

SPSS online Teil 2

1 Einfache Testverfahren

Auswahl eines geeigneten Tests zum Vergleich von Daten:

- welcher Art sind die Daten (nominal, ordinal, metrisch,...)

Nominal: ein geeigneter Test bei nominalen Daten ist z.B. der Chi-Quadrat-Test (vgl. Kreuztabellen im Kapitel Datenbeschreibung).

Ordinal: die Tests von den nominalen Daten dürfen auch verwendet werden, man verzichtet aber auf Informationen, die die Daten hergeben. Empfehlenswert sind i.d.R. sog. **nicht-parametrische Tests**, die meist auf Rängen beruhen.

Metrisch: sind die Daten normalverteilt (und varianzhomogen), dürfen sog. **parametrische Tests** verwendet werden; sind sie dies nicht, verwendet man **nicht-parametrische** Verfahren.

2

Einfache Testverfahren

Dann besteht auch die Frage nach dem geeigneten Testverfahren für den Vergleich:

- Hat man **eine Datenreihe**, bei der man untersuchen möchte, ob sie von einem bekannten Wert abweicht → Einstichproben-Test (T-Test bei NV, sonst Mann-Whitney-U Test (entspricht Wilcoxon-Test))
- Hat man eine **gepaarte Stichprobe** (z.B. vorher/ nachher Messungen, wiederholte Messung an einer Versuchseinheit) → Test für Stichproben mit paarigen Werten (T-Test bei NV, sonst MWU/ Wilcoxon Test)
- Hat man **zwei Datenreihen** (ordinal oder metrisch), zwischen denen man einen Zusammenhang vermutet → Korrelationsanalyse
- Hat man einen Wert in unterschiedlichen Gruppen gemessen und möchte **Gruppenunterschiede** aufzeigen → Test für unabhängige Stichproben (bei 2 Gruppen: T-Test bei NV, sonst MWU- Test; bei >2 Gruppen ANOVA bei NV, sonst Kruskal-Wallis-Test)

3

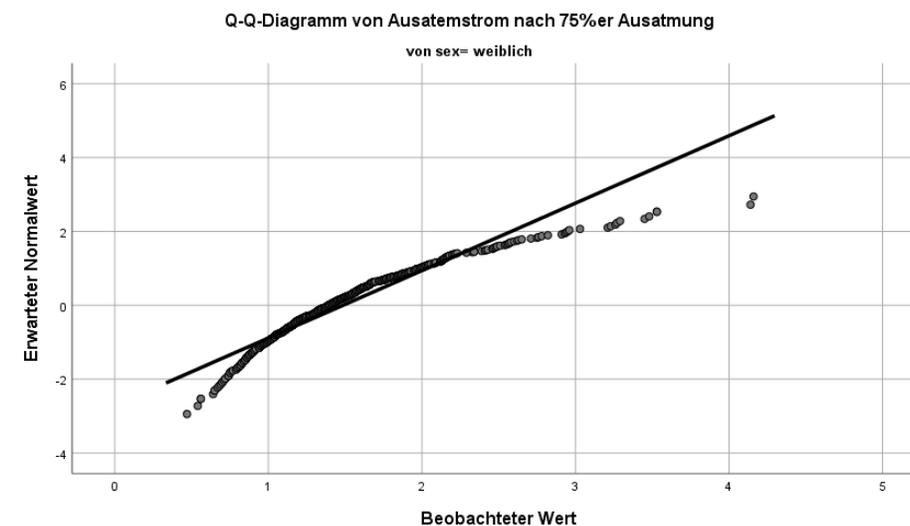
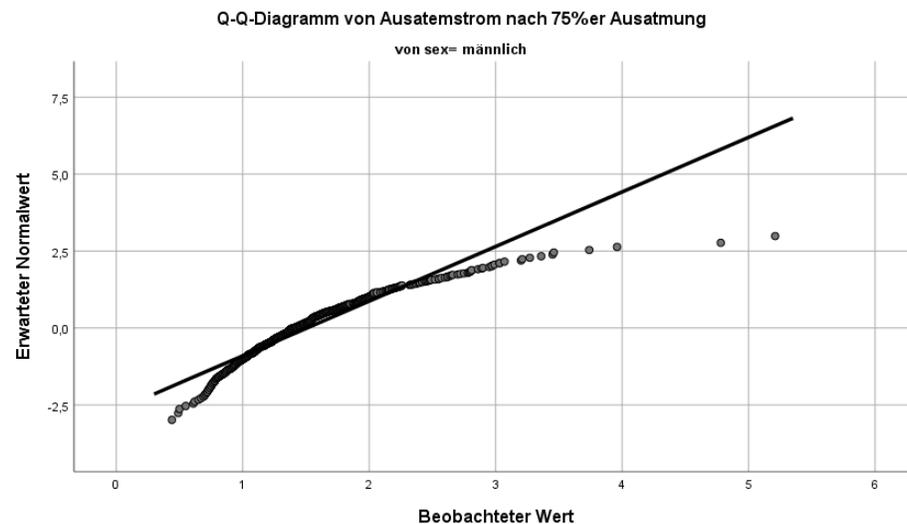
Einfache Testverfahren

Beispiel: Unterscheiden sich Jungen und Mädchen hinsichtlich des Ausatemstroms nach 75%iger Ausatmung (fef75)?

Der fef75 ist metrisch, also prüfen wir, ob dieser normalverteilt ist: **Analysieren** → **Deskriptive Statistiken** → **Explorative Datenanalyse**

Abhängige Variable: fef75, Faktorenliste: sex

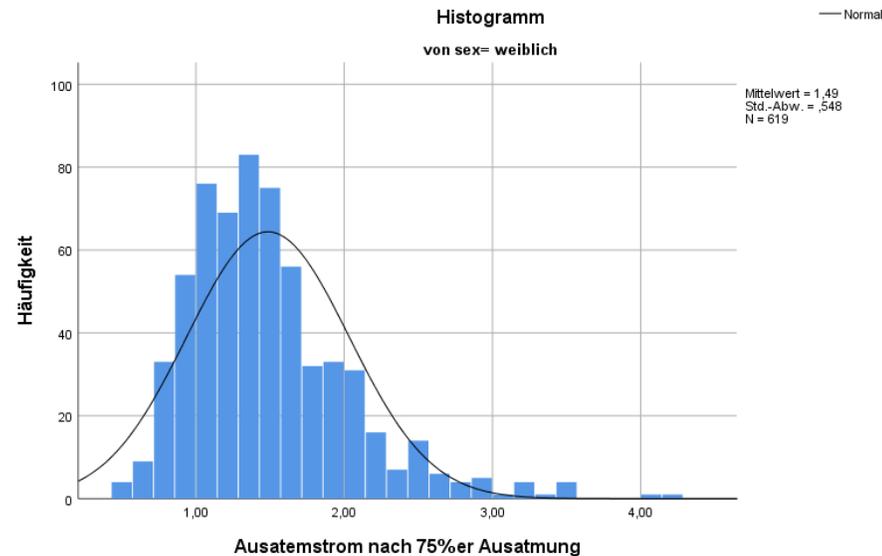
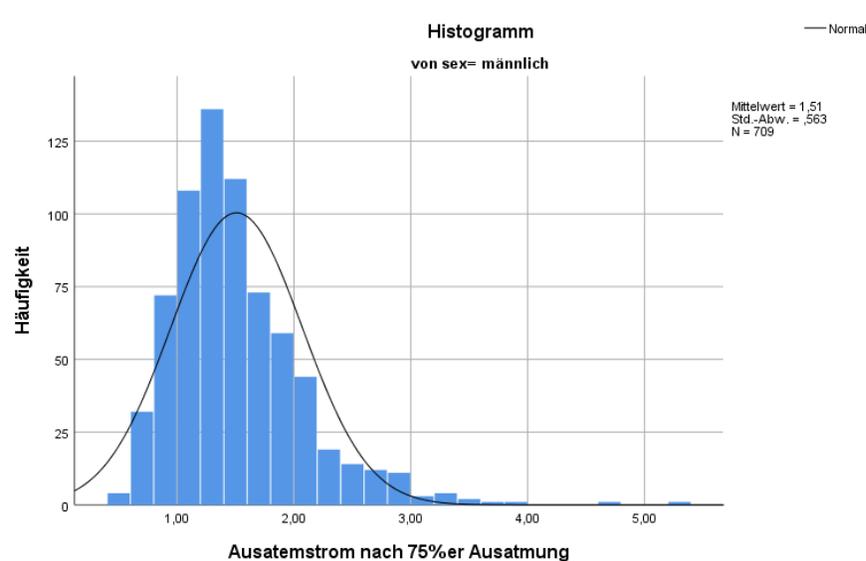
Auswertung: Beide **Q-Q-plots** zeigen deutliche Abweichung von der Normalverteilung (Punkte weichen von Linie ab)



4

Einfache Testverfahren

Die **Histogramme** zeigen eine rechtsschiefe („**Schiefe**“-Wert $\gg 0$) und steilgipflige („**Kurtosis**“-Wert $\gg 0$) Verteilung.



Ein weiterer Hinweis auf eine Abweichung von der Normalverteilung ist ein deutlicher Unterschied zwischen Mittelwert und Median (s. Tabelle Deskriptive St., hier findet man auch die Schiefe und Kurtosis-Werte)

5

Einfache Testverfahren

Man würde hier also nicht von normalverteilten Stichproben ausgehen, das Ergebnis des **Shapiro-Wilk** Tests bestätigt dies (p-Wert < 0,05).

!!!Beachten: Anwendung des zentralen Grenzwertsatzes wäre möglich!!!

Wir wählen einen nichtparametrischen Test:

Analysieren → Nichtparametrische Tests → unabhängige Stichproben.

Testvariable ist „fef75“, Gruppierungsvariable ist „sex“.

SPSS verwendet den Mann-Whitney-U-Test (entspricht Wilcoxon-Test) für den Vergleich der zwei Stichproben.

Ergebnis: Der p-Wert von 0,414 ist deutlich über dem Signifikanzniveau von 0,05, daher können keine geschlechterbedingten Unterschiede hinsichtlich des fef75 festgestellt werden.

6

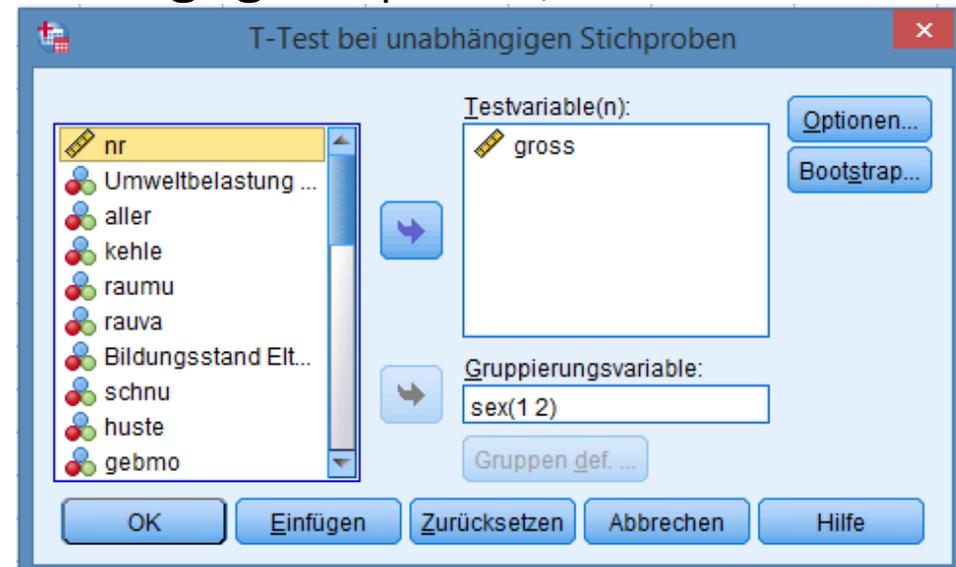
Einfache Testverfahren

Beispiel: Unterscheiden sich Jungen und Mädchen hinsichtlich der Körpergröße ?

Werte auf NV testen: **Analysieren** → **Deskriptive Statistik** → **Explorative Datenanalyse** Visuelle Beurteilung zeigt ungefähr Normalverteilung, auch wenn das Ergebnis des Shapiro-Wilk Tests dagegen spricht, würden wir hier ein parametrisches Verfahren wählen.

Analysieren → **Mittelwerte vergleichen** → **T-Test bei unabhängigen Stichproben**

Bei der Gruppierungsvariable noch die Gruppen definieren (1 und 2) (T-Test geht immer nur für zwei Gruppen!).



7

Einfache Testverfahren

Das Ergebnis sieht folgendermaßen aus:

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		t-Test für die Mittelwertgleichheit							
		F	Sig.	T	df	Signifikanz		Mittlere Differenz	Differenz für Standardfehler	95% Konfidenzintervall der Differenz	
						Einseitiges p	Zweiseitiges p			Unterer Wert	Oberer Wert
Körpergröße	Varianzen sind gleich	1,764	,184	4,782	1547	<,001	<,001	3,632	,760	2,142	5,122
	Varianzen sind nicht gleich			4,793	1546,724	<,001	<,001	3,632	,758	2,146	5,118



Sind Varianzen gleich (s. Levene-Test, Signifikanz $>0,2$), nimmt man das Ergebnis der oberen Zeile (Student's T-Test).

Sind Varianzen ungleich (Levene-Test, Signifikanz $<0,2$) nimmt man die untere Zeile (T-Test nach Welch).

Hier (Levene $<0,2$) würde man also das Ergebnis der unteren Zeile nehmen:
 „Mädchen und Jungen unterscheiden sich signifikant hinsichtlich der Körpergröße (p-Wert < 0.001)“.

8

Einfache Testverfahren

Vergleich eines Merkmals zwischen >2 Gruppen:
Daten in den Gruppen normalverteilt und varianzhomogen? **NEIN:**

Analysieren → Nicht-parametrische Tests → unabhängige Stichproben, das Programm verwendet automatisch den Kruskal-Wallis Test.

Beispiel: Unterscheidet sich die Vitalkapazität (fvc) der Kinder in den drei Wohngebieten (zone)?

Analysieren → Nicht-parametrische Tests → unabhängige Stichproben,
 Gruppierungsvariable zone,
 Testvariable fvc.

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von fvc ist über die Kategorien von Umweltbelastung am Wohnort identisch.	Kruskal-Wallis-Test bei unabhängigen Stichproben	,533	Nullhypothese beibehalten

Asymptotische Signifikanzniveaus werden angezeigt. Das Signifikanzniveau ist ,05.

In diesem Fall kann also kein Unterschied zwischen den Wohngebieten hinsichtlich der Vitalkapazität festgestellt werden (p-Wert > 0.05).

9

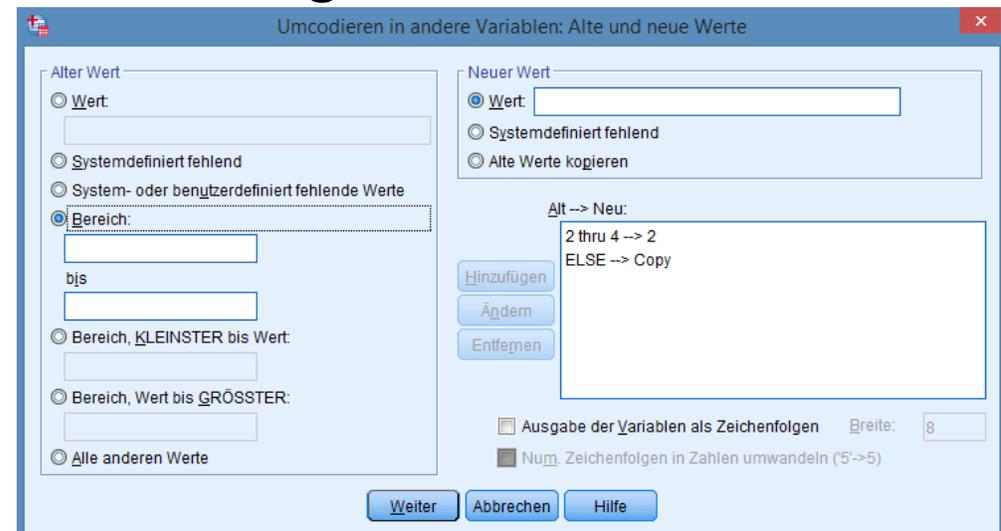
Einfache Testverfahren

Vergleich eines Merkmals zwischen >2 Gruppen:
Daten in den Gruppen normalverteilt und varianzhomogen? JA:

Analysieren → Mittelwerte vergleichen → Einfaktorielle Varianzanalyse

Beispiel: Zeigen die Gruppen mit unterschiedlicher Anzahl an Vorerkrankungen (Anz_Vorerkr) Unterschiede im max. Ausatemstrom pef?

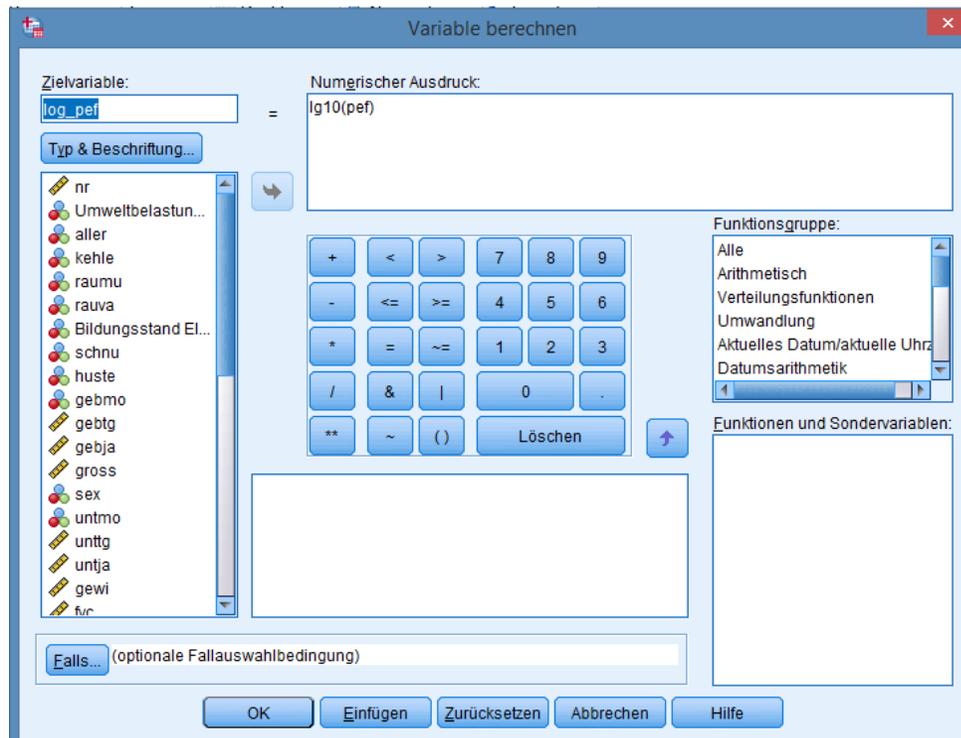
Vorarbeit 1: „balanciertes Design“: da nur sehr wenige Fälle mit 3 und 4 Vorerkrankungen existieren, werden diese mit der Gruppe „2 Vorerkrankungen“ zusammengefasst: **Transformieren → Umcodieren in andere Variablen**, neue Variable soll ANZ3 heißen; man möchte insgesamt die Gruppen 0, 1 und 2 erhalten.



10

Einfache Testverfahren

Vorarbeit 2: Die pef Werte sind nicht normalverteilt. Um dies zu erreichen, kann man die Daten transformieren, z.B. über eine Logarithmierung:



Dann untersucht man auf NV:
Analysieren → Deskriptive Statistiken
→ Explorative Datenanalyse,
abhängige Variable log_pef, Faktor
ANZ3

11

Einfache Testverfahren

Analysieren → Mittelwerte vergleichen → Einfaktorielle Varianzanalyse.

Testvariable log_pef, Faktor ANZ3.

Der p-Wert von 0,042 zeigt, dass signifikante Unterschiede zwischen den Gruppen hinsichtlich des log_pef Wertes vorliegen.

Welche Gruppen genau unterscheiden sich voneinander?

→ **Post-hoc Test:**

Zurück zu der Analyse, bei Schaltfläche „Post hoc“ geeigneten Test auswählen.

In unserem Fall würde man z.B. den Bonferroni-Test verwenden können.

Dieser zeigt signifikante Unterschiede nur zwischen den Gruppen 0 und 2 (p-Wert 0.039).

Übung 3

Folgender Versuch soll ausgewertet werden:

Datenarchiv des Instituts für Statistik und des SFB 386

Vasokonstriktion

Beschreibung des Datensatzes

[Seitenanfang](#)

Die Datei enthält binäre Daten. Drei Versuchspersonen atmeten bestimmte Luftmengen in verschiedenen Geschwindigkeiten ein. In einigen Fällen konnte daraufhin eine sog. Vasokonstriktion (von den Nerven gesteuerte Verengung der Blutgefäße) an deren Haut festgestellt werden. Ziel der Studie ist es, einen Zusammenhang zwischen Atmung und Vasokonstriktion aufzuzeigen. Die drei Versuchspersonen wiederholten diesen Test 9, 8 bzw. 22 mal, so daß die Datei insgesamt 39 Beobachtungen enthält.

Variablenbeschreibung

[Seitenanfang](#)

vol	Eingeatmete Luftmenge
rate	Geschwindigkeit, mit der die Luft eingeatmet wurde
vaso	Zustand der Blutgefäße
	1 : es wurde keine Vasokonstriktion beobachtet
	2 : es wurde Vasokonstriktion beobachtet

Übung 3

- 1) Öffnen Sie die Excel-Datei „Vaso.xlsx“ in SPSS
- 2) Überprüfen und vervollständigen Sie die Variablenübersicht
- 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?
- 4) Erzeugen Sie Grafiken zu „vol“ und „rate“, um die Unterschiede in den Werten bei den beiden „vaso“-Gruppen zu zeigen.
- 5) Kann statistisch ein Unterschied zwischen den beiden Gruppen mit/ ohne Vasokonstriktion abgesichert werden bei den Werten der Variablen „vol“ und „rate“?

12

Exkurs: Mehrfachantworten

In Fragebögen finden sich häufig Mehrfachantworten, d.h. bei einer Frage können mehrere Kreuze gesetzt werden. Jede Antwortmöglichkeit muss in SPSS als eigene Variable angelegt werden, es gibt dann wieder die Möglichkeit, diese als Mehrfachantwortset zusammenzufassen.

Im Atemwegsdatensatz bieten die „Vorerkrankungsfragen“ die Möglichkeit, diese zusammenzufassen, um sie so zusammen auswerten zu können.

Beachten: historisch bedingt gibt es zwei unterschiedliche Wege, Mehrfachantwortsets zu definieren!

13

Exkurs: Mehrfachantworten

Daten → Mehrfachantwortsets definieren. Diese Sets können dann z.B. für Diagramme verwendet werden.

Variablenkodierung: hier Dichotomien, gezählt werden soll der Wert „1“, also „ja“

Mehrfachantwortsets

Setdefinition

Variablen im Set:

Hier definierte Sets sind in den Prozeduren "Mehrfachantworten: Häufigkeiten" und "Mehrfachantworten: Kreuztabellen" nicht verfügbar.

Mehrfachantwortsets:

\$Vorerkr

Variablenkodierung

Dichotomien Gezählter Wert: 1

Kategorien

Quelle der Kategoriebeschriftungen

Variablenbeschriftungen

Beschriftungen des gezählten Werts

Variablenbeschriftung als Setbeschriftung verwenden

SetName: Vorerkr

Setbeschriftung: Vorerkrankungen

Hinzufügen

Ändern

Entfernen

OK Einfügen Zurücksetzen Abbrechen Hilfe

14

Exkurs: Mehrfachantworten

Analysieren → Mehrfachantworten → Variablen-Sets definieren.

Nach Definition eines Sets (Dialogfenster quasi genauso wie letzte Folie) kann man sich dann unter dem gleichen Menüpunkt Häufigkeiten und Kreuztabellen zu den erstellten Variablensets ausgeben lassen.

Beispiel:

Kreuztabelle rauva*\$Voerkr^a

			Vorerkrankungen ^b				
			Allergische Atemwegserk rankung	Kehlkopfentzü ndung	häufiger Schnupfen	häufiger Husten	Gesamt
Vater Raucher	nein	Anzahl	146	134	205	118	603
		Innerhalb rauva%	24,2%	22,2%	34,0%	19,6%	
	ja	Anzahl	41	33	77	43	194
		Innerhalb rauva%	21,1%	17,0%	39,7%	22,2%	
Gesamt		Anzahl	187	167	282	161	797

Prozentsätze und Gesamtwerte beruhen auf den Antworten.

a. Es sind nicht genug (weniger als 2) Mehrfachantworten-Gruppen für eine Paarbildung vorhanden. Die Prozentsätze beruhen auf den Antworten, es wird jedoch keine Paarbildung durchgeführt.

b. Dichotomie-Gruppe tabellarisch dargestellt bei Wert 1.

15

Exkurs 2: Fragebögen

Der folgende (fiktive!) Fragebogen soll in SPSS ausgewertet werden

Q1: Wie hoch ist Ihr monatliches Nettoeinkommen?		Q2: In welchen dieser Städte waren Sie bereits?	
<input type="checkbox"/>	Weniger als 800 €	<input type="checkbox"/>	Kassel
<input type="checkbox"/>	800 bis 1,200 €	<input type="checkbox"/>	Hannover
<input type="checkbox"/>	1,200 bis 1,500 €	<input type="checkbox"/>	Hamburg
<input type="checkbox"/>	1,500 bis 2,000 €	<input type="checkbox"/>	München
<input type="checkbox"/>	Mehr als 2,000 €	<input type="checkbox"/>	Berlin
		<input type="checkbox"/>	Paris
		<input type="checkbox"/>	Madrid

Q1 (Frage 1): Einfachantwort, man braucht eine Zeile in SPSS. Messung kategorisch, also Kategorie 1-5 möglich.

Q2 Frage 2: Mehrfachantwort möglich, man braucht also pro Abfrage (hier: Stadt) eine Spalte in SPSS (hier 7). Antwortmöglichkeiten sind nein (=0) und ja (=1).

16

Exkurs 2: Fragebögen

Q3: In welchem Maß treffen diese Aussagen auf Sie zu?

	Trifft voll zu			Trifft überhaupt nicht zu	
Ich reise gerne	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich mag Sport	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Alleine fühle ich mich unwohl	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich mag heißes Wetter	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich trinke regelmäßig Alkohol	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Ich lese gerne	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Q4: Wie groß sind Sie? Angaben in cm

Körpergröße in cm

Q5: Was ist Ihr Beruf?

Bezeichnung _____

Q3 (Frage 3): beinhaltet 6 Abfragen mit jeweils 5 Antwortkategorien, es werden also 6 Zeilen in SPSS benötigt. Hier als Wertebeschriftung „1= trifft voll zu“ bis „5=trifft überhaupt nicht zu“.

Q4 (Frage 4): offene Antwort, Angabe ist metrisch, also Skalenniveau.

Q5 (Frage 5): auch offene Antwort, aber nominal mit Zeichenfolge.

17

Exkurs 2: Fragebögen, Übung

1. Erstellen Sie in der Datei Fragebogen.sav eine neue Variable "Ausland", diese soll die Summe der im Ausland bereisten Städte (Paris, Madrid), Frage Q2, wider spiegeln.
2. Wie häufig wurden die einzelnen Städte bereist (Frage Q2) (Mehrfachantwortset!)?
3. 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.
4. Fügen Sie die Antwortmöglichkeit "weiß nicht" (6) bei den Fragen zu Q3 ein. Errechnen Sie den sinnigen (!) Durchschnittswert (?Mittelwert) für Q3_1 und Q3_2. Stellen Sie Q3_1 als Balkendiagramm dar.
5. Sie möchten die "Antwortverweigerer" besser erkennen. Erstellen Sie eine Variable, in der von Q3_1 und Q3_2 alle fehlenden Werte (Antwort=6 und fehlend) zusammengefasst werden.
6. Erstellen Sie eine Kreuztabelle der Einkommensgruppe zu der Anzahl der im Ausland besuchten Städte. Können Sie einen Trend/ Zusammenhang erkennen?