehlend" die "6" als fehlenden Wert definieren! Möchte man im folgenden Balkendiagramm für Q3_1 den Balken für die "6" mit angezeigt bekommen, muss man dieses wieder rückgängig machen!

6) Aufsummieren der nicht beantworteten Fragen:

Transformieren → Werte in Fällen zählen



Dann über die Schaltfläche "Werte definieren" als zu zählende Werte die "6" als einzelnen fehlenden Wert und "systemdefiniert fehlend" (SYSMIS) angeben.

7) Kreuztabelle zu der Einkommensgruppe und der Anzahl der im Ausland besuchten Städte; Trend/ Zusammenhang?

Analysieren → Deskriptive Statistiken → Kreuztabellen
Zeile "Q1", Spalte "Ausland", Häkchen setzen bei "Statistiken" →
ChiQuadrat und bei "Zellen" → Erwartet.

monatliches Nettoeinkommen * im Ausland bereiste Städte Kreuztabelle

			im Ausland bereiste Städte			
			0	1	2	Gesamt
monatliches Nettoeinkommen	weniger als 800€	Anzahl	2	2	0	4
		Erwartete Anzahl	,8	1,9	1,2	4,0
	800 bis 1200 €	Anzahl	5	2	0	7
		Erwartete Anzahl	1,4	3,4	2,2	7,0
	1200 bis 1500	Anzahl	1	5	5	11
		Erwartete Anzahl	2,3	5,4	3,4	11,0
	1500 bis 2000	Anzahl	0	5	2	7
		Erwartete Anzahl	1,4	3,4	2,2	7,0
	mehr als 2000	Anzahl	0	5	5	10
		Erwartete Anzahl	2,1	4,9	3,1	10,0
Gesamt		Anzahl	8	19	12	39
		Erwartete Anzahl	8,0	19,0	12,0	39,0

7ff) Der Chi-Quadrat-Test zeigt einen signifikanten Zusammenhang zwischen der Einkommensgruppe und der im Ausland besuchten Städte an (p<0,05).

Chi-Quadrat-Tests

	Wert	df	Asymptotisch e Signifikanz (zweiseitig)
Chi-Quadrat nach Pearson	21,450 ^a	8	,006
Likelihood-Quotient	24,236	8	,002
Zusammenhang linear- mit-linear	11,433	1	,001
Anzahl der gültigen Fälle	39		

a. 14 Zellen (93,3%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist .82.

Allerdings haben quasi alle Zellen eine erwartete Häufigkeit <5 und man sollte dieses Ergebnis nicht verwenden. Die Zusammenfassung von Einkommensgruppen könnte hier helfen.

1) Die Vermutung liegt nahe, dass die Körpergröße (gross) eng mit der Vitalkapazität (fvc) zusammenhängt. Zeigen Sie dies durch eine Korrelationsanalyse.

Bei vorherigen Analysen hatten wir festgestellt, dass die Werte der Vitalkapazität nicht normalverteilt sind. Man würde daher eher auf einen nicht-parametrischen Korrelationskoeffizienten

zurück greifen, z.B. Spearman.

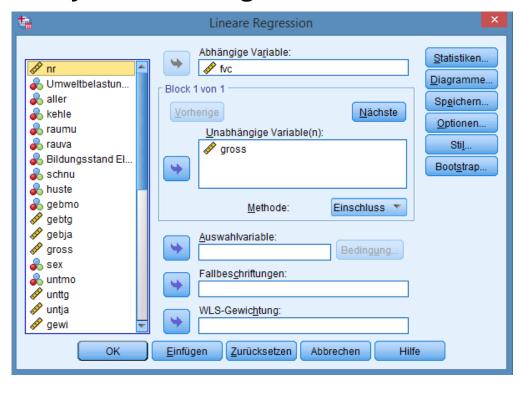
Analysieren → Korrelation → bivariat

Dies ergibt einen Korrelationskoeffizienten von 0,941, der auf dem 0,1%-Niveau signifikant ist. Es liegt also eine sehr starke, positive Korrelation vor.



Lineare Regression von gross und fvc

Analysieren → Regression → Linear



Unter "Diagramme" noch das Residuendiagramm (ZPRED gegen ZRESID) sowie ein Normalverteilungsdiagramm der standardisierten Residuen anfertigen lassen und unter "Statistiken" den Durbin-Watson Test auswählen.

Man erhält ein R² von 0,862, der Wert der Durbin-Watson Statistik von 1,853 weist nicht auf eine Autokorrelation der Daten hin.

Das Modell lautet fvc=-0,51*gross - 4,675

Das P-P-Diagramm zeigt, dass die Residuen relativ normal verteilt sind.

Das Streudiagramm der standardisierten Residuen gegen die standardisierten geschätzten Werte zeigt allerdings wieder eine gekrümmte Punktwolke.

Genau genommen empfiehlt es sich daher, es mit einer Linearisierung der Variable fvc zu versuchen.

Transformieren → **Variable berechnen**, Verwendung von z.B. lg10 als Funktion. Die Verwendung der Variable log_fvc in der Regression führt zu einem R² von 0,885 unter Erfüllung der Voraussetzung an die Residuen.