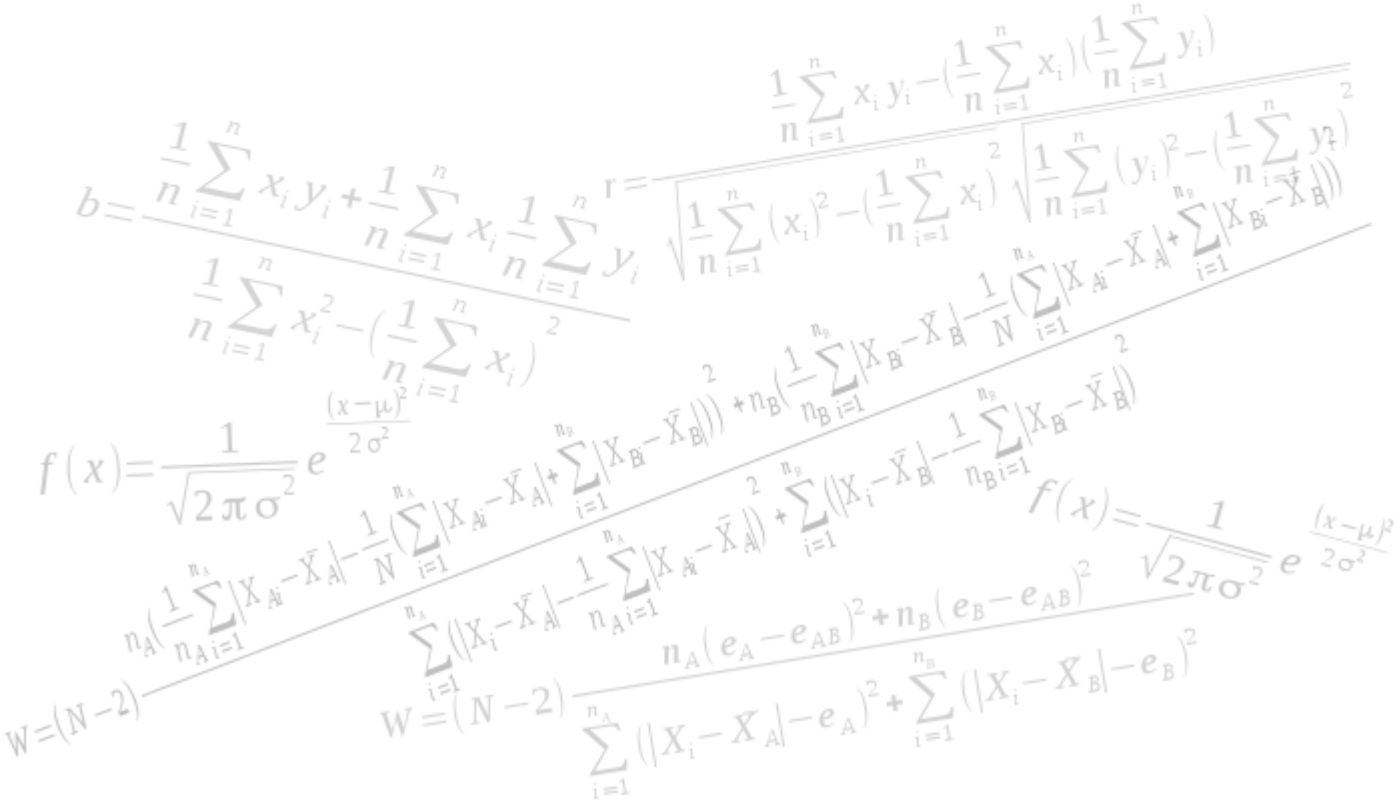
Übung WS 2023/2024

universität wien

Übung
Statistik
für Pflegewissenschaft

Isabella Hager

UE 1 SWS 2 ECTS



…zu dieser Lehrveranstaltung….

Die Übung „Statistik für Pflegewissenschaft“ gibt eine Einführung und Überblick über statistische Grundkenntnisse und Verfahren. Das **Handwerkszeug** sowie auch eine **kritische Interpretation** statistischer Ergebnisse stehen dabei im Mittelpunkt. Dabei werden Echtdaten aus einer Befragung 2009 von Langzeitarbeitslosen zur deren gesundheitlicher Situation (Gesundheit_AL_ue.sav) verwendet. Daten aus einer Studie über Patient*innen mit Stuhlinkontinenz am Rehab-Zentrum Weißer Hof (dm_gekürzt.sav) dienen zusätzlich als Übungsdatensatz.
Diese Unterlage ist als Arbeitsheft gedacht, bitte sämtliche Anmerkungen selbst ergänzen.

Viel Vergnügen in der Welt der Quantitäten !!!
Oktober 2023, Hager Isabella

Kennzeichnungen in dieser Arbeitsunterlage

SPSS-Befehl

Besonders **wichtige** Anmerkung

Interpretation

Pflichtaufgabe

Fleißaufgabe für Lernwütige mit **außerordentlichem** Lernerfahrungswert!

Termine der Fragestunden:

8. November 2023 – 13 bis 15 Uhr (bei Bedarf länger)
29. November 2023 – 13 bis 15 Uhr (bei Bedarf länger)
6. Dezember 2023 – 15:45 bis 17:45 Uhr (bei Bedarf länger)

Inhalt

UE 1

UE 2

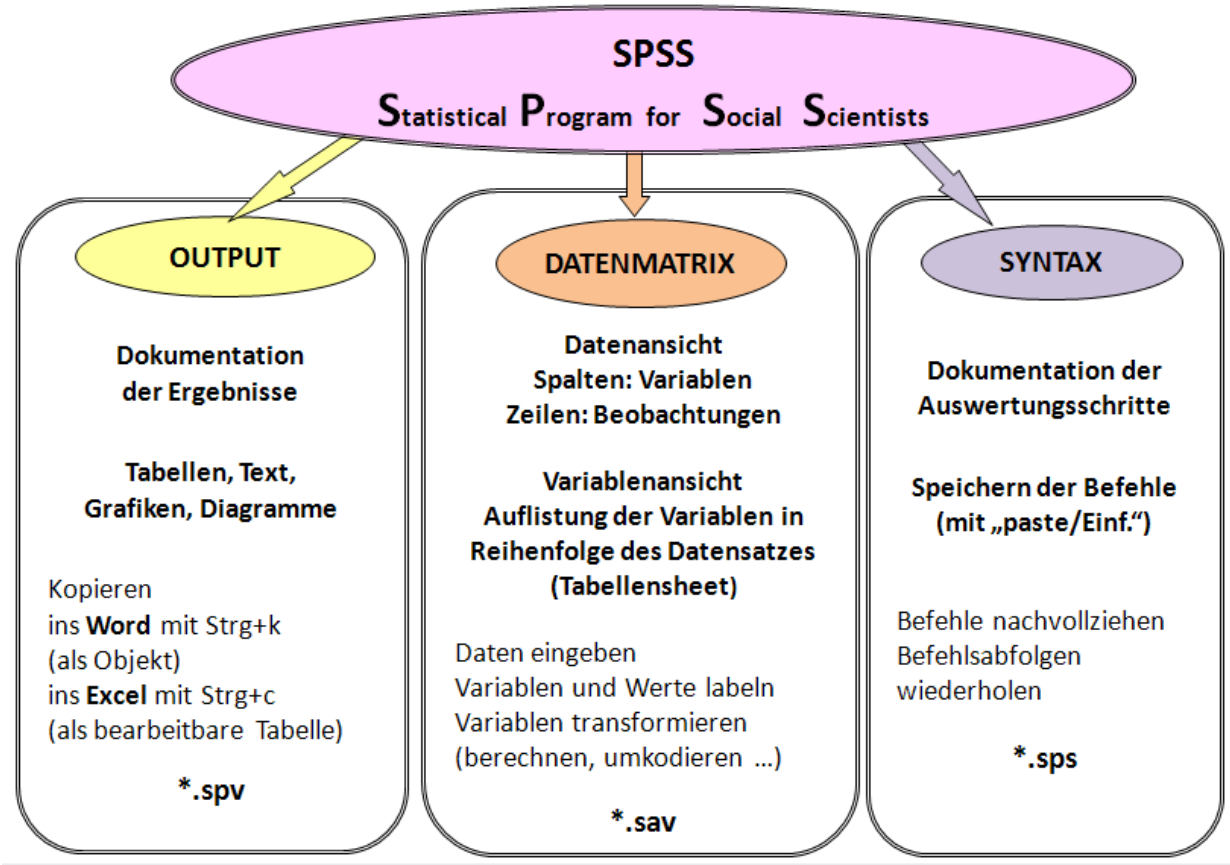
UE 3

UE 4

1. Arbeiten mit SPSS

1.1 Datentypen und Aufbau von SPSS

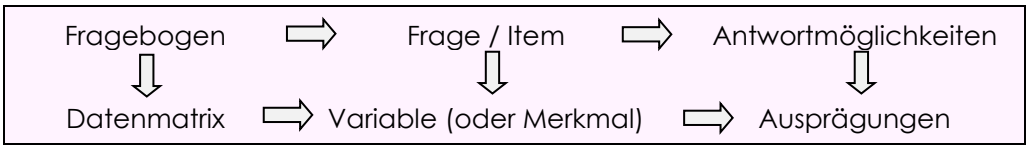
Grundsätzlich unterscheiden wir im SPSS drei **Datentypen**:



Datenmatrix Datenansicht	Dateneingabe: Hier wird der Inhalt der Datendatei angezeigt. Im Daten-Editor können neue Datendateien erstellt und vorhandene Datendateien bearbeitet werden. Das Fenster des Daten-Editors wird automatisch beim Start des SPSS-Programms geöffnet.
Datenmatrix Variablenansicht	Variablen definieren: Hier werden alle Variablen definiert und in der Reihenfolge des Dateneingabe angezeigt. Durch Kopieren eines Zelleninhaltes (Strg+C) können z.B. die Wertelabels einer Variable für eine andere übernommen werden (Strg+V).
Output/Viewer	Datenausgabe: Alle statistischen Ergebnisse, Tabellen, Diagramme werden im Viewer angezeigt. Ausgaben können hier bearbeitet und gespeichert werden. Das Viewer-Fenster wird automatisch geöffnet, wenn durch eine Prozedur die erste Ausgabe erzeugt wird.
Tabellen	Hier gibt es vielseitige Möglichkeiten um Ausgabetabellen zu bearbeiten.
Diagramme	In Diagrammfenstern können Diagramme und Grafiken bearbeitet werden.
Text	Textausgaben, die nicht in Pivot-Tabellen angezeigt werden
Syntax	Die Befehle, die in einem Dialogfeld ausgewählt werden, können mittels der Schaltfläche „Einfügen“ (engl. „Paste“) als Befehlssyntax direkt ins Syntaxfenster eingefügt werden. Das Befehlssyntax kann bearbeitet und gespeichert werden.

1.2 Die Datenmatrix

Aufbau der Datenmatrix:

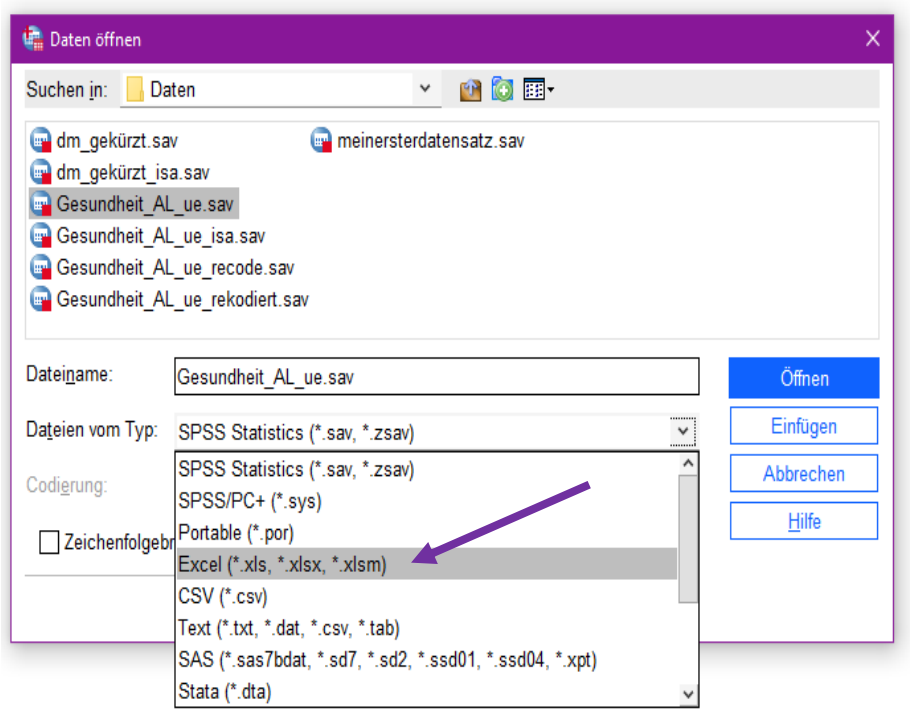


Allgemeine Schreibweise der Datenmatrix:

		Variablen					
		Var 1	Var 2	Var 3	.	.	Var k
Individuelle Beobachtungen (Personen) ↓	Nr. 1	X_{11}	X_{21}	X_{31}	.	.	X_{k1}
	Nr. 2	X_{12}	X_{22}	X_{32}	.	.	X_{k2}
	Nr. 3	X_{13}	X_{23}	X_{33}	.	.	X_{k3}
	Nr. 4	X_{14}	X_{24}	X_{34}	.	.	X_{k4}
	.	.	.	.	.	.	.
	.	.	.	.	.	.	.
	Nr. i	X_{1i}	X_{2i}	X_{3i}	.	.	X_{ki}

Zu Beginn jeder Auswertung erfolgt die Dateneingabe. Nachdem wir die Daten erhoben haben und die Fragebogen – nehmen wir an in Papierform – bei uns gelandet sind, zählen wir sie durch, nummerieren sie und geben damit jedem Fragebogen eine eindeutige ID. Etwaige Namen auf den Fragebögen werde auf diese Weise anonymisiert. Dann beginnen wir mit der Definition des Datenfiles und der Dateneingabe.

Import aus Excel

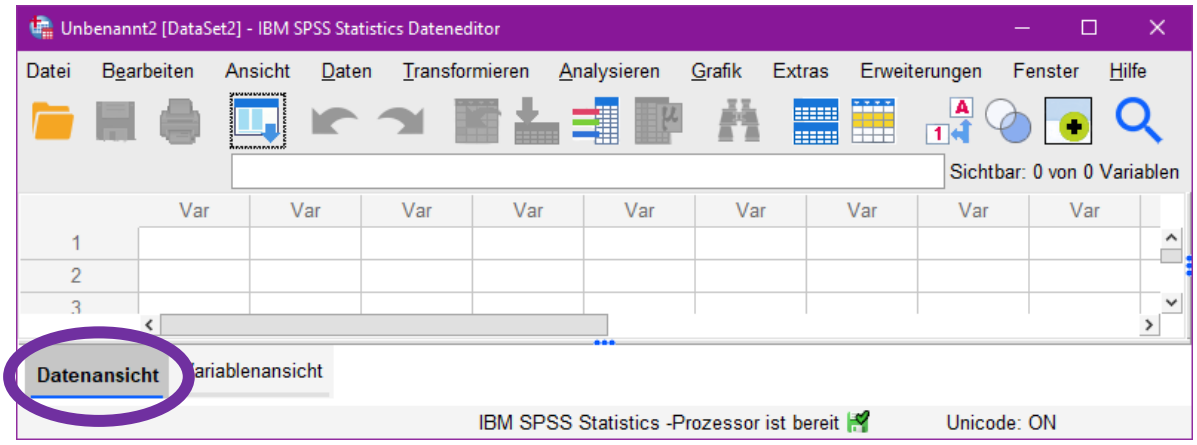


TIPP: Sie können die Daten auch im Excel eingeben und dann ins SPSS transferieren, das geht ganz einfach:

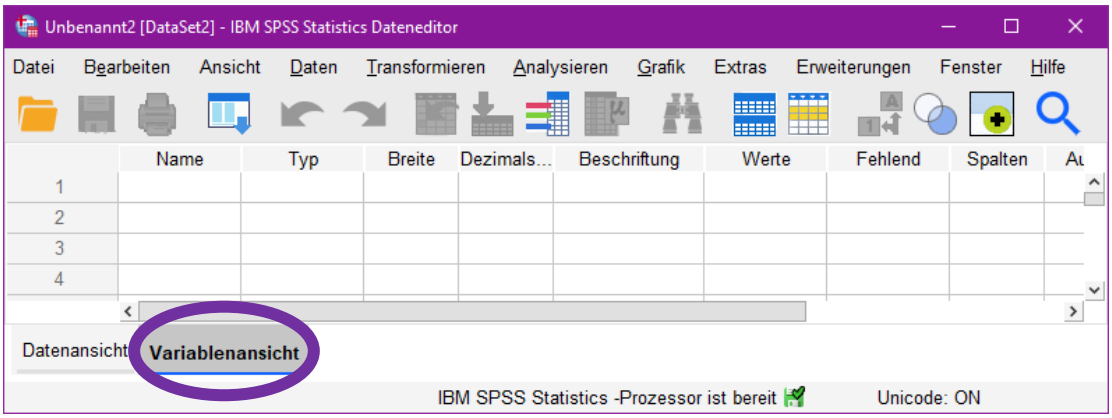
Öffnen eines **Excel-Datenfiles** in SPSS:
Sie haben SPSS gestartet und öffnen jetzt eine Excel-Datei:
Datei öffnen → Daten:
als Dateityp "***Excel***" auswählen und dem folgenden Dialog folgen.
Die in Excel eingegebenen Daten erscheinen dann in der SPSS-Datenmatrix.

1.3 Datenansicht und Variablenansicht

Öffnen von SPSS: Je nach Betriebssystem (Windows 7 oder 8) entweder im Startmenü oder unter "Apps" /"Programme" "IBM SPSS Statistics 28" doppelklicken (Tipp: Erstellen Sie sich eine Verknüpfung zum Desktop!). Es erscheint eine leere Datenmatrix in der **Datenansicht**. Hier werden die Daten eingegeben.



Vorher müssen wir allerdings die Variablen definieren, dazu wechseln wir in die **Variablenansicht**:

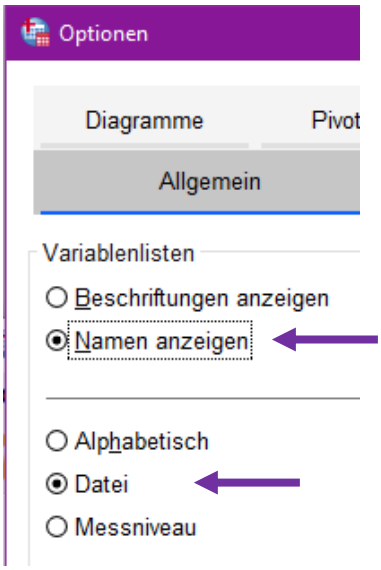


Variablen definieren in der Variablenansicht

Name	**Kurzbezeichnung der Variable – erstes Zeichen ein Buchstabe**
Typ	Numerisch oder String (Text)
Breite (1 bis 40)	Anzahl der Stellen vor dem Komma
Dezimalstellen (0 bis 16)	Anzahl der Dezimalstellen
Beschriftung	Beschreibung der Variable
Werte	Definition der einzelnen Ausprägungen (Werte)
Fehlende Werte	Definieren von fehlenden Werten
Spalten (1 bis 255)	Breite der Variable im Datenfenster
Ausrichtung	Links, Rechts, Mitte
Messniveau	Metrisch, Ordinal, Nominal

1.4 Ändern der Voreinstellungen in SPSS

Vorher ändern wir noch einige **Voreinstellungen**, damit das Arbeiten leichter geht. Achten Sie darauf, dass Sie vor jeder SPSS-Sitzung diese Voreinstellungen kontrollieren.

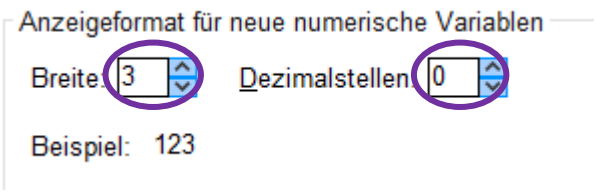


Anzeigen der Variablen in den Variablenlisten

Bearbeiten → Optionen → ***Allgemein*** → „Variablenlisten“
⊙ „Namen anzeigen“ ⊙ „Datei“ auswählen

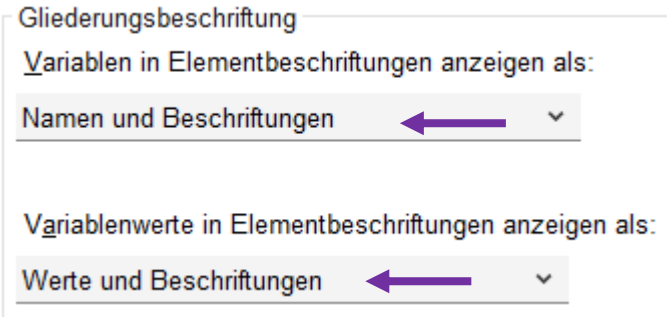
Standardformat für neue Variablen festlegen

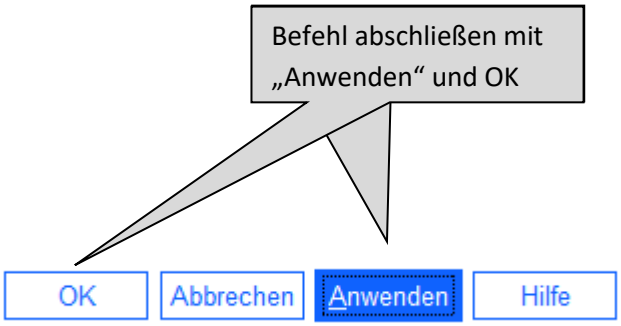
Bearbeiten → Optionen → ***Daten*** →
Breite auf „3“ einstellen
„Dezimalstellen“ auf „0“ (empfohlen)

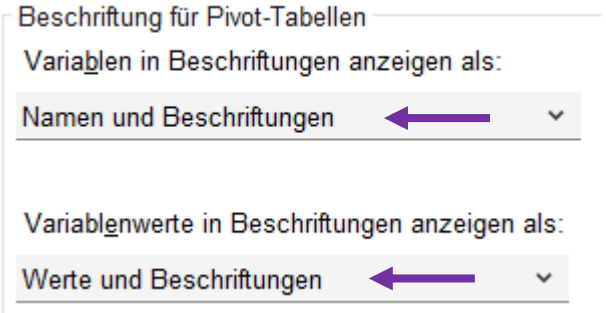


Ändern der Beschriftung der Ausgabe

Bearbeiten → Optionen → ***Ausgabe***
"Gliederungsbeschriftung und
„Beschriftung für Pivot-Tabellen“
→ Variablen in Beschriftungen anzeigen als
„ Namen und Beschriftungen“;
→Variablenwerte in Beschriftungen anzeigen als
„Werte und Beschriftungen“







Achten Sie darauf, dass Sie vor jeder SPSS-Sitzung diese Voreinstellungen kontrollieren.

Hinweis: Ich habe für dieses Skriptum unter „Allgemein“ – „Windows“ – „Erscheinungsbild“ *SPSS-Light* ausgewählt – das ist eine farblose und kontrastreiche Version, besser für die Darstellung besser geeignet.

1.5 Variablen und Werte labeln

Folgenden Auszug aus dem ***Codeplan*** des FB social survey definieren wir im SPSS:

s1 Geschlecht des/der Befragten (von InterviewerIn auszufüllen)

männlich.....1 weiblich.....2 divers.....3 keine Angabe.....99

s2 Alter ..... Jahre *keine Angabe.....0*

s3 Wo wohnen Sie?

in einer Mietwohnung..... 1
in einer Eigentumswohnung..... 2
in einem Eigenheim/ gemietetes Einfamilienhaus..... 3
anderes..... 5 **s3_txt** *anderes, und zwar: _____*
keine Angabe..... 99

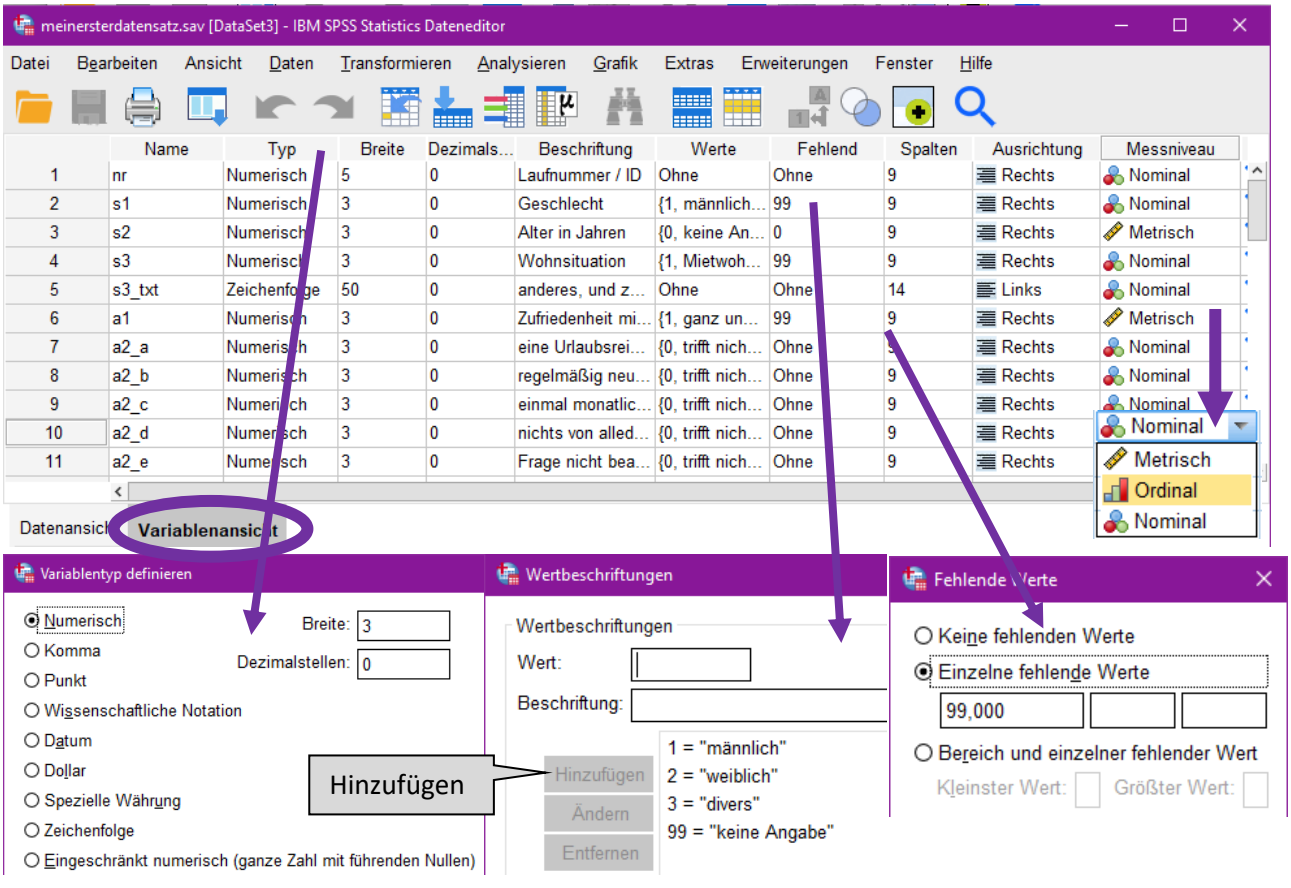
a1 Wie zufrieden sind Sie mit Ihrem Lebensstandard? Ich meine, was Güter und Dienstleistungen betrifft, die man kaufen kann, wie Wohnen, Kleidung, Auto, Urlaub, Reisen. Stufen Sie sich auf folgender Skala ein zwischen „ganz unzufrieden" 1 Punkt) bis ganz zufrieden (10 Punkte) (KARTE vorlegen)

ganz un-zufrieden	keine Angabe 99								ganz zufrieden
1	2	3	4	5	6	7	8	9	10

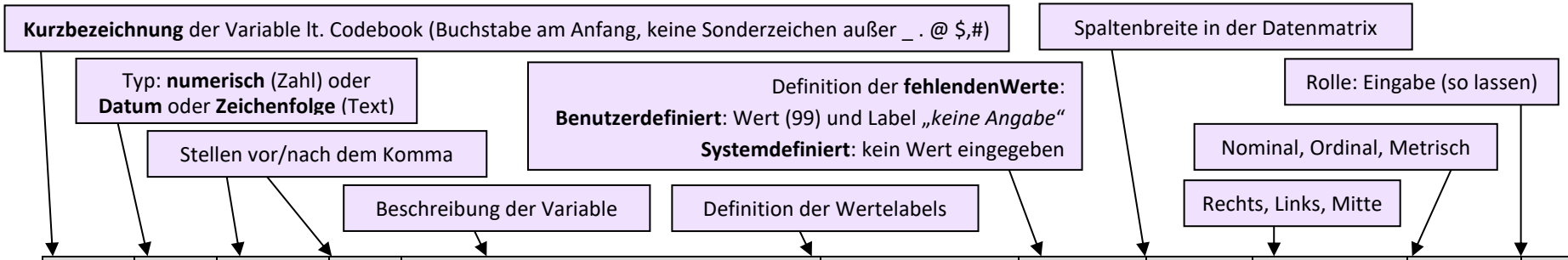
a2 Was können/könnten Sie sich leisten? (DURCHFRAGEN, MEHRFACHNENNUNG)

a eine Urlaubsreise im Jahr ..... 1=trifft zu / 0=trifft nicht zu
b regelmäßig neue Kleidung kaufen ..... 1= trifft zu / 0=trifft nicht zu
c einmal monatlich mit der Familie/Freunden auswärts essen gehen ..... 1= trifft zu / 0=trifft nicht zu
d nichts von alledem ..... 1= trifft zu / 0=trifft nicht zu
e Frage nicht beantwortet..... 1= trifft zu / 0=trifft nicht zu

Jeder Frage im Fragebogen wird ein **Name** (Kurzbezeichnung), ein **Typ** (numerisch, Datum oder Text), ein **Variablenlabel** (Beschreibung der Variable/Fragenwortlaut), **Werte** und **Wertelabels, fehlende Werte** sowie **Messniveau** zugewiesen. Siehe dazu die Übersicht auf der folgenden Seite.



Variablenansicht: Variablen definieren in SPSS



Name	Typ	Breite	Dez.	Beschriftung (=Variablenlabel)	Werte	Fehlende Werte	Spalten	Ausrichtung	Mess-niveau	Rolle
nr	num	3	0	Fragebogennummer/Laufnummer/ID	Keine	Keine	8	Rechts	Nominal	Eingabe
s1	num	3	0	Geschlecht	1= männlich 2= weiblich 3= divers 99= keine Angabe	99	8	Rechts	Nominal	Eingabe
s2	num	3	0	Alter in Jahren	0 = keine Angabe	0	8	Rechts	Metrisch	Eingabe
s3	num	3	0	Wohnsituation	1= Mietwohnung 2= Eigentumswohnung 3= Eigenheim/gemietetes Einfamilienhaus 4= anderes 99= keine Angabe	99	8	Rechts	Nominal	Eingabe
s3_txt	Zeich	30	0	anderes, und zwar…	Keine	Keine	8	Links	Nominal	Eingabe
a1	num	3	0	Zufriedenheit mit Lebensstandard	1 = ganz unzufrieden 10= ganz zufrieden	99	8	Rechts	Metrisch	Eingabe
a2_a	num	3	0	eine Urlaubsreise im Jahr	0= trifft nicht zu 1= trifft zu	Keine	8	Rechts	Nominal	Eingabe
a2_b	num	3	0	regelmäßig neue Kleidung kaufen	0= trifft nicht zu 1= trifft zu	Keine	8	Rechts	Nominal	Eingabe
a2_c	num	3	0	einmal monatlich auswärts mit Fam./Freunden essen gehen	0= trifft nicht zu 1= trifft zu	Keine	8	Rechts	Nominal	Eingabe
a2_d	num	3	0	nichts von alledem	0= trifft nicht zu 1= trifft zu	Keine	8	Rechts	Nominal	Eingabe
a2_e	num	3	0	Frage nicht beantwortet	0= trifft nicht zu 1= trifft zu	Keine	8	Rechts	Nominal	Eingabe

Das **Codebook/Codeplan** enthält die Variablennamen, die Codierung der Antworten und die Codierung der fehlende Werte. Jeder Frage im Fragebogen wird ein **Name** (Kurzbezeichnung), ein **Typ** (numerisch, Datum oder Text), ein **Variablenlabel** (Beschreibung der Variable oder Fragenwortlaut), **Werte** und **Wertelabels, fehlende Werte** sowie **Messniveau** zugewiesen.

Variablenname	Möglichst kurze und eindeutige Bezeichnung der Variable; In der Regel wird die Fragennummerierung aus dem Fragebogen übernommen: z.B. f1 (erste Frage: Geschlecht), f2 (zweite Frage: Alter), f3 (dritte Frage: Bildung) etc.; Immer mit Buchstaben beginnen; die übrigen Zeichen können Buchstaben, Ziffern, Punkt oder Symbolen (@, #, _, $); das letzte Zeichen darf kein Punkt sein;
Variablentyp	**Datentypen**: Numerisch, Komma, Punkt, wissenschaftliche Notation, Datum, Währung, String (Textvariable); Standard: Numerisch; **Maß**: Skala (=metrisch), ordinal oder nominal;
Labels	**Variablenlabels**: den kurzen Variablennamen werden beschreibende Variablenlabels zugeordnet **Wertelabels**: jedem Wert einer Variable soll ein beschreibendes Wertelabel zugeordnet werden; die unterschiedlichen Ausprägungen werden so definiert; z.B.: Bei Variable f3 (Lebenssituation) Wert 4 bedeutet: „andere Lebensform“
Fehlende Werte	Angegebene Datenwerte können als **benutzerdefinierte** fehlende Werte definiert werden; z.B.: als Wertelabel für die Ausprägung 99 „Antwort verweigert“ oder „Frage trifft auf die befragte Person nicht zu“ ; Werden diese Ausprägungen als fehlende Werte definiert, werden sie nicht bei statistischen Berechnungen miteinbezogen; (siehe s1 „99“ = keine Angabe) **Systemdefinierte** fehlende Werte sind jene, wo bereits bei der Dateneingabe keine Werte eingegeben werden;
Formate	Breite der **Spalten** und **Ausrichtung** der Datenwerte

Dateneingabe in der Datenansicht

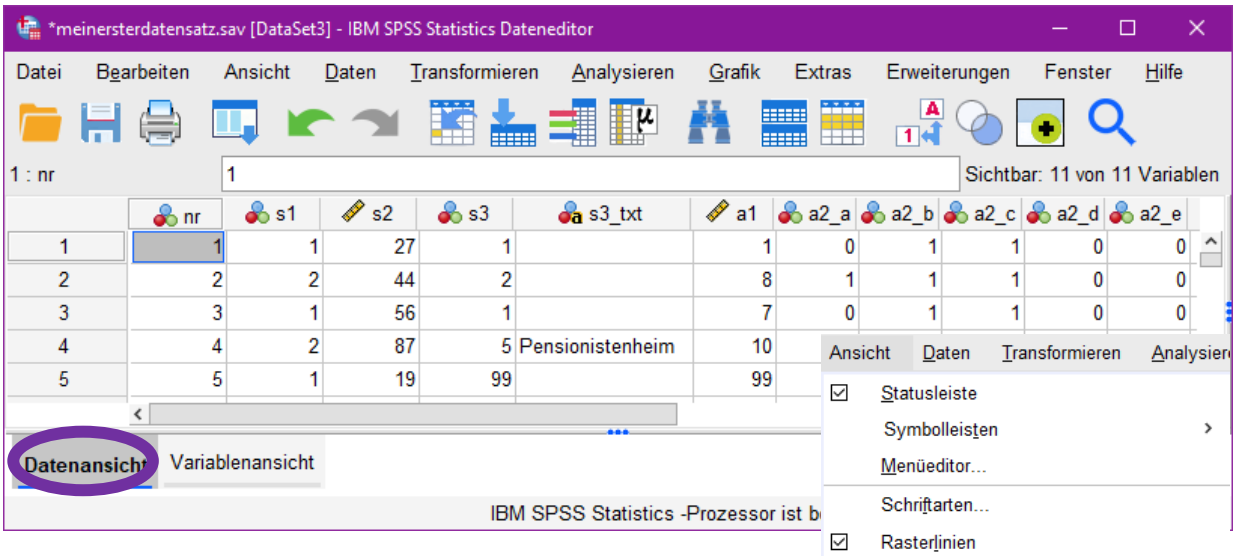
- Aktive Zelle wird durch fetten Rand hervorgehoben.
- In der linken oberen Ecke werden der Variablenname und die Zeilennummer der aktiven Zelle angezeigt.
- Mit Tabulatortaste können Variablenwerte zeilenweise eingegeben werden.
- Nach der Eingabe des letzten Wertes einer Zeile kann mit der Pos1-Taste zum Beginn des Datenfiles und mit Cursor-Pfeil in die zweite Zeile gewechselt werden.

Zeilen bzw. Fall löschen: Fall durch Anklicken der SPSS-Fallnummer markieren
⇨ Entf-Taste oder Bearbeiten ⇨Löschen

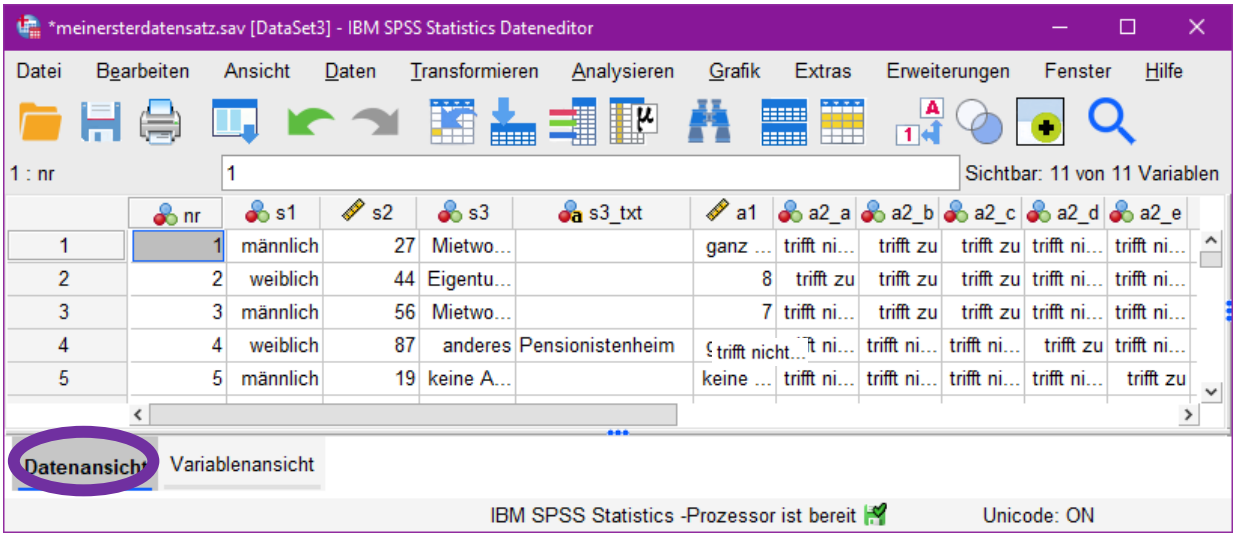
Spalte bzw. Variable löschen: Variable durch Anklicken des Variablennamens markieren
⇨ Entf-Taste oder Bearbeiten ⇨Löschen

Vor der Dateneingabe werden alle Fragebögen durchnummeriert. **Die erste Variable im Datensatz ist IMMER die ID-Variable**, die jedem eigegebenen Fall eindeutig zuordenbar ist (Laufnummer oder Code). Warum? Der Grundsatz der Nachvollziehbarkeit muss bei wissenschaftlichen Arbeiten gewähr-leistet sein. Und: Wenn Fehler bei der Dateneingabe passieren, können diese zurückverfolgt und ausgebessert werden. Beachte: **Die ID-Variable hat nichts mit der Aufhebung der Anonymität zu tun!**
Personenbezogene Daten werden zu Auswertungszwecken grundsätzlich anonym (ohne Namen oder Sozialversicherungsnummern) gespeichert.

Nun geben wir 5 fiktive Fälle in der Datenansicht ein.



Im Menü "**Ansicht**" können wir zwischen der Ansicht der Codes (siehe oben) und der Ansicht mit den Wertebeschriftungen (siehe unten) switchen: (Wertebeschriftungen an- oder ausklicken)



- Jede **Zeile** stellt einen Fall oder eine Beobachtung dar (= jede Person, die einen FB ausgefüllt hat).
- Jede **Spalte** stellt eine Variable oder eine gemessene Eigenschaft dar (z.B. das Alter der Personen).
- Jede **Zelle** enthält einen einzelnen Wert einer Variable für einen Fall (z.B. Alter der Befragten Nr.4)
- Spalten (Variablen) und Zeilen/Fälle können markiert und anschließend an eine beliebige Stelle im Datenfile verschoben werden (anklicken, halten und verschieben).

1.6 Richtlinien bei der Dateneingabe

→ Jeder FB erhält eine **Laufnummer**, die auch im Datensatz enthalten ist: So garantiert man die Nachvollziehbarkeit der Ergebnisse und ermöglicht die Kontrolle bei Eingabefehlern.

→ **Codeplan** erstellen: Jede Variable erhält einen (kurzen) **Variablennamen** und ein (beschreibendes) **Variablenlabel**, jeder Antwortmöglichkeit wird eine Zahl zugeordnet (**Werte und Wertelabels**).
Tipp: Am besten bereits bei der FB-Erstellung die Fragen nummerieren und die Codes darstellen.

→ Während der Eingabe: zwischendurch immer wieder speichern!

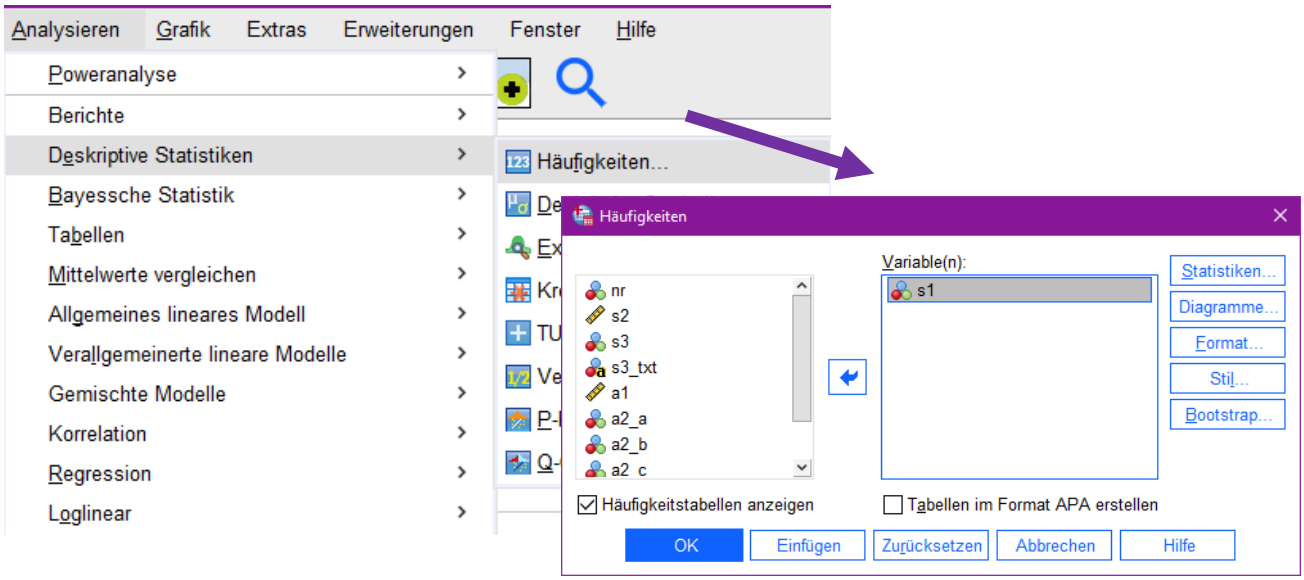
→ Regeln zum **Umgang mit fehlenden Eintragungen** aufstellen:
Entweder leer lassen oder festgelegte Codes eintragen (z.B. 99).

1.7 Datenkontrolle

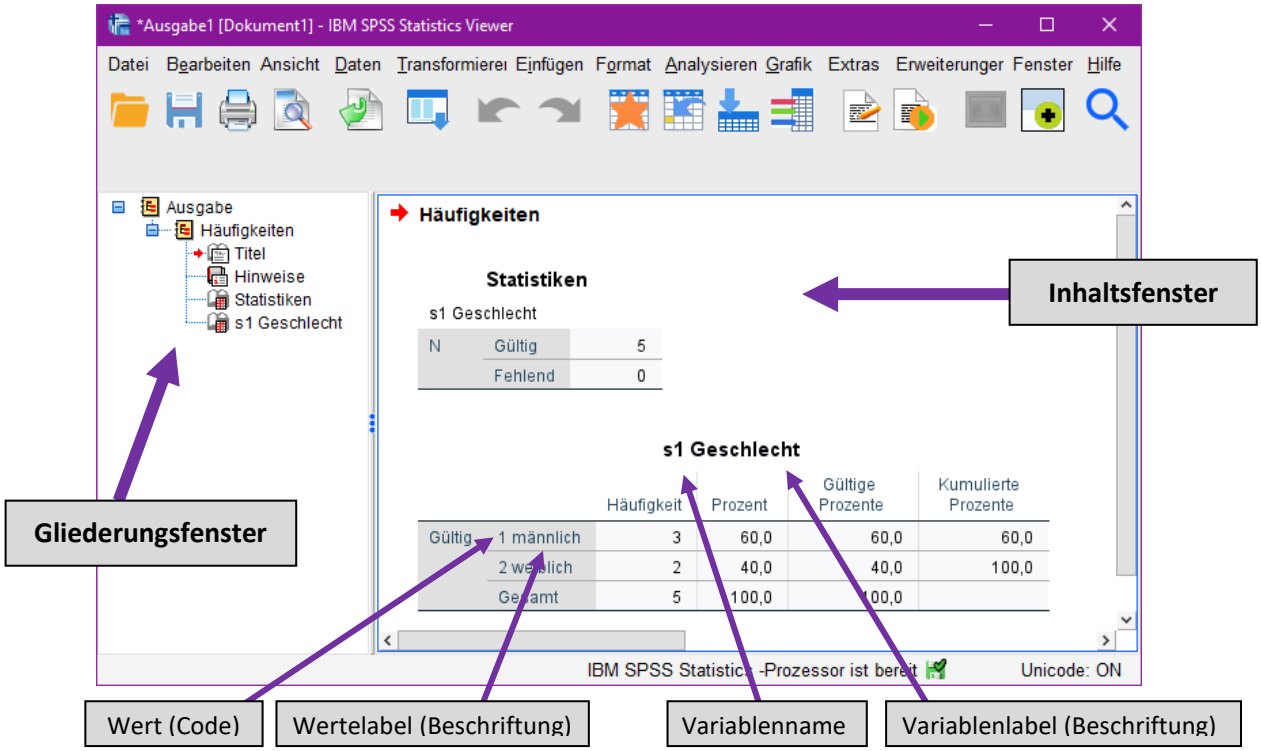
Der erste Schritt nach der Dateneingabe ist die Datenkontrolle. Wir erstellen für alle Variablen eine Häufigkeitstabelle und kontrollieren, ob wir keine Fehler bei der Dateneingabe gemacht haben.

Erstellen einer Häufigkeitstabelle:

Analysieren → Deskriptive Statistiken → Häufigkeiten .... → s1 (Geschlecht) → OK



Sobald wir eine Tabelle erstellt haben, öffnet sich das **Ausgabefenster**. Dieses teilt sich in:



→ Das **Gliederungsfenster** enthält eine Gliederungsansicht des Inhalts. Hier kann man mit einem Klick zwischen den Tabellen wechseln, einzelne Tabellen ein- oder ausblenden (klicken auf das Minus- bzw. Plus-Symbol), verschieben oder löschen.

→ Das **Inhaltsfenster** enthält Tabellen mit Statistiken, Diagramme und Textausgaben. Mit Doppelklicken auf die Objekte können Sie diese bearbeiten.

→ **Kopieren** mit Strg+c, **einfügen** ins Excel oder Word als bearbeitbare Tabellen mit Strg+v, einfügen mit "Inhalte einfügen" als "geräteunabhängige Bitmap (Objekt) in einen Bericht in Word.

Datenkontrolle: Anhand der Häufigkeitstabellen fehlerhafte Einträge aufgefunden werden.

Eingabe eines falschen Codes

s1 Geschlecht

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 männlich	2	40,0	40,0	40,0
	2 weiblich	2	40,0	40,0	80,0
	5	1	20,0	20,0	100,0
	Gesamt	5	100,0	100,0	

Auffinden des falschen Eintrags in der Datenmatrix.
Der Fehler ist beim Fragebogen mit der Nr.10 passiert.

	nr	s1	s2	s3
1	1	1	27	1
2	2	2	44	2
3	3	1	56	1
4	4	2	87	5
5	5	5	19	99

Wir sehen nochmal nach im FB Nr. 5 und stellen fest, dass der Befragte männlich ist.
Wir bessern den Fehler in der Datenmatrix aus und kontrollieren nochmals die Tabelle.

s1 Geschlecht

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 männlich	3	60,0	60,0	60,0
	2 weiblich	2	40,0	40,0	100,0
	Gesamt	5	100,0	100,0	

In jedem Datensatz steht an erster Stelle eine **ID-Variable**, die die Fälle der Datenquelle zuordnet, nicht aber den konkreten Personen. Die ID ist wichtig, um eventuelle **Fehler richtigzustellen** bzw. den Grundsatz der **Nachvollziehbarkeit** in der wissenschaftlichen Forschung zu gewährleisten.

Anschließend speichern Sie den Datensatz unter „*meinersterdatensatz.sav*“.

Datei → Speichern unter → Ordner "UE_statistik" erstellen → *"meinerstersdatensatz.sav"*
→ Speichern

1.8 Hausübung 1: Datenfile erstellen und Date eingeben

Auf der letzten Seite dieses Skriptums finden Sie die formalen Kriterien für das Verfassen der Hausarbeit für diese Lehrveranstaltung.
Kopieren Sie alle relevanten Outputs aus dem SPSS ins word (Kopieren = Strg+C, Einfügen = Strg+V), und schreiben Sie dort Ihre Interpretationen.

Pflichtaufgabe 1:

Erstellen eines Datenfiles in SPSS:
Arbeiten Sie mit dem **Fragebogen** zum Darmmanagement.
Wählen Sie jeweils drei Variablen mit nominalem – ordinalem – metrischen Datenniveau aus, und definieren Sie dazu ein Datenfile in SPSS. Labeln Sie die Variablen und die Codes.

Dateneingabe und Datenkontrolle:
Geben Sie **10 fiktive Fälle** ein. Führen Sie eine Datenkontrolle mittels Häufigkeitstabellen für alle Variablen aus ihrem Datensatz durch.

Dokumentation:
In der Hausübung (in word) präsentieren Sie die Häufigkeitstabellen aller Variablen.

Fleißaufgabe für Lernwütige: Geben Sie insgesamt 50 Fälle ein!

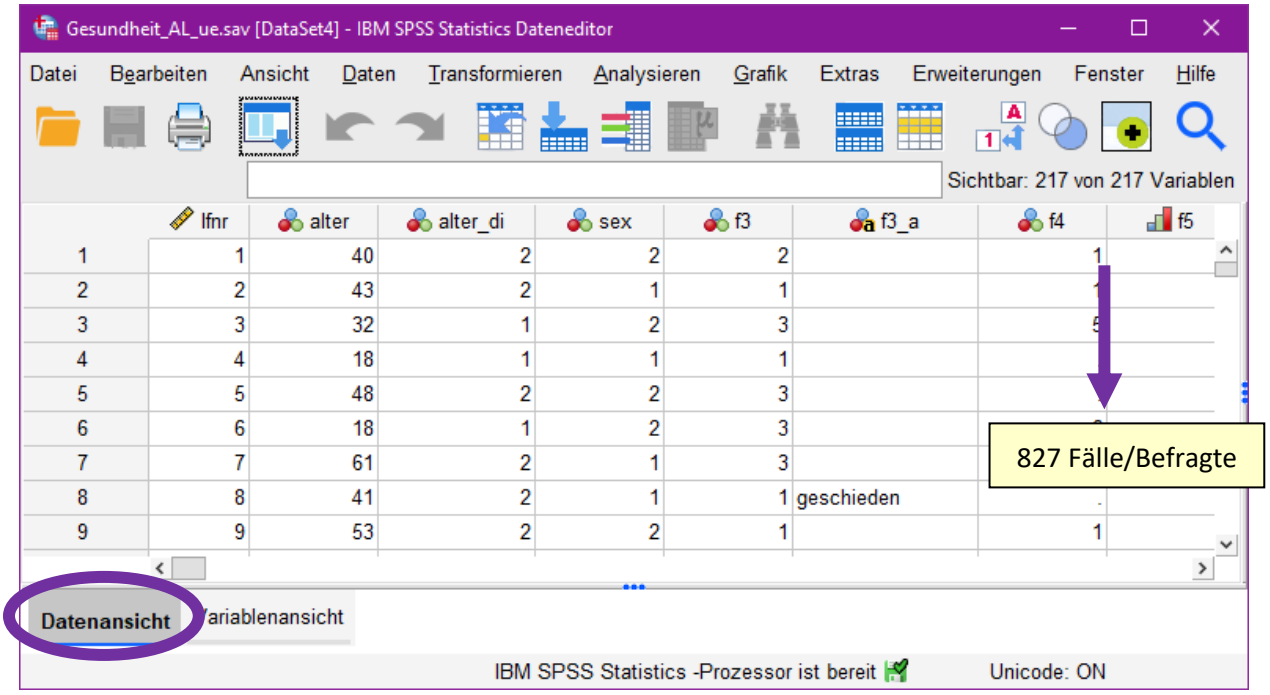
2 Erstellen und Interpretieren einer Häufigkeitstabelle

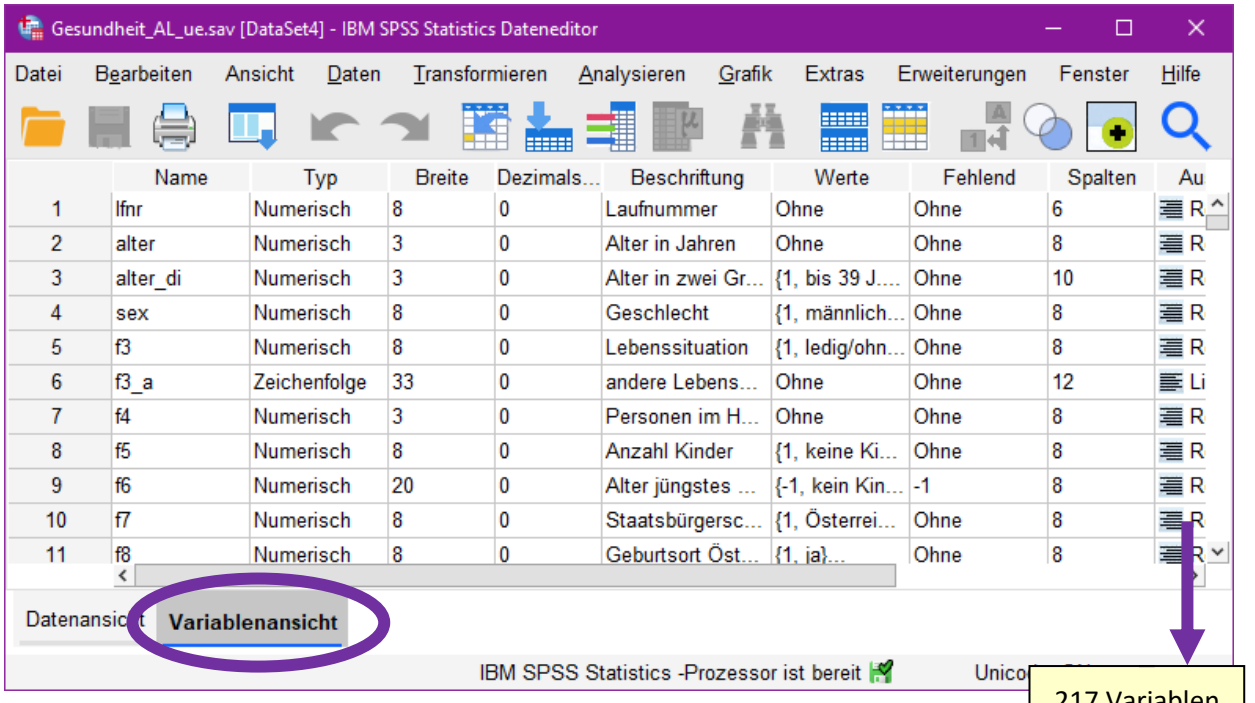
Wir arbeiten in diesem Skriptum mit dem Datensatz zur Gesundheit von Arbeitslosen.

Daten öffnen

Datei → Öffnen → Daten… (.sav *Datensatz auswählen*) [Gesundheit_AL_ue.sav]

Daten- und Variablenansicht





Häufigkeitstabelle erstellen

Analysieren → deskriptive Statistiken → Häufigkeiten → *Variablen auswählen….* z.B. **f5**
→ OK oder „Einfügen“ um den Befehl zunächst in eine Syntax-Datei (.sps) zu schreiben

f5 Anzahl Kinder

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 keine Kinder	334	40,4	42,0	42,0
	2 ein Kind	159	19,2	20,0	62,0
	3 zwei Kinder	167	20,2	21,0	83,0
	4 drei Kinder	76	9,2	9,6	92,6
	5 mehr als drei Kinder	46	5,6	5,8	98,4
	88 Kind(er), Anzahl nicht angegeben	13	1,6	1,6	100,0
	Gesamt	795	96,1	100,0	
Fehlend	System	32	3,9		
Gesamt		827	100,0		

Fehlende Werte definieren:

Variablenansicht → in Zeile der betreffenden Variable und in der Spalte „Fehlend“ →
→ das Feld anklicken → hier den jeweiligen Wert eintragen [für Variable f5 ist es der Code „88“].

f5 Anzahl Kinder

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 keine Kinder	334	40,4	42,7	42,7
	2 ein Kind	159	19,2	20,3	63,0
	3 zwei Kinder	167	20,2	21,4	84,4
	4 drei Kinder	76	9,2	9,7	94,1
	5 mehr als drei Kinder	46	5,6	5,9	100,0
	Gesamt	782	94,6	100,0	
Fehlend	88 Kind(er), Anzahl nicht angegeben	13	1,6		
	System	32	3,9		
	Gesamt	45	5,4		
Gesamt		827	100,0		

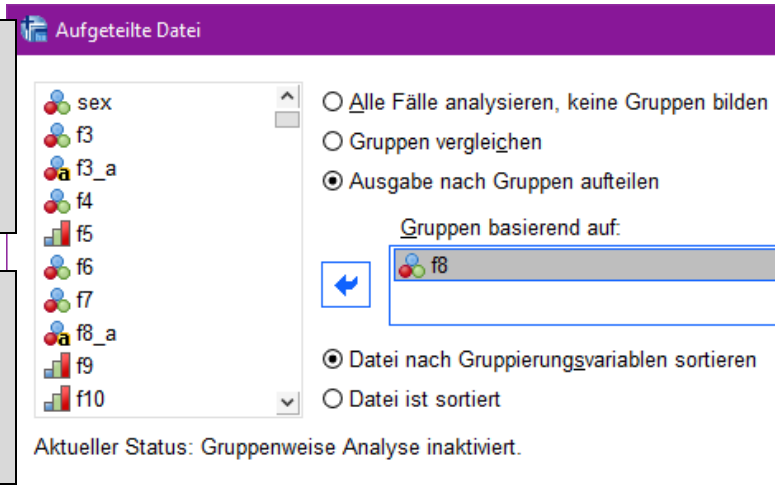
Interpretation:
42% der Befragten haben keine Kinder, jeweils rund ein Fünftel haben ein bzw. zwei Kinder.
Der Anteil der Befragten mit mehr als zwei Kindern beträgt insgesamt etwa 15%.

2.1 Datei aufteilen: Vergleich von Gruppen

Mit einer einfachen Funktion im Menü "*Daten*" kann diese Häufigkeitstabelle getrennt für zwei Vergleichsgruppen dargestellt werden. Wir wollen untersuchen, ob sich die Anzahl der Kinder nach Migrationshintergrund (f8 Geburtsort) unterscheidet.

Daten → Datei aufteilen…
⊙ **Ausgabe nach Gruppen aufteilen → f8**
⊙ Datei nach Gruppenvariablen sortieren
→ Analysieren→ deskriptive Statistiken
→ Häufigkeiten → f5 → OK

Hinweis: Es gibt zwei Optionen:
⊙ Gruppen vergleichen → **eine Tabelle** für alle Vergleichsgruppen (ev. unübersichtlich)
⊙ Ausgabe nach Gruppen aufteilen → für jede Vergleichsgruppen eine **separate Tabelle**



f5 Anzahl Kinder[a]

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 keine Kinder	247	50,9	53,0	53,0
	2 ein Kind	102	21,0	21,9	74,9
	3 zwei Kinder	74	15,3	15,9	90,8
	4 drei Kinder	28	5,8	6,0	96,8
	5 mehr als drei Kinder	15	3,1	3,2	100,0
	Gesamt	466	96,1	100,0	
Fehlend	88 Kind(er), Anzahl nicht angegeben	8	1,6		
	System	11	2,3		
	Gesamt	19	3,9		
Gesamt		485	100,0		

a. f8 Geburtsort Österreich = 1 ja

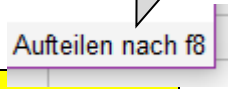
BEACHTE:
Mit dem Befehl "Datei aufteilen" wird nur die Teilung aktiviert!
Die Befehle für die Tabellen müssen neu ausgeführt werden!

f5 Anzahl Kinder[a]

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 keine Kinder	85	26,1	27,2	27,2
	2 ein Kind	56	17,2	17,9	45,2
	3 zwei Kinder	93	28,5	29,8	75,0
	4 drei Kinder	47	14,4	15,1	90,1
	5 mehr als drei Kinder	31	9,5	9,9	100,0
	Gesamt	312	95,7	100,0	
Fehlend	88 Kind(er), Anzahl nicht angegeben	4	1,2		
	System	10	3,1		
	Gesamt	14	4,3		
Gesamt		326	100,0		

a. f8 Geburtsort Österreich = 2 nein

BEACHTE:
In der rechten unteren Leiste in der Datenansicht steht ein Hinweis, wenn eine Aufteilung aktiv ist.

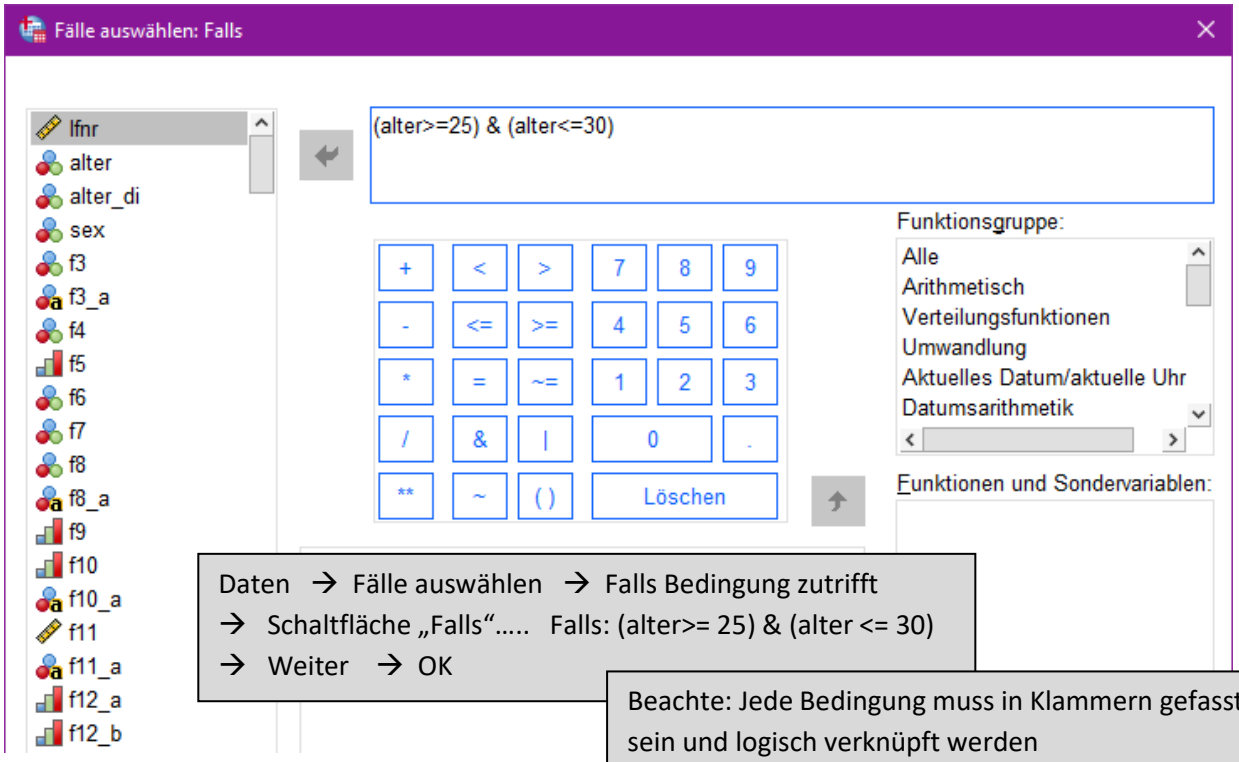


Wichtig: nach aufgeteilter Datei die **Aufteilung wieder ausschalten**!
Daten → Aufgeteilte Datei → Alle Fälle analysieren, keine Gruppen bilden → OK

Interpretation:
Mehr als die Hälfte der Befragten ohne Migrationshintergrund haben keine Kinder (53%), etwa ein Fünftel hat ein Kind und das restliche Viertel der Befragten hat zwei und mehr Kinder.
Von den Befragten mit Migrationshintergrund hat lediglich ein gutes Viertel der Befragten (27%) keine Kinder, die meisten (29%) haben zwei Kinder und das restliche Viertel hat drei und mehr Kinder.

2.2 Fälle auswählen: Auswahl einer Subgruppe

Nun wählen wir eine bestimmte Subgruppe aus, um für diese spezielle Subgruppe Auswertungen durchzuführen. Wir wählen Befragte aus, die zwischen 25 und 30 Jahre alt sind.



Daten → Fälle auswählen → Falls Bedingung zutrifft
→ Schaltfläche „Falls“….. Falls: (alter>= 25) & (alter <= 30)
→ Weiter → OK

Beachte: Jede Bedingung muss in Klammern gefasst sein und logisch verknüpft werden
→ logisches "UND" wird mit **&** gekennzeichnet
→ logisches "ODER" wird mit | gekennzeichnet

f5 Anzahl Kinder

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 keine Kinder	46	64,8	65,7	65,7
	2 ein Kind	11	15,5	15,7	81,4
	3 zwei Kinder	9	12,7	12,9	94,3
	4 drei Kinder	3	4,2	4,3	98,6
	5 mehr als drei Kinder	1	1,4	1,4	100,0
	Gesamt	70	98,6	100,0	
Fehlend	System	1	1,4		
Gesamt		71	100,0		

Interpretation:
71 Befragte in der Stichprobe sind zwischen 25 und 30 Jahre alt. Fast zwei Drittel dieser Altersgruppe haben (noch) keine Kinder (66%). Jeweils etwas mehr als 10% haben bereits ein oder zwei Kinder.

Wichtig: nach Fallauswahl wieder ausschalten!
Daten → Fälle auswählen → Alle Fälle → Ok

Bei der Auswahl von Fällen erstellt SPSS eine neue Variable (filter_$) mit den Codes
1 = Selected und 0 = Not Selected.
Alle Fälle, auf die die Filterbedingung zutrifft, werden "Selected", alle anderen nicht.

Die Personen mit den Laufnummern 11 und 21 und 24 sind zwischen 25 und 30 Jahre alt und daher in unserem Filter drinnen.

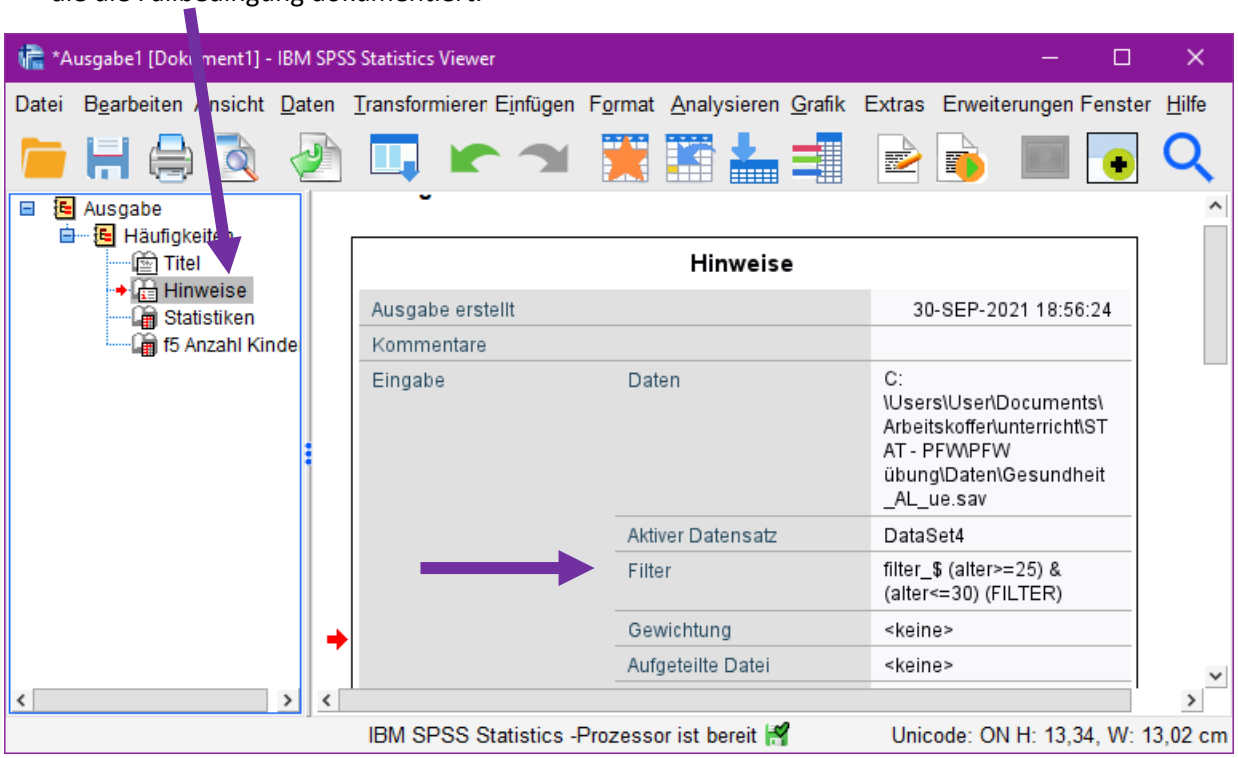
Die Nicht-Ausgewählten werden währenddessen der Filter aktiv ist, als durchgestrichen gekennzeichnet.

	lfnr	alter	filter_$
1	1	40	Not Selected
2	2	43	Not Selected
3	3	32	Not Selected
4	4	18	Not Selected
5	5	48	Not Selected
6	6	18	Not Selected
7	7	61	Not Selected
8	8	41	Not Selected
9	9	53	Not Selected
10	10	17	Not Selected
11	11	27	Selected
12	12	46	Not Selected
13	13	38	Not Selected
14	14	38	Not Selected
15	15	34	Not Selected
16	16	41	Not Selected
17	17	49	Not Selected
18	18	21	Not Selected
19	19	49	Not Selected
20	20	55	Not Selected
21	21	29	Selected
22	22	21	Not Selected
23	23	45	Not Selected
24	24	30	Selected
25	25	54	Not Selected

Unicode: ON Filter aktiv

Achtung SPSS28 – in meiner Version ist ein Bug: Die gefilterten Fälle werden ausgeblendet. Wird die Spalte jedoch sortiert, dann werden die gefilterten Fälle wieder angezeigt, sind nicht durchgestrichen – wie hier – die Daten sind aber trotzdem gefiltert (Filtervariable hinten ist vorhanden mit 0/1).

Ein Hinweis in der rechten unteren Leiste zeigt an, dass ein Filter aktiv ist.
Beachte: An der Häufigkeitstabelle ist nicht direkt ersichtlich, um welchen Filter es sich handelt.
Klicken Sie dazu im Gliederungsfenster auf "**Hinweise**". Dann erscheint eine Tabelle, die die Fallbedingung dokumentiert.



2.3 Hausübung 2: Häufigkeitstabelle, Datei aufteilen, Fälle auswählen

Pflichtaufgabe 2:
Wählen Sie aus der Fragebatterie f18 zu den körperlichen Beschwerden eine Beschwerde aus, die Sie am meisten interessiert. Erstellen Sie die Häufigkeitstabelle und beschreiben Sie die Ergebnisse.

Welche Unterschiede könnten sich anhand eines sozialen Merkmals zeigen?
Wählen Sie eine geeignete Variable für einen Gruppenvergleich aus. Teilen Sie die Datei nach dieser Variable auf und untersuchen Sie nochmals dieselbe Häufigkeitstabelle.
Beschreiben Sie die charakteristischen Unterschiede. Welches Ergebnis haben Sie erwartet? Welches Ergebnis hat Sie überrascht?

Wählen Sie nun 20 bis 40-jährige Befragte aus und erstellen Sie dieselbe Häufigkeitstabelle. Nun wählen Sie 45 bis 55-jährige Befragte aus und erstellen wieder dieselbe Häufigkeitstabelle. Beschreiben Sie die Unterschiede der beiden Altersgruppen hinsichtlich der Häufigkeit der von Ihnen gewählten Beschwerden.

Fleißaufgabe für Lernwütige: Führen Sie die obige Aufgabe für eine andere Beschwerde aus dem Frageblock f18 durch. Beschreiben Sie die Ergebnisse!

Fleißaufgabe für Lernwütige: Gruppenvergleiche: **Erstellen Sie dazu einen Boxplot**:
Analysieren → Deskriptive Statistiken → Explorative Datenanalyse
→ Abhängige Variable: Variable auswählen aus Frageblock f18
Faktorenliste: beispielsweise Variable sex
Anzeigen: Nur Diagramme: Boxplot: Faktorstufen zusammen, Stengel-Blatt ausklicken
Probieren Sie noch eine weitere Variable in der Faktorenliste aus!

Fleißaufgabe für Lernwütige: Üben Sie die Funktionen "Datei aufteilen" und "Fälle auswählen" mit Ihrem eigenen fiktiven Datensatz!

3 Deskriptive Statistik - Statistische Kennzahlen

Hier eine Übersicht über die **Interpretation** der wichtigsten Lage-, Streuungs- und Formmaße:

Lagemaße	**Modus** (mode)	Häufigster Wert	ab Nominal
	Median(Zentralwert, median)	50 % der Fälle weisen Werte unter Median auf	ab Ordinal
	1. Quartil	25% der Fälle weisen einen Wert bin zum 1. Quartil auf.	ab Ordinal
	3. Quartil	75% der Fälle weisen einen Wert bin zum 3. Quartil auf.	ab Ordinal
	Perzentile (percentile)	eine oder mehrere Marken, unterhalb deren ein bestimmter Anteil von Fällen liegt	ab Ordinal
	Arithmetisches Mittel (mean)	Mittelwert, Durchschnitt	ab Metrisch
Streuungsmaße	**IQA** (Interquartilabstand)	Intervall zwischen 1. und 3. Quartil	ab Ordinal
	Spannweite (range)	Differenz zwischen größtem und kleinstem Wert	ab Metrisch
	Standardabweichung (standarddeviation)	Maß für die Stärke der Abweichung der Werte um den Mittelwert (Wurzel aus der Varianz)	ab Metrisch
	Varianz (variance)	Quadrat der Standardabweichung	ab Metrisch
	Standardfehler (standarderror)	Maß für die Variabilität (Zuverlässigkeit) des Stichprobenmittelwerts	ab Metrisch
Formmaße	**Schiefe** (Skewness)	Symmetrie einer Verteilung: 0 = symmetrisch rechtsschiefe Verteilung = positiver Wert linksschiefe Verteilung = negativer Wert	ab Metrisch
	Wölbung (Kurtosis)	Steilheit der Verteilung: 0 = wie Normalverteilung Steiler als Normalverteilung = positiver Wert Flacher als Normalverteilung = negativer Wert	ab Metrisch

Hier eine Übersicht über die wichtigsten Lage-, Streuungs- und Formmaße nach **Datenniveau**:

<table>
<tr><th></th><th>nominal</th><th>ordinal</th><th>metrisch</th></tr>
<tr><td rowspan="3">Zuordnung</td><td colspan="3">Alle Fälle sind explizit EINER/M (und nur einer/m) Kategorie/Wert zuzuordnen</td></tr>
<tr><td colspan="2">Jede Kategorie ist bezeichnet</td><td>nur Endpunkte der Skala sind bezeichnet</td></tr>
<tr><td></td><td>Kategorien sind gereiht</td><td>Punkteskala oder Messeinheit</td></tr>
<tr><td rowspan="3">Merkmal</td><td>gleich – ungleich</td><td>gleich – ungleich</td><td>gleich – ungleich</td></tr>
<tr><td rowspan="2"></td><td>mehr – weniger</td><td>mehr – weniger</td></tr>
<tr><td></td><td>gleiche Abstände</td></tr>
<tr><td rowspan="3">Lagemaß</td><td>Modus</td><td>Modus</td><td>Modus</td></tr>
<tr><td rowspan="2"></td><td>Median, Perzentile</td><td>Median, Perzentile</td></tr>
<tr><td></td><td>Mittelwert</td></tr>
<tr><td rowspan="2">Streuungs-maß</td><td rowspan="2"></td><td>IQA (Interquartilabstand)
Quartilabstände</td><td>IQA (Interquartilabstand)
Quartilabstände</td></tr>
<tr><td></td><td>Standardabweichung (Varianz)</td></tr>
<tr><td>Formmaße</td><td></td><td></td><td>Schiefe (Skewness) + Wölbung (Kurtosis)</td></tr>
</table>

3.1 Kennzahlen bei nominalem Datenniveau

Analysieren → deskriptive Statistiken → Häufigkeiten → **f3**
→ **Diagramme** → ⊙ Balkendiagramm mit ⊙ Prozentwerten
→ **Statistik**: Kennzahlen auswählen ☑ *Modalwert* → OK

Statistiken

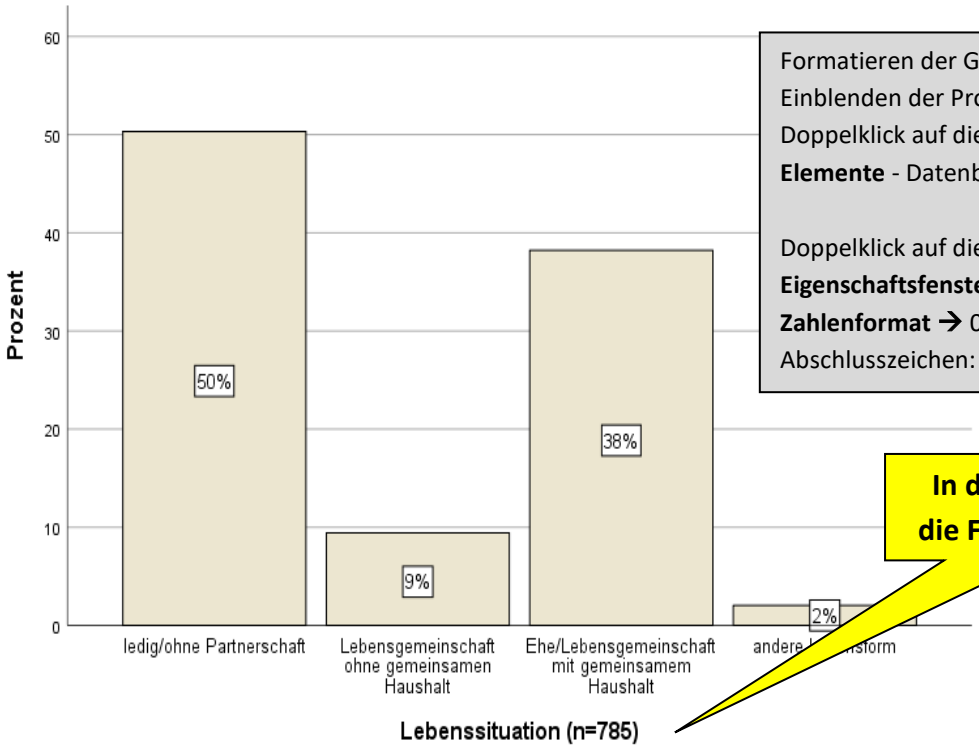
f3 Lebenssituation

N	Gültig	785
	Fehlend	42
Modus		1

Interpretation:
Die meisten Befragten in dieser Stichprobe sind ledig bzw. ohne Partnerschaft (50%).
Etwas mehr als ein Drittel der Befragten (38%) lebt mit einer Partnerschaft im gemeinsamen Haushalt.

f3 Lebenssituation

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 ledig/ohne Partnerschaft	395	47,8	50,3	50,3
	2 Lebensgemeinschaft ohne gemeinsamen Haushalt	74	8,9	9,4	59,7
	3 Ehe/Lebensgemeinschaft mit gemeinsamem Haushalt	300	36,3	38,2	98,0
	4 andere Lebensform	16	1,9	2,0	100,0
	Gesamt	785	94,9	100,0	
Fehlend	System	42	5,1		
Gesamt		827	100,0		



Formatieren der Grafik:
Einblenden der Prozentwerte in die Balken:
Doppelklick auf die Grafik →
Elemente - Datenbeschriftung einblenden

Doppelklick auf die Datenbeschriftung -
Eigenschaftsfenster erscheint:
Zahlenformat → 0 Dezimalstellen,
Abschlusszeichen: % → Zuweisen

In der Grafik immer die Fallzahl anführen!

3.2 Kennzahlen bei ordinalem Datenniveau

Analysieren → deskriptive Statistiken → Häufigkeiten → **f12_b**
→ **Diagramme** → ⊙ Balkendiagramm mit ⊙ Prozentwerten
→ **Statistik**: Kennzahlen auswählen ☑ *Quartile, Median, Modalwert* → OK

Statistiken

f12_b seelischen Belastungen am letzten Arbeitsplatz

N	Gültig	764
	Fehlend	63
Median		3,00
Modus		3
Perzentile	25	3,00
	50	3,00
	75	4,00

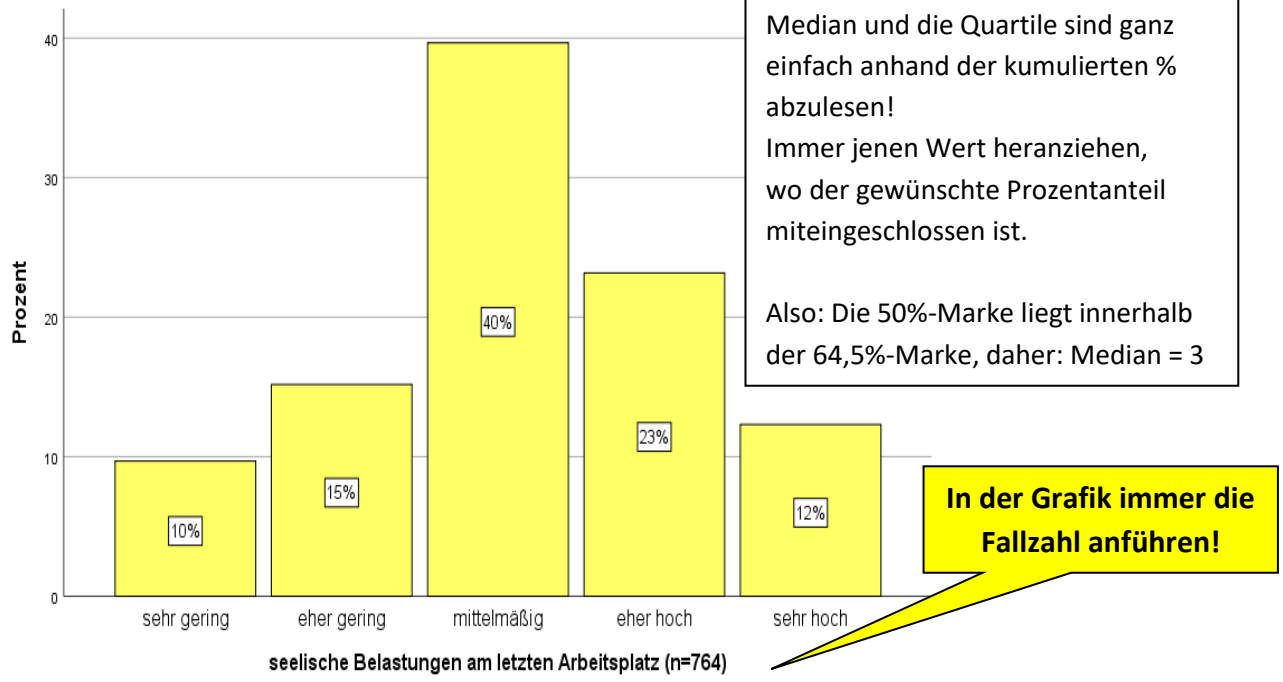
Interpretation:
Die meisten Befragten in dieser Stichprobe gaben eine mittelmäßige seelische Belastung an ihrem letzten Arbeitsplatz an (40%).
Ein Viertel war gering belastet, etwa die Hälfte (hier 64%!) war gering bis mittelmäßig belastet.

f12_b seelischen Belastungen am letzten Arbeitsplatz

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 sehr gering	74	8,9	9,7	9,7
	2 eher gering	116	14,0	15,2	24,9
	3 mittelmäßig	303	36,6	39,7	64,5
	4 eher hoch	177	21,4	23,2	87,7
	5 sehr hoch	94	11,4	12,3	100,0
	Gesamt	764	92,4	100,0	
Fehlend	System	63	7,6		
Gesamt		827	100,0		

TIPP:
Median und die Quartile sind ganz einfach anhand der kumulierten % abzulesen!
Immer jenen Wert heranziehen, wo der gewünschte Prozentanteil miteingeschlossen ist.

Also: Die 50%-Marke liegt innerhalb der 64,5%-Marke, daher: Median = 3



In der Grafik immer die Fallzahl anführen!

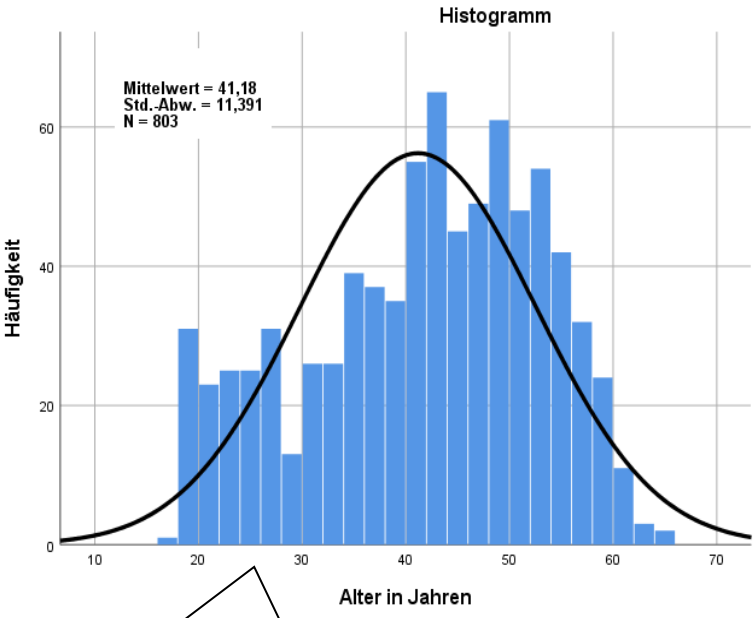
3.3 Kennzahlen bei metrischem Datenniveau

Analysieren → deskriptive Statistiken → Häufigkeiten → **alter**
→ optional: ☐ *Häufigkeitstabelle anzeigen* → ausklicken
→ Diagramme → ⊙ Histogramm mit ☑ Normalverteilungskurve
→ Statistik: Kennzahlen auswählen ☑ *Quartile, Median, Modalwert, Mittelwert, Stdabw., Varianz, Bereich (=Spannweite), Minimum, Maximum, Std.Fehler, Schiefe, Kurtosis* → OK

Statistiken

alter Alter in Jahren

N	Gültig	803
	Fehlend	24
Mittelwert		41,18
Standardfehler des Mittelwertes		,402
Median		43,00
Modus		46
Standardabweichung		11,391
Varianz		129,748
Schiefe		-,364
Standardfehler der Schiefe		,086
Kurtosis		-,796
Standardfehler der Kurtosis		,172
Spannweite		47
Minimum		17
Maximum		64
Summe		33064
Perzentile	25	33,00
	50	43,00
	75	50,00



Beachte: Die Normalverteilungskurve kann optional eingeblendet werden, um die untersuchte Verteilung mithilfe der NV als Schablone zu vergleichen. Auf diese Weise wird leichter erkannt, ob die Verteilung schief oder symmetrisch, steil oder flach ist.

Interpretation:
Modus: Die meisten Befragten sind 46 Jahre alt. (Bei metrischem Datenniveau nicht aussagekräftig!)
Mittelwert: Die Befragten dieser Stichprobe sind durchschnittlich 41 Jahre alt.
Median: Die Hälfte der Stichprobe ist jünger als 43 Jahre und die andere Hälfte ist älter als 43 Jahre.
Quartile: Ein Viertel der Befragten ist bis zu 33 Jahre alt, und drei Viertel sind bis zu 50 Jahre alt.
Spannweite: Die Spannweite der Verteilung reicht von 17 bis 64 Jahre, dies umfasst 47 Jahre.
Standardabweichung: Die durchschnittliche Streuung um den Mittelwert beträgt 11 Jahre, was bei der gegebenen Spannweite von 47 Jahren (17 bis 64) eine relativ breite Streuung darstellt.
Schiefe: Die Verteilung ist leicht rechtsgipfelig (=linksschief), dies zeigt sich anhand der negativen Schiefe =-0,364). Ein weiterer Hinweis ist, dass der Mittelwert etwas kleiner ist als der Median. Es gibt demnach vermehrt ältere Personen in dieser Stichprobe.
Kurtosis/Wölbung: Die negative Kurtosis (= -0,798) weist darauf hin, dass die Verteilung insgesamt betrachtet etwas flacher als die Normalverteilung ist.

3.4 Hausübung 3: Kennzahlen

Pflichtaufgabe 3:
Wählen Sie aus dem Datensatz zur Gesundheit von Arbeitslosen oder aus dem Datensatz zum Darmmanagement je eine Variable mit nominalem, ordinalem und metrischen Datenniveau aus. Erstellen Sie die Häufigkeitstabellen mit den jeweils dazugehörigen Kennzahlen und das passende Diagramm. Beschreiben und interpretieren Sie die Ergebnisse!
Wichtig: Bei der Interpretation der Kennzahlen immer ganze, verständliche Sätze bilden!

Fleißaufgabe für Lernwütige:
Erstellen Sie dieselbe Aufgabe wie oben aus Ihrem eigenen fiktiven Datensatz!

Fleißaufgabe für Lernwütige: Konfidenzintervall des Mittelwertes:
Analysieren → deskriptive Statistiken → Explorative Datenanalyse → Abhängige Variable alter → Anzeigen: nur Statistiken: Deskriptive Statistik;
Jetzt: Interpretieren Sie das Konfidenzintervall in Worten!
(Untergrenze - Obergrenze)

Fleißaufgabe für Lernwütige: Konfidenzintervall des Anteilswertes:
Berechnen Sie den Anteilswert der Befragten, die nicht in Österreich geboren sind (f8) und interpretieren Sie es in Worten!
p = Anteilswert = Stichprobenwahrscheinlichkeit (%-Wert/100); n = Stichprobengröße (gültige n);
95% Konfidenzintervall:

$$KI_{95\%} = p - 1{,}96 * S_p ; p + 1{,}96 * S_p \quad \text{wobei:} \quad S_p = \sqrt{\frac{p*(1-p)}{n}}$$

4 Neue Variablen erstellen

4.1 Rekodieren

Sehr oft werden Variablen rekodiert, das heißt, die original vorliegenden Ausprägungen der Antwortskala werden weiter zusammengefasst, um damit eine einfache erste Auswertungsmethode anzuwenden, nämlich die Kreuztabelle.

Machen Sie keine Auswertung ohne einer Fragestellung im Hintergrund!
Fragestellung: Besteht ein Zusammenhang zwischen Rauchen und Atemnot?

Vorgangsweise bei der Rekodierung:

1. Schritt: Festlegen der neuen **Gruppeneinteilung**:

Bestimmen Sie anhand der **Häufigkeitstabelle**, am besten anhand der kumulierten Prozent, welche Gruppeneinteilung für Ihre Auswertung am sinnvollsten ist. Dabei sind ausreichende Fallzahlen zu beachten.

WICHTIG: Vor jeder Rekodierung: Häufigkeitstabelle machen!
Anhand der Häufigkeitsverteilung und anhand inhaltlicher Kriterien werden die Kategorien bestimmt (Notizen machen, welche Gruppe zu welcher Kategorie zusammengefasst werden sollen!).

Statistiken

		f20_a Rauchen	f18_d Atemnot
N	Gültig	793	773
	Fehlend	34	54

f20_a Rauchen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 nie	344	41,6	43,4	43,4
	2 1-10 Zigaretten am Tag	122	14,8	15,4	58,8
	3 11-20 Zigaretten am Tag	203	24,5	25,6	84,4
	4 21-40 Zigaretten am Tag	105	12,7	13,2	97,6
	5 mehr als 40 Zigaretten am Tag	19	2,3	2,4	100,0
	Gesamt	793	95,9	100,0	
Fehlend	System	34	4,1		
Gesamt		827	100,0		

Systemdefiniert fehlende Werte

f18_d Atemnot

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 (fast) täglich	71	8,6	9,2	9,2
	2 alle paar Tage	73	8,8	9,4	18,6
	3 alle paar Wochen	80	9,7	10,3	29,0
	4 alle paar Monate	109	13,2	14,1	43,1
	5 nie	440	53,2	56,9	100,0
	Gesamt	773	93,5	100,0	
Fehlend	0	54	6,5		
Gesamt		827	100,0		

Benutzerdefiniert fehlende Werte

2. Schritt: Bevor wir Rekodieren machen wir eine **Dokumentation** über diesen Arbeitsschritt:

Variable f**20_a** (Rauchen)	→	Variable **f20_a_kat3** Rauchgewohnheiten in 3 Gruppen
1	→	1 "nie"
2-3	→	2 "1-20 Zig."
4-5 (4 thru Highest)	→	3 "21 Zig. ++"
Systemdefiniert fehlend (SYSMIS)	→	99 "keine Angabe"

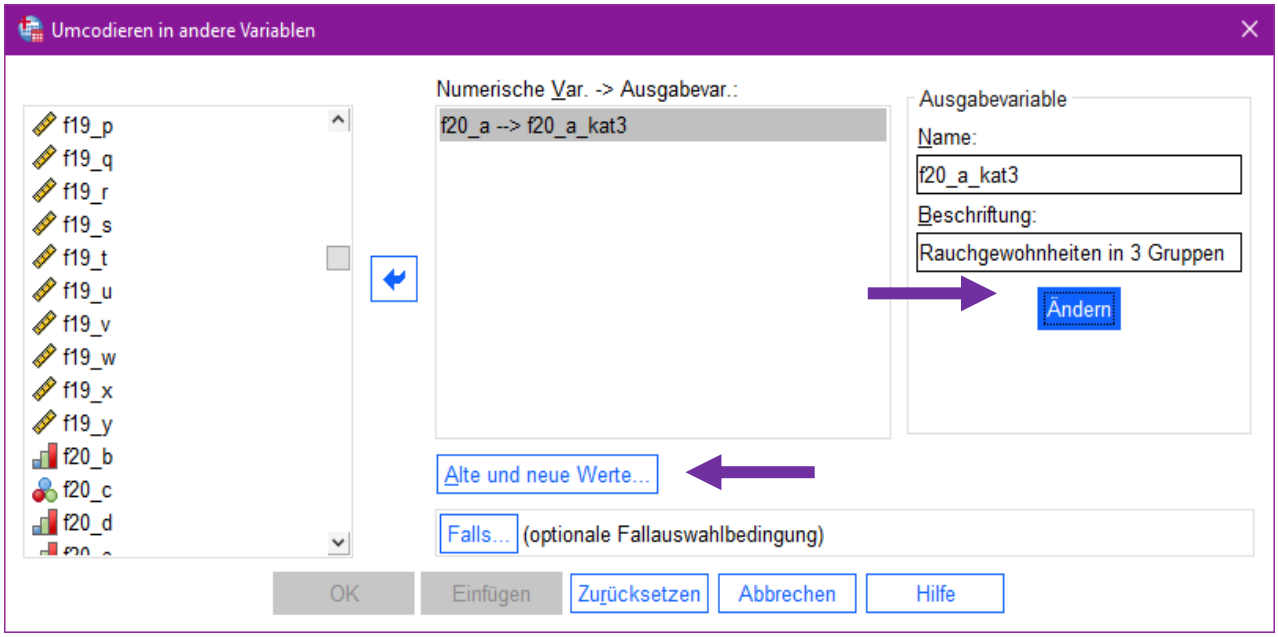
Variable **f18_d** (Atemnot)	→	Variable **f18_d_kat3** Atemnot in 3 Gruppen
1-2 (Lowest thru 2)	→	1 "tgl./alle paar Tage"
3-4	→	2 "alle paar Wo/Mo"
5	→	3 "nie"
System- oder benutzerdefiniert fehlende Werte (MISSING)	→	99 "keine Angabe"

3. Schritt: Durchführen der Rekodierung:

Transformieren →Umkodieren in **andere** Variablen…

1.) Fenster:

→ *jeweilige Variablen auswählen* [**f20_a**]→ *Ausgabevariable: Variablennamen eingeben* [**f20_a_kat3**]
→ *Beschriftung eingeben [„Rauchgewohnheiten in 3 Gruppen]* → Ändern

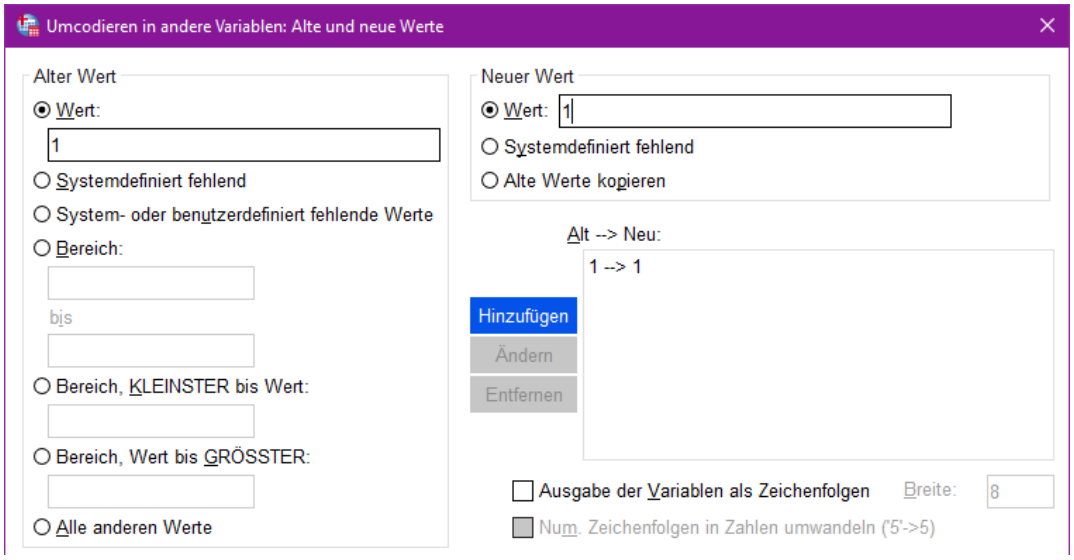


2.) Fenster:

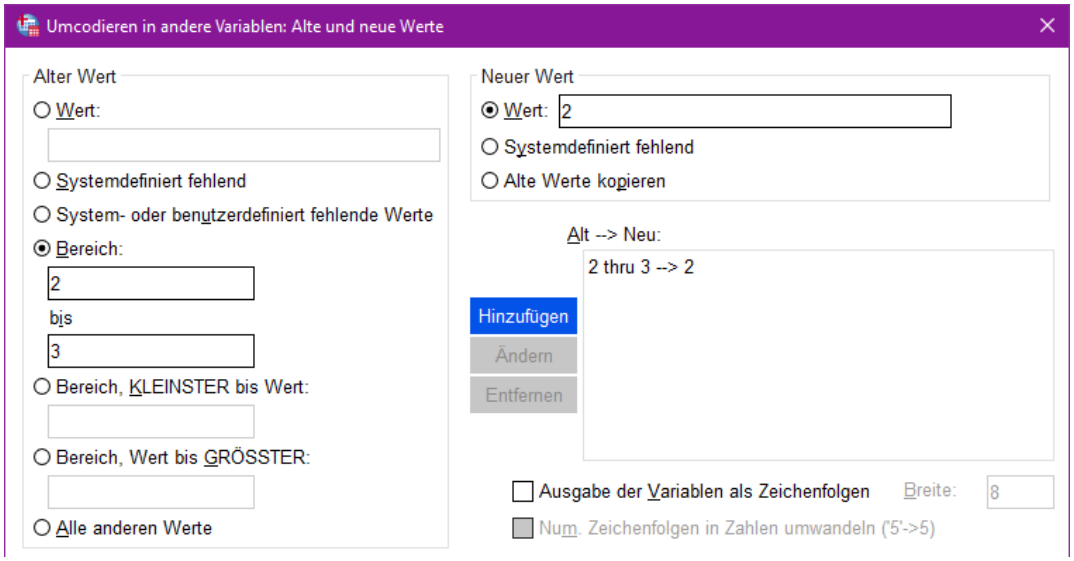
Alte und neue Werte → *Alte und entsprechende neue Werte eintragen*:

Alter Wert: 1 → *Neuer Wert:* ⊙ *Wert: 1* → Hinzufügen
Alter Wert: "Bereich" 2 bis 3 → *Neuer Wert:* ⊙ *Wert: 2* → Hinzufügen
Alter Wert: „Bereich Wert bis GRÖSSTER“ 4 → *Neuer Wert:* ⊙ *Wert: 3* → Hinzufügen
Alter Wert: ⊙ Systemdefiniert fehlend → *Neuer Wert:* ⊙ *Wert: 99* → Hinzufügen
→ Weiter → OK

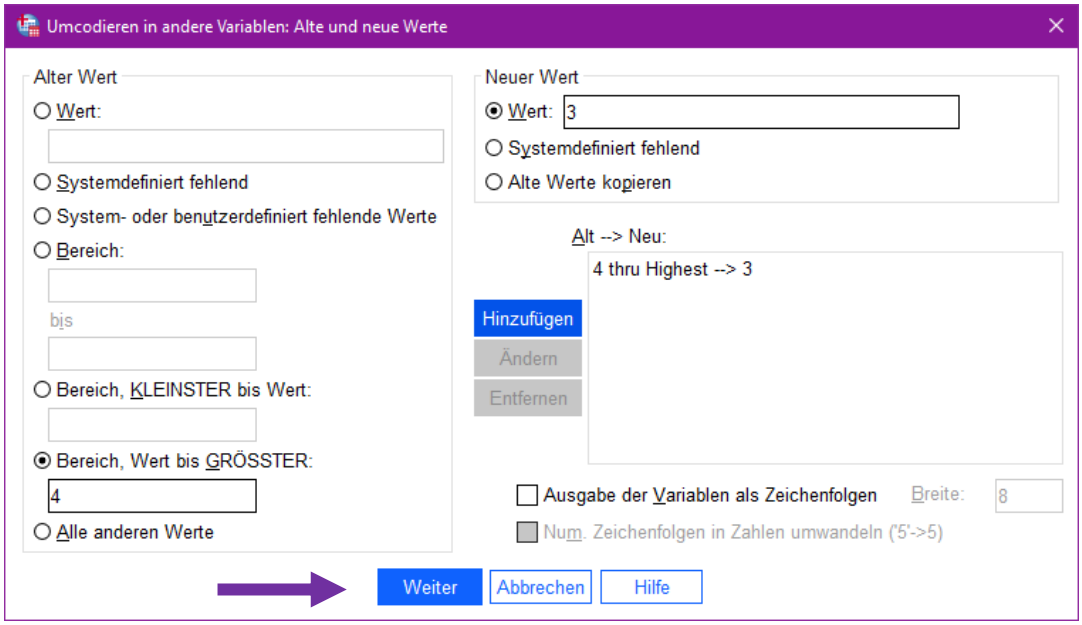
Umkodieren: Einzelne Werte = alter Wert → neuer Wert → Hinzufügen



Umkodieren: Bereich = alle Werte in dem angegebenen Intervall



Umkodieren: Bereich, Wert bis GRÖSSTER = alle Werte AB dem angegebenen Wert
Umkodieren: Bereich, KLEINSTER bis Wert = alle Werte BIS zu dem angegebenen Wert



4. Schritt: Labeln, kontrollieren, speichern!

In der Variablenansicht sehen wir ganz unten die neuen Variable f20_a_kat3 und f18_d_kat3.

→ neu gebildete Variablen sind immer ganz unten in der Liste
→ in der Spalte “Wertelabels” jeweilige Werte und Labels eintragen
→ in der Spalte "Fehlend" definieren der fehlenden Werte (99)

WICHTIG: Wir **labeln die Codes** laut unserer Dokumentation
WICHTIG: Wir **speichern** den Datensatz und
WICHTIG: **Kontrolle! Nach JEDER Rekodierung**: Häufigkeitstabelle erstellen! Vergleichen mit der ursprünglichen Häufigkeitstabelle und sicherstellen, dass der Rekodierung kein Fehler passiert ist:

Stimmt die Anzahl der fehlenden Fälle? Stimmen die Anteilswerte kumulierten Prozentwerte der neuen Variable mit den der ursprünglichen Variable überein?

Statistiken

		f20_a_kat3 Rauchgewohn-heiten in 3 Gruppen	f18_d_kat3 Atemnot in 3 Gruppen
N	Gültig	793	773
	Fehlend	34	54

f20_a_kat3 Rauchgewohnheiten in 3 Gruppen

		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 nie	344	41,6	43,4	43,4
	2 1-20 Zig.	325	39,3	41,0	84,4
	3 21 Zig. ++	124	15,0	15,6	100,0
	Gesamt	793	95,9	100,0	
Fehlend	99 keine Angabe	34	4,1		
Gesamt		827	100,0		

f18_d_kat3 Atemnot in 3 Gruppen

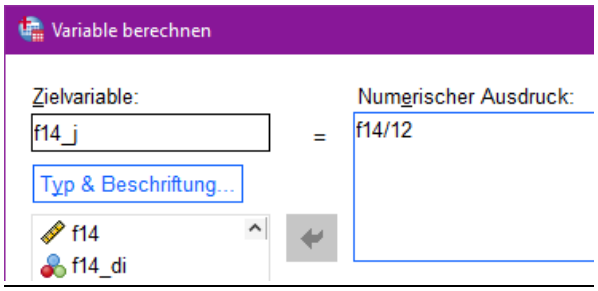
		Häufigkeit	Prozent	Gültige Prozente	Kumulierte Prozente
Gültig	1 tgl./alle paar Tage	144	17,4	18,6	18,6
	2 alle paar Wo/Mo	189	22,9	24,5	43,1
	3 nie	440	53,2	56,9	100,0
	Gesamt	773	93,5	100,0	
Fehlend	99 keine Angabe	54	6,5		
Gesamt		827	100,0		

WICHTIG: Speichern Sie Ihren bearbeiteten Datensatz **IMMER** ab!

4.2 Variable berechnen

Im Datensatz findet sich die Variable f14 "Dauer der Arbeitslosigkeit in Monaten". Manchmal ist es sinnvoll, hier eine andere Skala zu verwenden. Wir rechnen nun die Dauer der Arbeitslosigkeit (die in Monaten erfasst wurde) in Jahre um.

Transformieren → Variable berechnen →*Namen für Zielvariable eingeben* [**f14_j**]
→*Variable auswählen [f14] und in das Feld „numerischer Ausdruck" übertragen*
„f14/12" (Division durch 12) → OK



Jetzt wieder kontrollieren, ob alles fehlerfrei ist.

Statistiken

		f14 Dauer der Arbeitslosigkeit in Monaten	f14_j Dauer AL in Jahren
N	Gültig	747	747
	Fehlend	80	80
Mittelwert		34,02	2,8354
Minimum		1	,08
Maximum		240	20,00

WICHTIG: Speichern Sie Ihren bearbeiteten Datensatz **IMMER** ab!

4.3 Hausübung 4: Rekodieren

Pflichtaufgabe 4: Rekodieren
Wählen Sie einen der Datensätze für diese Aufgabe aus (Arbeitslose oder Darmmanagement). Überlegen Sie sich eine relevante **Fragestellung**, indem Sie zwei Variablen miteinander in Verbindung bringen, die etwas miteinander zu tun haben könnten. Begründen Sie inhaltlich, warum Sie glauben, dass die beiden Merkmale miteinander zusammenhängen. Welches Ergebnis erwarten Sie?

Erstellen Sie die **Häufigkeitstabellen** dieser beiden Variablen und **rekodieren** Sie beide Variablen derart, dass beide in zwei bis drei sinnvollen Kategorien vorliegen. Begründen Sie Ihre Entscheidung bei der Zusammenfassung der Kategorien! (Labeln und Daten speichern nicht vergessen!)

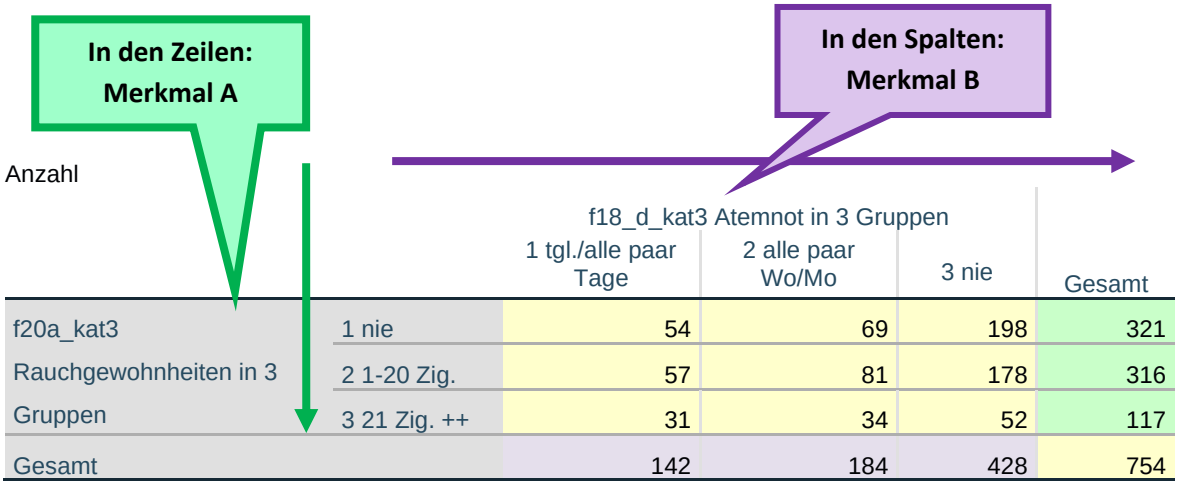
Fleißaufgabe für Lernwütige:
Erstellen Sie dieselbe Aufgabe wie oben aus Ihrem eigenen fiktiven Datensatz!
Für eine optimale Lernerfahrung verwenden Sie dazu eine metrische Variable!

5 Kreuztabelle

Eine sehr **einfache und elementare Auswertungsmethode** stellt die Kreuztabelle dar. Dabei werden – wie der Name schon sagt – zwei Merkmale gekreuzt, also miteinander in Zusammenhang gebracht. Mithilfe einer Kreuztabelle können einfache Fragestellungen sehr rasch beantwortet werden. Kreuztabellen mit zwei Variablen, die dichotom sind (also nur zwei Ausprägungen haben) nennt man 2x2-Kreuztabellen oder Vierfeldertafeln.
Bei einer Kreuztabelle kann angezeigt werden:

Anzahl	Anzahl der tatsächlich beobachteten Fälle in einer Zelle
% von Zeilenvariable bzw. Randsummen Zeile	Zeilensummen: einfache Häufigkeits- od. Randverteilung der Zeilenvariable
% von Spaltenvariable bzw. Randsummen Spalte	Spaltensummen: einfache Häufigkeits- od. Randverteilung der Spaltenvariable
% von Zeilenvariable (Zeilenprozent)	Fälle in den Zellen als Prozentanteile an den Fällen der zugehörigen Zeile (Zeilensumme) ausgedrückt: Anzahl / Zeilensumme *100
% von Spaltenvariable (Spaltenprozent)	Fälle in den Zellen als Prozentanteile an den Fällen der zugehörigen Spalte (Spaltensumme) ausgedrückt: Anzahl / Spaltensumme *100
% von Gesamtzahl (Gesamtprozent)	Fälle in den Zellen werden als Prozentanteile an allen gültigen Fällen ausgedrückt
Erwartete Anzahl	Anzahl der Fälle in einer Zelle, die erwartet werden, wenn kein Zusammenhang zwischen den beiden Variablen besteht, d.h. wenn sie unabhängig voneinander sind

Die **Freiheitsgrade** (degrees of freedom) einer Kreuztabelle sind ein Maß für die Größe der Kreuztabelle. Die Anzahl der Zeilen minus 1 x der Anzahl der Spalten minus 1 = ergibt die Freiheitsgrade. Eine Tabelle mit 3 Zeilen und 3 Spalten (exklusive der Randverteilung) hat somit 4 Freiheitsgrade (2 x 2 = 4).



		f18_d_kat3 Atemnot in 3 Gruppen			
		1 tgl./alle paar Tage	2 alle paar Wo/Mo	3 nie	Gesamt
f20a_kat3 Rauchgewohnheiten in 3 Gruppen	1 nie	54	69	198	321
	2 1-20 Zig.	57	81	178	316
	3 21 Zig. ++	31	34	52	117
Gesamt		142	184	428	754

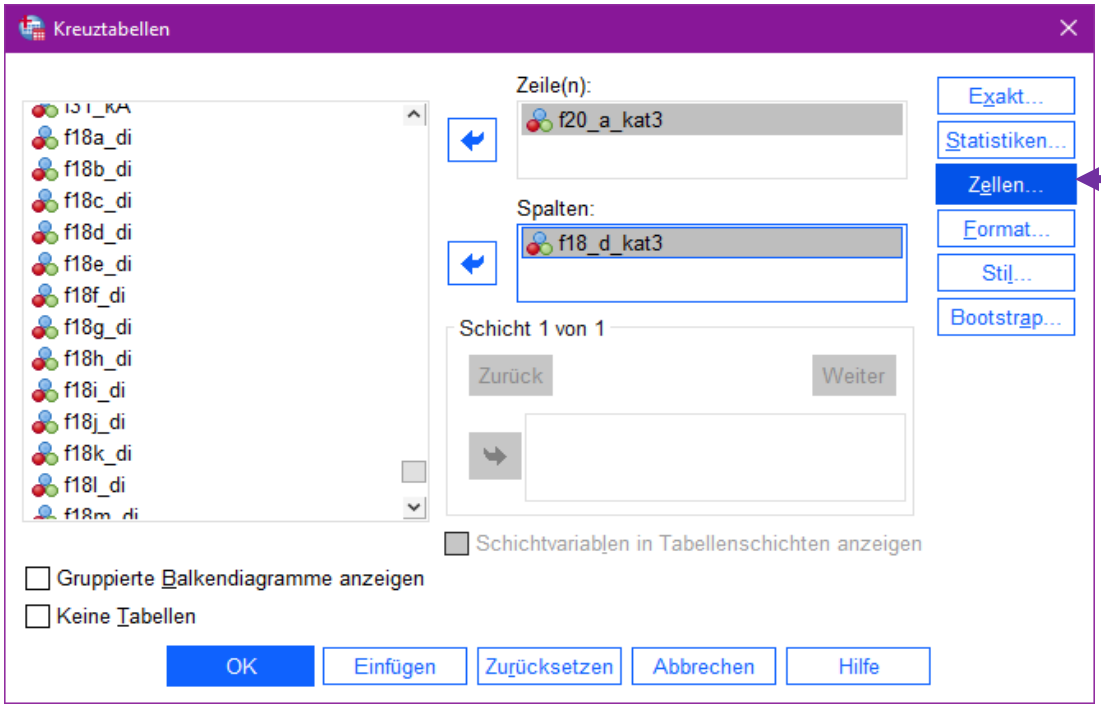
5.1 Erstellen und Interpretieren einer Kreuztabelle

Machen Sie keine Auswertung ohne einer Fragestellung im Hintergrund!

Fragestellung: Besteht ein Zusammenhang zwischen Rauchen und Atemnot?

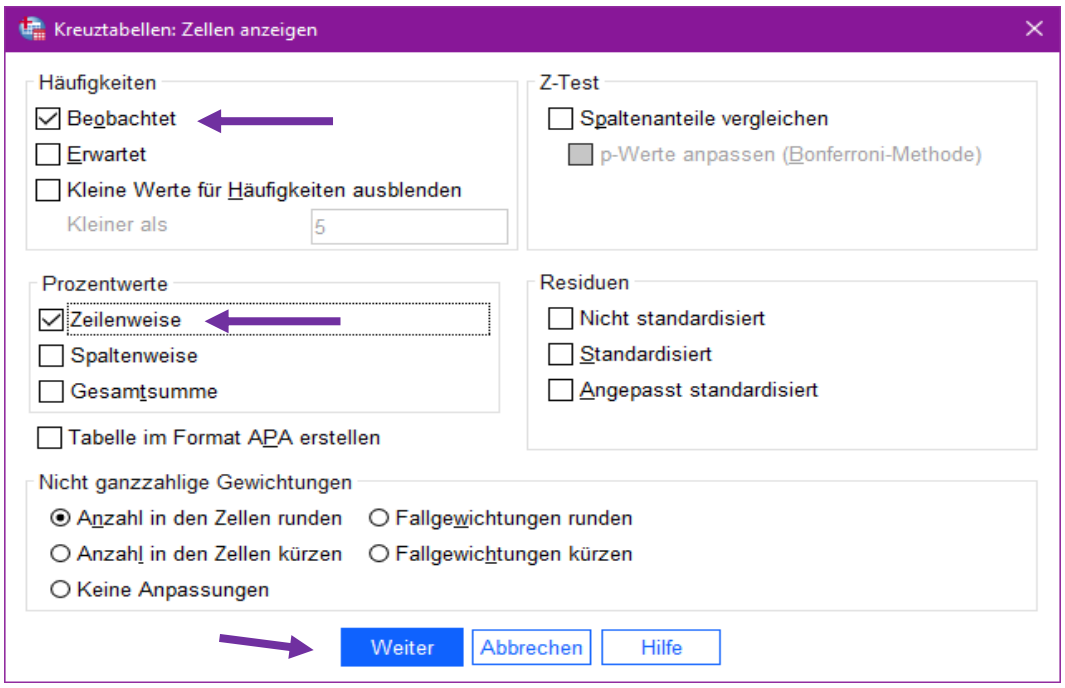
VOR jeder Kreuztabelle müssen die verwendeten Variablen mit Häufigkeitstabelle „inspiziert" werden! Bevor die Kreuztabelle gemacht wird, müssen die Variablen zumeist in weniger Kategorien rekodiert werden, um eine genügende Fallzahl in jeder Zelle der Kreuztabelle zu gewährleisten.

Analysieren → Deskriptive Statistiken → **Kreuztabellen**… Zeilen: **f20_a_kat3**
Spalten: **f18_d_kat3**



Zeilenprozent:

Zellen → beobachtete Werte und **Zeilenprozentwerte** → Weiter → OK



Zeilenprozent:

f20_a_kat3 Rauchgewohnheiten in 3 Gruppen * f18_d_kat3 Atemnot in 3 Gruppen Kreuztabelle

			f18_d_kat3 Atemnot in 3 Gruppen			
			1 tgl./alle paar Tage	2 alle paar Wo/Mo	3 nie	Gesamt
f20_a_kat3 Rauchgewohn-heiten in 3 Gruppen	1 nie	Anzahl	54	69	198	321
		% innerhalb f20_a_kat3	16,8%	21,5%	61,7%	100,0%
	2 1-20 Zig.	Anzahl	57	81	178	316
		% innerhalb f20_a_kat3	18,0%	25,6%	56,3%	100,0%
	3 21 Zig. ++	Anzahl	31	34	52	117
		% innerhalb f20_a_kat3	26,5%	29,1%	44,4%	100,0%
Gesamt		Anzahl	142	184	428	754
		% innerhalb f20_a_kat3	18,8%	24,4%	56,8%	100,0%

ZEILENPROZENT Zelle Nummer 1: 54 Befragte, die nie rauchen und täglich/alle paar Tage unter Atemnotbeschwerden leiden:

Von allen Befragten, die nie rauchen, leiden etwa 17% täglich oder alle paar Tage an Atemnot.

$$\frac{n_1}{n_{row}} * 100 = \frac{54}{321} * 100 = 16{,}8\%$$

ZEILENPROZENT

Interpretation der Zeilenprozent:

Insgesamt leiden knapp 19% der befragten Langzeitarbeitslosen täglich oder alle paar Tage an Atemnot-beschwerden.

Dieser Anteil variiert nach der Häufigkeit des Rauchens: Je mehr Zigaretten die Befragten pro Tag rauchen, desto häufiger leiden sie an Atemnot:

Der Anteil derjenigen Befragten, die täglich oder alle paar Tage an Atemnot leiden steigt von 17% bei den Nichtraucher*innen auf 18% bei den Befragten, die bis zu 20 Zigaretten pro Tag rauchen, auf bis zu 26% bei jenen Befragten, die 21 und mehr Zigaretten täglich rauchen.

Umgekehrt sinkt der Anteil derjenigen Befragten, die nie unter Atemnot leiden mit steigendem Zigarettenkonsum: Insgesamt leiden 57% der Befragten niemals unter Atemnot.

Unter den Nichtraucher*innen ist der Anteil jener, die niemals Atemnot haben, 62%, unter jenen, die 1-20 Zigaretten pro Tag rauchen haben, leiden 56% niemals an Atemnot und schließlich ist der Anteil jener, die niemals Atemnot haben bei den starken Raucher*innen (mehr als 20 Zigaretten pro Tag) am geringsten, nämlich 44%.

Spaltenprozent:

Analysieren → Deskriptive Statistiken → **Kreuztabellen**…	Zeilen: **f20_a_kat3**
	Spalten: **f18_d_kat3**
Zellen→ beobachtete Werte und **Spaltenprozentwerte**	→ Weiter → OK

f20_a_kat3 Rauchgewohnheiten in 3 Gruppen * f18_d_kat3 Atemnot in 3 Gruppen Kreuztabelle

			f18_d_kat3 Atemnot in 3 Gruppen			
			1 tgl./alle paar Tage	2 alle paar Wo/Mo	3 nie	Gesamt
f20_a_kat3 Rauchgewohn-heiten in 3 Gruppen	1 nie	Anzahl	54	69	198	321
		% innerhalb f18_d_kat3	38,0%	37,5%	46,3%	42,6%
	2 1-20 Zig.	Anzahl	57	81	178	316
		% innerhalb f18_d_kat3	40,1%	44,0%	41,6%	41,9%
	3 21 Zig. ++	Anzahl	31	34	52	117
		% innerhalb f18_d_kat3	21,8%	18,5%	12,1%	15,5%
Gesamt		Anzahl	142	184	428	754
		% innerhalb f18_d_kat3	100,0%	100,0%	100,0%	100,0%

SPALTENPROZENT Zelle Nummer 1: 54 Befragte, die nie rauchen und tgl/alle paar Tage Atemnot-beschwerden haben: Von allen Befragten, die täglich oder alle paar Tage unter Atemnot leiden, sind 38% Nichtraucher*innen.

SPALTENPROZENT

$$\frac{n_1}{n_{col}} * 100 = \frac{54}{142} * 100 = 38{,}0\%$$

Interpretation der Spaltenprozent:
Insgesamt gibt es in der Stichprobe 43% Befragte, die angeben, Nicht-Raucher*innen zu sein. Dieser Anteil ist unter jenen, die an Atemnot leiden, geringer, nämlich etwa 38%. Unter jenen Befragten, die niemals an Atemnot leiden ist dieser Anteil jedoch höher, nämlich 46%.
Insgesamt sind in der Stichprobe 15% starke Raucher*innen, die täglich mehr als 20 Zigaretten rauchen. Dieser Anteil ist unter jenen, die täglich oder alle paar Tage an Atemnot leiden, höher, nämlich 22%. Unter jenen Befragten, die niemals Atemnot haben, ist dieser Anteil am geringsten, nämlich 12%.

Gesamtprozent:

Analysieren → Deskriptive Statistiken → **Kreuztabellen**… Zeilen: **f20_a_kat3**
Spalten: **f18_d_kat3**
Zellen → beobachtete Werte und **Gesamtprozentwerte** → Weiter → OK

f20_a_kat3 Rauchgewohnheiten in 3 Gruppen * f18_d_kat3 Atemnot in 3 Gruppen Kreuztabelle

			f18_d_kat3 Atemnot in 3 Gruppen			
			1 tgl./alle paar Tage	2 alle paar Wo/Mo	3 nie	Gesamt
f20_a_kat3 Rauchgewohn-heiten in 3 Gruppen	1 nie	Anzahl	54	69	198	321
		% der Gesamtzahl	7,2%	9,2%	26,3%	42,6%
	2 1-20 Zig.	Anzahl	57	81	178	316
		% der Gesamtzahl	7,6%	10,7%	23,6%	41,9%
	3 21 Zig. ++	Anzahl	31	34	52	117
		% der Gesamtzahl	4,1%	4,5%	6,9%	15,5%
Gesamt		Anzahl	142	184	428	754
		% der Gesamtzahl	18,8%	24,4%	56,8%	100,0%

GESAMTPROZENT: Zelle 1: 54 Befragte: In der Stichprobe gibt es 7% Befragte, die nie rauchen und täglich/alle paar Tage Atemnot haben.

$$\frac{n_1}{n_{total}} * 100 = \frac{54}{754} * 100 = 7{,}2\%$$

GESAMTPROZENT

Interpretation der Gesamtprozent:
Die meisten Befragten, nämlich 26% der Stichprobe, rauchen nicht und leiden auch nicht an Atemnot-beschwerden. Fast ebensoviele, nämlich 24% der Befragten, sind "mäßige" Raucher*innen, die eine bis 20 Zigaretten pro Tag rauchen und niemals Atemnot haben.
Jeweils rund 10% der Stichprobe sind Nicht-Raucher*innen oder "gemäßigte" Raucher*innen, die alle paar Wochen unter Atemnot leiden. Starke Raucher*innen, die mehr als 20 Zigaretten pro Tag rauchen und niemals an Atemnot leiden sind 7% der Stichprobe. Ebenfalls jeweils 7% der Stichprobe sind Nicht-Raucher*innen und "mäßige" Raucher*innen mit täglich bis alle paar Tage Atemnotbeschwerden.
Sehr klein sind die Anteile jener Befragten, die starke Raucher*innen sind und unter Atemnot leiden: nämlich 4% jener, die alle paar Wochen Atemnot haben und ebenso 4%, die täglich oder alle paar Tage unter Atemnot leiden.

WICHTIG: Beachte: Gesamtprozent sind **rein DESKRIPTIV** und sagen **NICHTS** über den Zusammenhang der beiden Merkmale aus !!

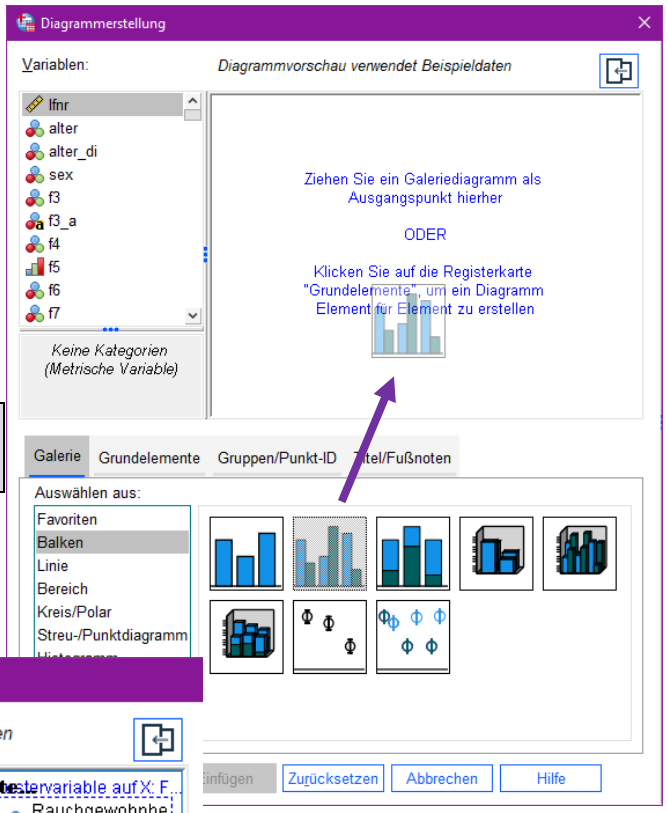
5.2 Balkendiagramm

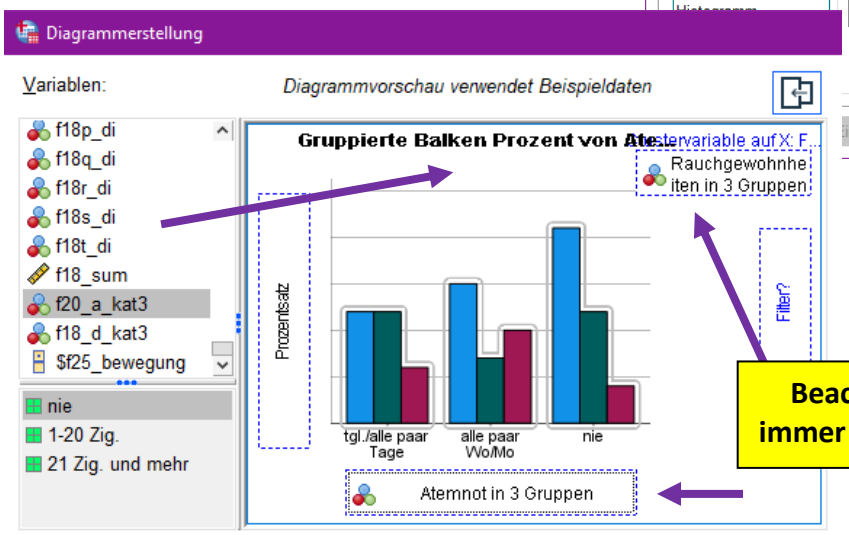
Grafik → **Diagrammerstellung**
Galerie: Balken → "Gruppierte Balken"
in die Diagrammvorschau ziehen

X-Achse = abhängige Variable

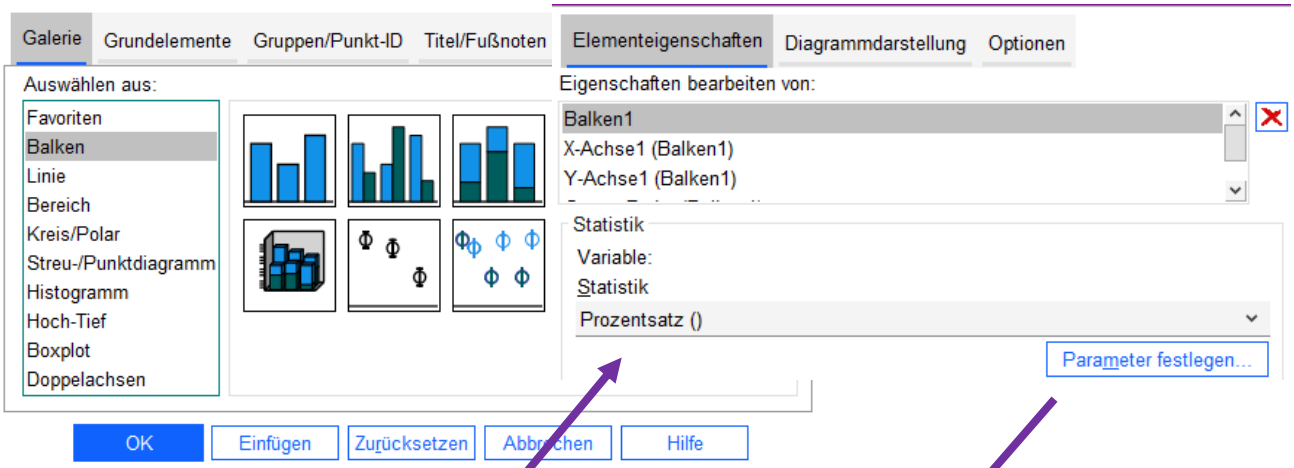
Variable **f18_d_kat3** (Atemnot) in die x-Achse
Variable **f20a_kat3** (Rauchen) in die Clustervariable

Clustervariable (Gruppenvariable) = unabhängige Variable

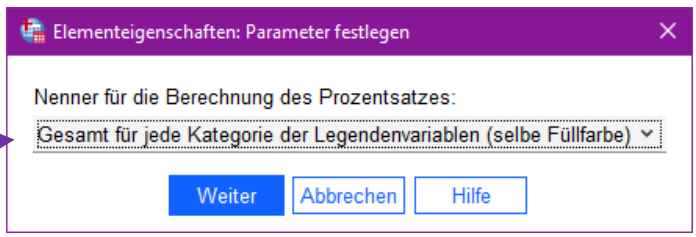


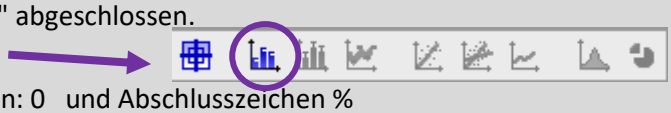


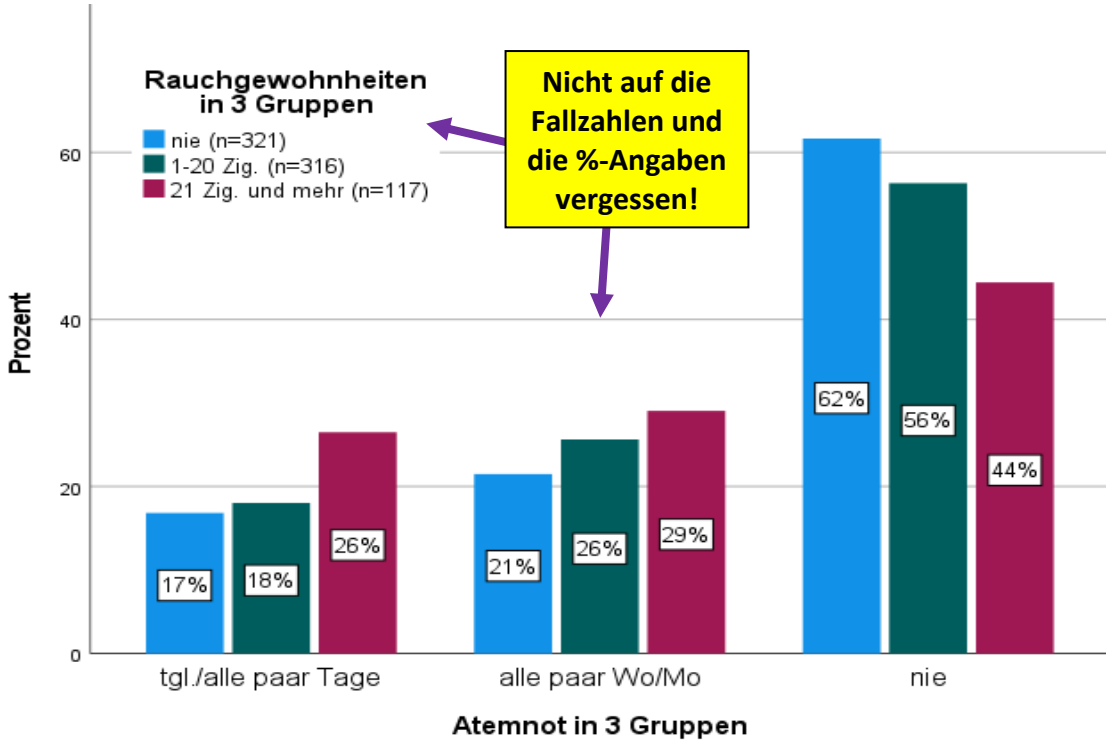
Beachte: Die Clustervariable ist immer die, wo die 100% sein sollen!



Im Fenster **Elementeigenschaften**
Balken1 → Statistik "**Prozentsatz ()**"
→ Parameter festlegen: wählen:
"**Gesamt für jede Kategorie der Legendenvariable (selbe Füllfarbe)**"
→ Weiter → Anwenden



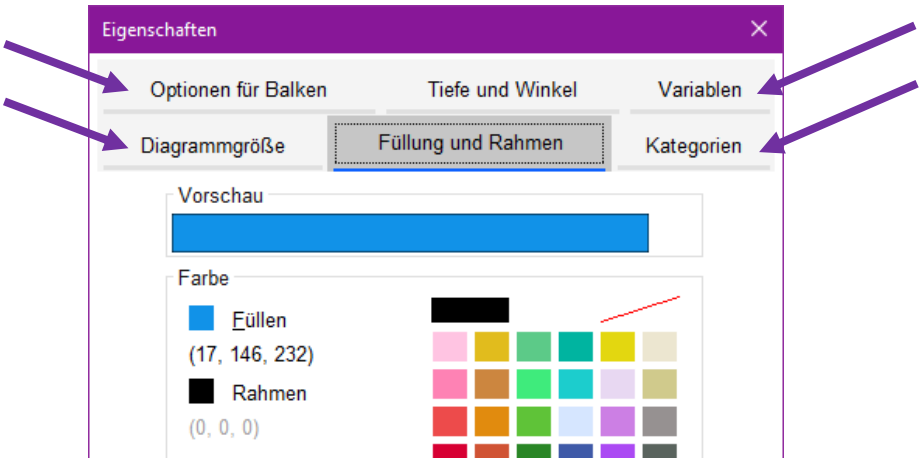
Mit Doppelklick auf die Grafik kann diese in einem Eigenschaftsfenster bearbeitet werden.
Jede Änderung wird mit dem Befehl "Zuweisen" abgeschlossen.
→ **Elemente**: Datenbeschriftung einblenden 
→ **Eigenschaften**: Zahlenformat: Dezimalstellen: 0 und Abschlusszeichen %
→ Doppelklick auf Y-**Skalierung**: Zahlenformat: Dezimalstellen: 0



Interpretation (Vergleiche Kreuztabelle auf Seite 32 (Zeilenprozent):

Die Anteile jener Befragten, die sehr oft unter Atemnot leiden, nehmen mit der Rauchhäufigkeit zu: Beispielsweise leiden etwas mehr als ein Viertel (26%) der starken Raucher*innen täglich bis alle paar Tage unter Atemnot. Demgegenüber nehmen die Anteile jener, die keine Atemnot haben, mit der Rauchhäufigkeit ab: 62% Nichtraucher*innen, 56% der gemäßigten Raucher*innen und lediglich 44% der starken Raucher*innen leiden nie unter Atemnot.

Doppelklick auf Balken..... → ..... Bearbeiten nach Belieben
Vorgangsweise: Zuerst das zu ändernde Objekt mit **Doppelklick** markieren, dann im **Eigenschaftsfenster** den entsprechenden **Reiter auswählen**, Füllung / Rahmen, Kategorien..... etc. formatieren → Befehl mit "**Anwenden**" abschließen.



5.3 Dreidimensionale Kreuztabelle: Einfügen einer Kontrollvariable

Fragestellung: *Besteht der Zusammenhang zwischen Rauchen und Atemnot unabhängig vom Geschlecht?*

Analysieren → Deskriptive Statistiken → Kreuztabellen… Zeilen: f20_a_3kat
Spalten: f18_d_3kat
Schicht: sex
Zellen→ beobachtete Werte und Zeilenprozentwerte → Weiter → OK

				f18_d_kat3 Atemnot in 3 Gruppen			
sex Geschlecht				1 tgl./alle paar Tage	2 alle paar Wo/Mo	3 nie	Gesamt
1 männlich	f20_a_kat3 Rauch-gewohn-heiten in 3 Gruppen	1 nie	Anzahl	24	29	113	166
			%	14,5%	17,5%	68,1%	100,0%
		2 1-20 Zig.	Anzahl	21	39	116	176
			%	11,9%	22,2%	65,9%	100,0%
		3 21 Zig. ++	Anzahl	23	26	36	85
			%	27,1%	30,6%	42,4%	100,0%
	Gesamt		Anzahl	68	94	265	427
			%	15,9%	22,0%	62,1%	100,0%
2 weiblich	f20_a_kat3 Rauch-gewohn-heiten in 3 Gruppen	1 nie	Anzahl	29	40	82	151
			%	19,2%	26,5%	54,3%	100,0%
		2 1-20 Zig.	Anzahl	36	41	59	136
			%	26,5%	30,1%	43,4%	100,0%
		3 21 Zig. ++	Anzahl	8	7	15	30
			%	26,7%	23,3%	50,0%	100,0%
	Gesamt		Anzahl	73	88	156	317
			%	23,0%	27,8%	49,2%	100,0%
Gesamt	f20_a_kat3 Rauch-gewohn-heiten in 3 Gruppen	1 nie	Anzahl	53	69	195	317
			%	16,7%	21,8%	61,5%	100,0%
		2 1-20 Zig.	Anzahl	57	80	175	312
			%	18,3%	25,6%	56,1%	100,0%
		3 21 Zig. ++	Anzahl	31	33	51	115
			%	27,0%	28,7%	44,3%	100,0%
	Gesamt		Anzahl	141	182	421	744
			%	19,0%	24,5%	56,6%	100,0%

Interpretation:
Bei den **Männern** zeigen sich die Gruppenunterschiede zwischen den Nicht-, den gemäßigten und den starken Rauchern sehr deutlich:
Die Anteile jener, die häufig unter Atemnot leiden, nehmen je nach Rauchhäufigkeit zu, und zwar von 14% auf 27%.
Dies gilt auch für jene, die alle paar Wochen bzw. Monate unter Atemnot leiden: der Anteil jener Gruppe nimmt ebenfalls mit der Rauchhäufigkeit zu, und zwar von 17% auf 31%.
Ebenso nehmen die Anteile jener, die keine Atemnot haben, nach Rauchhäufigkeit kontinuierlich ab, und zwar von 68% auf 42%.

Bei den **Frauen** zeigen sich diese Zusammenhänge viel weniger deutlich.
Die Anteile jener Frauen, die häufig unter Atemnot leiden, nehmen auch je nach Rauchhäufigkeit zu, und zwar von 19% auf 27%.
Die Anteile jener, die keine Atemnot haben, nehmen nach Rauchhäufigkeit nicht kontinuierlich ab, die Anteile schwanken hier zwischen 40% und 50%.
Auch bei jenen, die alle paar Wochen bzw. Monate unter Atemnot leiden schwanken die Anteile zwischen 23% und 30% und nehmen nicht mit der Rauchhäufigkeit ab.
Zu beachten ist außerdem die besonders geringe Anzahl der starken Raucherinnen (30 Frauen).

5.4 Hausübung 5: Kreuztabelle

Pflichtaufgabe 5: Kreuztabelle
Erstellen Sie mit den von Ihnen in Aufgabe 4 rekodierten Variablen eine **Kreuztabelle**.
Zuerst mit den Zeilenprozent, dann mit den Spaltenprozent, dann mit den Gesamtprozentwerten.
Beschreiben und interpretieren Sie die Ergebnisse anhand der Zeilen-, der Spalten-, und der Gesamtprozentwerte. Was bedeuten die Ergebnisse inhaltlich? Hat sich Ihre Vermutung bestätigt? Wenn nicht, haben Sie eine Erklärung dafür?

Fleißaufgabe für Lernwütige: Erstellen Sie zu dieser Aufgabe das zweidimensionale Balkendiagramm!

Fleißaufgabe für Lernwütige: Fügen Sie weiters eine dichotome Schichtvariable zu Ihrer Frage-stellung ein (nicht die Variable Geschlecht!) Gegebenenfalls rekodieren Sie die Variable Ihrer Wahl hierfür auf 2 Kategorien. Beschreiben und interpretieren Sie die Ergebnisse.

Fleißaufgabe für Lernwütige: Erstellen Sie dieselbe Aufgabe mit Ihrem eigenen fiktiven Datensatz!

5.5 Chi-Quadrat-Test bei Kreuztabellen

Erwartete Anzahl	Anzahl der Fälle in einer Zelle, die erwartet werden, wenn kein Zusammenhang zwischen den beiden Variablen besteht, d.h. wenn sie unabhängig voneinander sind
Residuen	Differenzen zwischen beobachteten und erwarteten Werten
Standardisierte Residuen	Diese Differenzen als standardisierte Werte
Chi-Quadrat	Testverfahren für die Analyse von Häufigkeitsunterschieden im Auftreten bestimmter Merkmale durch den Vergleich von beobachteten und erwarteten Häufigkeiten
Signifikanz (=WS des Prüfergebnisses unter der Annahme der Nullhypothese) (= Irrtums-WS bei Annahme eines Zusammenhangs)	ermittelt auf Basis des Chi-Quadrat-Wertes und der Freiheitsgrade *größer als α:* Nullhypothese (es besteht kein Zusammenhang) wird vorläufig beibehalten. *kleiner als α:* Nullhypothese kann abgelehnt und die Alternativhypothese (es besteht ein Zusammenhang)angenommen werden.
Zellen mit erwarteter Häufigkeit kleiner 5	Anzahl und Anteil der Zellen an der gesamten Kreuztabelle mit Erwartungswerten < 5 → Anteil darf nicht > 20% sein!

Wir testen nun mit der von uns bereits erstellen Kreuztabelle mit den Prozentangaben folgende Fragestellung:

Fragestellung:
Ist der Zusammenhang zwischen Rauchen und Atemnot signifikant?

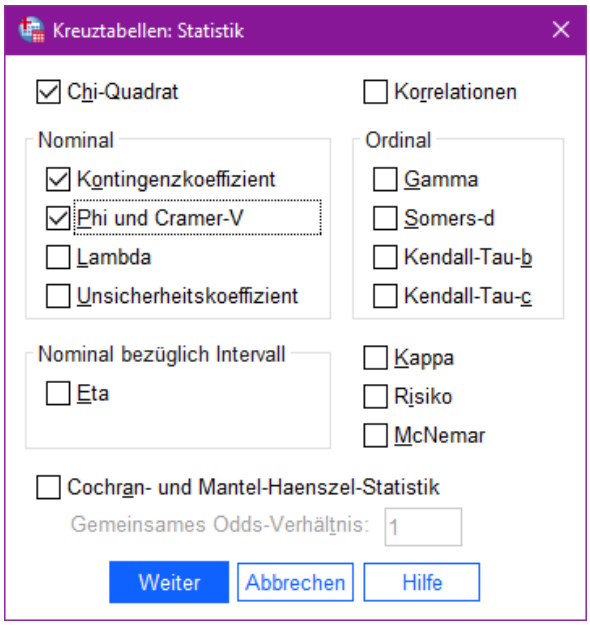
Dazu erstellen wir wieder dieselbe Kreuztabelle, diesmal mit den erwarteten Häufigkeiten und den Residuen.

Analysieren → Deskriptive Statistiken → Kreuztabellen… Zeilen: f20_a_kat3
Spalten: f18_d_kat3
Zellen→ beobachtete und erwartete Häufigkeiten → Weiter → OK

			f18_d_kat3 Atemnot in 3 Gruppen			
			1 tgl./alle paar Tage	2 alle paar Wo/Mo	3 nie	Gesamt
f20_a_kat3 Rauchgewohn-heiten in 3 Gruppen	1 nie	Anzahl	54	69	198	321
		Erwartete Anzahl	60,5	78,3	182,2	321,0
	2 1-20 Zig.	Anzahl	57	81	178	316
		Erwartete Anzahl	59,5	77,1	179,4	316,0
	3 21 Zig. ++	Anzahl	31	34	52	117
		Erwartete Anzahl	22,0	28,6	66,4	117,0
Gesamt		Anzahl	142	184	428	754
		Erwartete Anzahl	142,0	184,0	428,0	754,0

Erwartete Häufigkeit: Zelle 1: 54 Befragte: Wenn zwischen den Rauchgewohnheiten und Atemnot kein Zusammenhang besteht, dann werden in dieser Zelle 60 Personen erwartet, die nie rauchen und täglich/alle paar Tage unter Atemnot leiden. Somit befinden sich in dieser Zelle 6 Personen weniger als erwartet.

Analysieren → Deskriptive Statistiken → Kreuztabellen… Zeilen: f20_a_3kat
Spalten: f18_d_3kat
Zellen→ **Häufigkeiten** beobachtet und erwartet
Residuen: nicht standardisiert und standardisiert → Weiter
Statistiken → Chi-Quadrat, Kontingenzkoeffizient, Phi und Cramer's V → Weiter → OK



Standardisierte Residuen - wozu?
Wenn kein Zusammenhang besteht, dann sind die standardisierten Residuen standardnormal-verteilt. Das heißt, wenn sie nahe bei 0 sind, dann spricht das für den Zufall.
Wenn sie hingegen stark von 0 abweichen (also **kleiner als <-2 oder größer als > +2** sind), dann spricht dies für eine auffällige Abweichung.
Beachte: Ob die tatsächlichen Häufigkeiten von den erwarteten Häufigkeiten auffällig abweichen kann nur anhand der standardisierten Residuen festgestellt werden, nicht anhand der nicht standardisierten Residuen!

			f18_d_kat3 Atemnot in 3 Gruppen			
			1 tgl./alle paar Tage	2 alle paar Wo/Mo	3 nie	Gesamt
f20_a_kat3 Rauchgewohn-heiten in 3 Gruppen	1 nie	Anzahl	54	69	198	321
		Erwartete Anzahl	60,5	78,3	182,2	321,0
		Residuen	-6,5	-9,3	15,8	
		Stdd. Residuum	-,8	-1,1	1,2	
	2 1-20 Zig.	Anzahl	57	81	178	316
		Erwartete Anzahl	59,5	77,1	179,4	316,0
		Residuen	-2,5	3,9	-1,4	
		Stdd. Residuum	-,3	,4	-,1	
	3 21 Zig. ++	Anzahl	31	34	52	117
		Erwartete Anzahl	22,0	28,6	66,4	117,0
		Residuen	9,0	5,4	-14,4	
		Stdd. Residuum	1,9	1,0	-1,8	
Gesamt		Anzahl	142	184	428	754
		Erwartete Anzahl	142,0	184,0	428,0	754,0

Residuen: Zelle 1: In dieser Zelle befinden sich 6 Personen weniger als erwartet. Dieses Residuum von -6,5 Personen ergibt übertragen auf die Standardnormalverteilung ("standardisiert") -0,8.
Die deutlichsten Abweichungen zum Zufall können hier bei den starken Raucher*innen (über 21 Zigaretten) beobachtet werden: Es gibt 9 Personen mehr als erwartet, die stark rauchen und häufig unter Atemnot leiden (Stdd. Res. +1,9) und 14 Personen weniger als erwartet, die stark rauchen und nie unter Atemnot leiden (Stdd. Res. -1,8)

Beispiel anhand der Zelle Rauchen "nie" und tgl./alle paar Tage Atemnot: 54 Personen

Erwartete Anzahl = $f_e = \frac{Zeilenrandsumme * Spaltenrandsumme}{Gesamtsumme}$ → $f_e = \frac{321*142}{754} = 60{,}5$

Die ***erwarteten Häufigkeiten*** geben an, wie viele Fälle in der jeweiligen Zelle vorliegen (würden), wenn zwischen den beiden Merkmalen kein Zusammenhang besteht. Kein Zusammenhang wird auch als Unabhängigkeit oder Zufall bezeichnet. Jene Häufigkeit also, die wir erwarten würden, wenn es hier keine Gesetzmäßigkeit gibt und die Zellhäufigkeiten somit zufällig zustande gekommen sind. Hier sind 6 Personen weniger in der Zelle, als wir erwarten würden:

Residuum (Abweichung) = $f_o\,(tatsächliche\ Anzahl) - f_e\,(erwartete\ Anzahl)$ → $\mathrm{Re}\,s = 54 - 60{,}5 = -6{,}6$

Das ***Residuum*** ist die Differenz zwischen den beobachteten Häufigkeiten und den erwarteten Häufigkeiten, also wie weit die Zufallshäufigkeiten von den tatsächlichen Häufigkeiten entfernt sind. Diese Residuen kann man mit Hilfe der Rechenschablone der Standardnormalverteilung standardisieren:

Standardisiertes Residuum = $Stdd.\mathrm{Re}\,s = \frac{abs.\mathrm{Re}\,s}{\sqrt{erw.Anz.}}$ → $Stdd.\mathrm{Re}\,s = \frac{6{,}5}{\sqrt{60{,}5}} = 0{,}8$

Die Testtabelle zum Chi-Quadrat-Test zeigt, wie wahrscheinlich das Prüfmaß Chi-Quadrat unter der Voraussetzung der Zufallsverteilung ist.

Prüfmaß $\chi^2 = 11{,}3$

Chi-Quadrat-Tests

	Wert	df	Asymptotische Signifikanz (zweiseitig)
Chi-Quadrat nach Pearson	11,297[a]	4	,023
Likelihood-Quotient	11,111	4	,025
Zusammenhang linear-mit-linear	8,871	1	,003
Anzahl der gültigen Fälle	754		

a. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 22,03.

Signifikanz oder "Fehlerwahrscheinlichkeit" **= p * 100 = 2,3%**

Symmetrische Maße

		Wert	Näherungsweise Signifikanz
Nominal- bzgl. Nominalmaß	Phi	,122	,023
	Cramer-V	,087	,023
	Kontingenzkoeffizient	,121	,023
Anzahl der gültigen Fälle		754	

Assoziationsmaß CV = 0,087 wird nur interpretiert, wenn signifikant!

Interpretation: Das Prüfmaß Chi-Quadrat beträgt 11,3 und besitzt eine Wahrscheinlichkeit von 2,3% (p= 0,023) unter Voraussetzung der Unabhängigkeit. Der Zusammenhang zwischen Rauchen und Atemnot ist signifikant (p <= 0,05). Die Stärke des Zusammenhangs ist allerdings gering (Cramer's V = 0,087). Das Symptom Atemnot tritt also signifikant häufiger auf, wenn die Befragten mehr Zigaretten pro Tag rauchen, doch ist der hier beobachtete Gruppenunterschied gering.

Wir testen nun abschließend, ob dieser signifikante Zusammenhang unabhängig vom Geschlecht besteht, also für Frauen und Männer gleichermaßen gilt.

Fragestellung: ***Ist dieser signifikante Zusammenhang zwischen Rauchen und Atemnot unabhängig vom Geschlecht signifikant?***

Analysieren → Deskriptive Statistiken → Kreuztabellen… Zeilen: f20_a_3kat
Spalten: f18_d_3kat
Schicht: sex
Statistiken → Chi-Quadrat, Kontingenzkoeffizient, Phi und Cramer's V → Weiter → OK

Chi-Quadrat-Tests

sex Geschlecht		Wert	df	Asymptotische Signifikanz (zweiseitig)
1 männlich	Chi-Quadrat nach Pearson	19,707[b]	4	,001
	Anzahl der gültigen Fälle	427		
2 weiblich	Chi-Quadrat nach Pearson	4,148[c]	4	,386
	Anzahl der gültigen Fälle	317		
Gesamt	Chi-Quadrat nach Pearson	11,178[a]	4	,025
	Anzahl der gültigen Fälle	744		

Der Zusammenhang ist nur für die Männer signifikant.

Aber auch hier ist der Zusammenhang eher schwach,
Cramers V = 0,152.

a. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 21,79.

b. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 13,54.

c. 0 Zellen (0,0%) haben eine erwartete Häufigkeit kleiner 5. Die minimale erwartete Häufigkeit ist 6,91.

Beachte: Wenn in mehr als 20% der Zellen eine erwartete Häufigkeit kleiner als 5 vorliegt, dann sollte der Chi-Quadrat-Test nicht durchgeführt werden. Fallzahl in den Zellen ist dann zu gering. Wenn geht, in gröbere Kategorien zusammenfassen.

Symmetrische Maße

sex Geschlecht			Wert	Näherungsweise Signifikanz
1 männlich	Nominal- bzgl. Nominalmaß	Phi	,215	,001
		Cramer-V	,152	,001
		Kontingenzkoeffizient	,210	,001
	Anzahl der gültigen Fälle		427	
2 weiblich	Nominal- bzgl. Nominalmaß	Phi	,114	,386
		Cramer-V	,081	,386
		Kontingenzkoeffizient	,114	,386
	Anzahl der gültigen Fälle		317	
Gesamt	Nominal- bzgl. Nominalmaß	Phi	,123	,025
		Cramer-V	,087	,025
		Kontingenzkoeffizient	,122	,025
	Anzahl der gültigen Fälle		744	

Beachte: Das Assoziationsmaß "Phi" wird nur bei 2x2-Kreuztabellen verwendet (genannt: Vierfeldertafel) Hier liegt eine 3x3-Kreuztabelle vor.

Interpretation:
Die Fragestellung, ob die Rauchgewohnheiten mit dem Auftreten von Atemnot zusammenhängen, zeigt für Männer und Frauen unterschiedliche Ergebnisse.

Bei den **Männern** ist der Zusammenhang signifikant (χ^2=19,7, df=4, p = 0,001): Je mehr Zigaretten die männlichen Befragten rauchen, desto häufiger leiden sie unter Atemnot.

Bei den **Frauen** ist der Zusammenhang allerdings nicht signifikant (χ^2=4,1, df=4, p = 0,386): Zwar geben Frauen, die mehr als 21 Zigaretten pro Tag rauchen, an, häufiger unter Atemnot zu leiden als jene, die weniger rauchen, doch ist dieser Unterschied nicht signifikant, da die Anzahl der starken Raucher*innen sehr gering ist (30 Personen).

5.6 Hausübung 6: Chi-Quadrat-Test bei Kreuztabellen

Pflichtaufgabe 6: Testen Sie, ob die von Ihnen aufgestellte Fragestellung von Aufgabe 5 signifikant ist. Beschreiben und interpretieren Sie die erwarteten Häufigkeiten, die Residuen und das Test-ergebnis. Was bedeutet das Ergebnis inhaltlich? Haben Sie eine Erklärung dafür?

Fleißaufgabe für Lernwütige: Testen Sie weiters die Signifikanz nach der von Ihnen in Hausübung 5 eingeführten Schichtvariable!

Fleißaufgabe für Lernwütige: Erstellen Sie dieselbe Aufgabe mit Ihrem eigenen fiktiven Datensatz!

6 Signifikanztests

Exkurs: Anwendung von Signifikanztests

Signifikanztests dienen zur statistischen Überprüfung von Hypothesen. Zunächst wird davon ausgegangen, dass die Nullhypothese (H_0) in der Grundgesamtheit (Population) gilt. Unter dieser Annahme lässt sich für die Population eine Stichprobenkennwerteverteilung konstruieren, die angibt, mit welcher Wahrscheinlichkeit mögliche Stichprobenergebnisse auftreten können. Mit dieser Stichprobenkennwerteverteilung wird nun das konkret in der Untersuchung ermittelte Stichprobenresultat verglichen. Ist das gefundene Stichprobenergebnis ein wahrscheinliches Ergebnis, so steht es in Einklang mit der H_0. Ist das Stichprobenergebnis ein unwahrscheinliches Ergebnis, das unter Gültigkeit der H_0 nur extrem selten auftreten kann, wird die Nullhypothese als unplausibel verworfen. Ein solches, im Sinne der H_0unplausibles Ergebnis wird als "signifikantes Ergebnis" bezeichnet (H_0 wird abgelehnt und H_1 wird angenommen).
Je nach Signifikanztest sind Voraussetzungen erforderlich sein (z.B. metrisches Skalenniveau und Normalverteilung der Testvariablen).

Vorgehen:

1. Formulierung der Nullhypothese (H_0) und der Alternativhypothese (H_1).
2. Ermittlung einer statistischen Prüfgröße.
3. Festlegung des Signifikanzniveaus (üblicherweise 5%-Niveau).
4. Annahme der H_1, wenn Irrtumswahrscheinlichkeit p kleiner <0,05, ansonsten wird H_0 beibehalten.

Fragestellung:

Zweiseitige Fragestellung: Wenn über die Richtung des vermuteten Zusammenhangs keine sichere Annahme getroffen werden kann.
Z.B. $\bar{x}_1 \neq \bar{x}_2$: *"Die durchschnittliche Tagestrinkmenge unterscheidet sich bei Männern und Frauen."*

Einseitige Fragestellung: Wenn die Richtung des vermuteten Zusammenhangs angegeben werden kann. In diesem Fall darf die Wahrscheinlichkeit bei der Signifikanz (p) halbiert werden.
Z.B. $\bar{x}_1 < \bar{x}_2$: *"Frauen trinken durchschnittlich pro Tag mehr als Männer."*

Unterschied zwischen Prüfgröße und Signifikanz:

Grundsätzlich wird bei jedem statistischen Test zwischen der **Prüfgröße** (z.B. der Chi-Quadrat-Wert) und der **Signifikanz der Prüfgröße** unterschieden. Während die Prüfgröße Chi-Quadrat theoretisch Werte bis unendlich annehmen kann, liegt die **Signifikanz (=Wahrscheinlichkeit der Prüfgröße bei angenommener Unabhängigkeit) immer zwischen 0 und 1**.

Signifikanz:

- ein Wert **nahe bei 0** bedeutet:
  der berechnete Wert der Prüfgröße ist bei angenommener Unabhängigkeit sehr **un**wahrscheinlich → ist dieser Wert **gleich oderkleiner als das gewählte Signifikanzniveau** (üblicherweise 0,05 oder 0,01), dann wird konventionell die H_0 verworfen und die H_1 (Annahme von Abhängigkeit oder **Zusammenhang**) angenommen;
- ein Wert **nahe bei 1** bedeutet:
  der berechnete Wert der Prüfgröße ist bei angenommener Unabhängigkeit sehr wahrscheinlich → ist dieser Wert **größer als das gewählte Signifikanzniveau** (üblicherweise 0,05 oder 0,01), dann wird konventionell die H_0 (Annahme von Unabhängigkeit oder **keinem Zusammenhang**) beibehalten;

Test	Chi-Quadrat-Test	t - Test		ANOVA (Varianzanalyse)	U - Test	Wilcoxon-Test
Verfahren	Vergleich von tatsächlichen und erwarteten Häufigkeiten	Vergleich von zwei Mittelwerten		Vergleich mehrerer Mittelwerte	Vergleich von zwei Verteilungen (anhand der Rangplätze)	
Anwendung bei	Kreuztabellen	unabhängige Gruppen/Stpr.	abhängige/ gepaarte Stpr.	mehrere Gruppen/Stpr.	unabhängige Gruppen/Stpr.	abhängige/ gepaarte Stpr.
Prüfgröße	χ^2	t bei homogenen/gleichen Varianzen bei ungleichen Varianzen (Prüfgröße F)	t	F	z	z
Vorrausetzung an die abhängige Variable	nominal, ordinal	metrisch, annähernd normalverteilt (n ≥ 30)			ordinal bzw. metrisch und nicht normal-verteilt (bzw. zu geringe Fallzahl)	
Vorrausetzung an die unabhängige Variable		zwei Ausprägungen	zwei Messwerte derselben Person (vorher/nachher, Befr./PartnerIn)	nominal, ordinal	zwei Ausprägungen	zwei Messwerte der-selben Person (vorher/nachher) (Befr./PartnerIn)
Nullhypothese	Es besteht kein Zusammenhang zwi-schen den beiden Variablen in der Grundgesamtheit	Die Mittelwerte in den zwei Gruppen sind gleich, bzw. Differenz = 0	Die mittlere Differenz der Messwerte der Ver-gleichspaare (= durch-schnittliche Abweichung zwischen vorher/nachher oder Befr./Partn.) ist 0	Die Varianz zwi-schen den Gruppen ist gleich bzw. kleiner als die Vari-anz innerhalb der Subgruppen	Die Variable hat in beiden Grup-pen in der GG die gleiche (Rang-) Verteilung.	Die Variable hat vor-her/nachher bzw. bei Befr. und deren PartnerInnen die gleiche (Rang-) Verteilung
Typische Fragestellung	Unterscheidet sich die Parteipräferenz nach Geschlecht	Unterscheidet sich das Durchschnitts-einkommen von Männern und Frauen	Unterscheidet sich der durchschnittliche Blutzu-ckerspiegel vor und nach einem Medikament.	Unterscheidet sich das Durch-schnittseinkommen nach Bildung	Unterscheiden sich die Noten in Mathe bei Burschen und Mädchen	Unterscheiden sich die Noten der Befragten in Mathe und Deutsch

6.1 t-Test für unabhängige Stichproben

Beim t-Test wird untersucht, ob sich die Mittelwerte einer Variable zwischen zwei Gruppen (Gruppe A und Gruppe B) signifikant unterscheiden. Für diesen Test brauchen wir:

Voraussetzungen für den t-Test für unabhängige Stichproben

1.) Eine **abhängige, metrische Testvariable** und
2.) Eine **dichotome Gruppenvariable** (Gruppe A und Gruppe B werden verglichen)
3.) Annähernde Normalverteilung in beiden Gruppen der Testvariable:
Grafische Analyse mittels Histogramms und Normalverteilungskurve:
Kriterien: Gipfel ist annähernd in der Mitte, Schiefe ist zwischen 0 und ±1 (nahe bei 0).
4.) Der **Levene-Test auf Varianzgleichheit** testet, ob die Varianzen in Gruppe A und Gruppe B gleich ("homogen") oder ungleich ("heterogen") sind.
Die **Nullhypothese**: $H_0: s_A^2 = s_B^2$ — Die Varianzen sind homogen/gleich
Die **Alternativhypothese**: $H_A: s_A^2 \neq s_B^2$ — Die Varianzen sind heterogen/ungleich

Bei einer Signifikanz von….
p > 0,05 sind die **Varianzen gleich** und es wird die Teststatistik in der **1. Zeile** abgelesen.
p <= 0,05 sind die **Varianzen ungleich** und es wird die Teststatistik in der **2. Zeile** abgelesen.

Berechnung des t-Test für unabhängige Stichproben

Die **Nullhypothese**: $H_0: \bar{x}_A = \bar{x}_B$ — Mittelwert von Gruppe A entspricht dem Mittelwert von Gruppe B

Die **Alternativhypothese**: $H_A: \bar{x}_A \neq \bar{x}_B$ — Mittelwert von Gruppe A unterscheidet sich signifikant von Mittelwert von Gruppe B

Die **Rechenlogik**: $t = \frac{\bar{x}_A - \bar{x}_B}{s_{\bar{x}_A - \bar{x}_B}}$ — Getestet wird der **Unterschied** (die Differenz) **von zwei Mittelwerten** (Gruppe A und Gruppe B).

Dieser Unterschied wird ins Verhältnis gesetzt zum geschätzten **Standardfehler der Differenz der beiden Mittelwerte**.

Es gibt zwei Formeln:

$$t = \frac{\bar{x}_A - \bar{x}_B}{\frac{s_A^2(n-1)+s_B^2(m-1)}{(n-1)+(m-1)} * \sqrt{\frac{1}{n}+\frac{1}{m}}}$$ — t wenn Varianzen homogen

$$t = \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\frac{s_A^2}{n}+\frac{s_B^2}{m}}}$$ — t wenn Varianzen heterogen

Vor der Durchführung des t-Tests ist eine **grafische Analyse der Testvariable** (Häufigkeitstabelle mit Histogramm mit Normalverteilungskurve) notwendig, um die Normalverteilung zu beurteilen!
Wenn die Normalverteilung nicht gegeben ist, dann einen U-Test durchführen!

Fragestellung: *Unterscheidet sich das durchschnittliche körperliche Wohlbefinden von Befragten, die unter 2 bzw. über 2 Jahre arbeitslos sind?*

Überprüfung der Normalverteilung

Wir teilen den Datensatz nach der Gruppenvariable auf (**f14_di**) und erstellen ein Histogramm mit Normalverteilungskurve und lassen uns Mittelwert, Median, Schiefe und Kurtosis angeben.

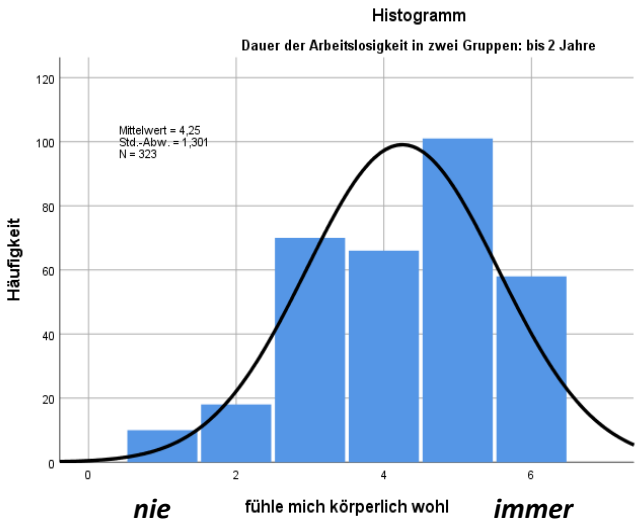
Daten → Datei aufteilen → Ausgabe nach Gruppen aufteilen: **f14_di** → OK
Analysieren → deskriptive Statistiken → Häufigkeiten → **f22_a**
→ optional: *Häufigkeitstabelle ausklicken*
→ Diagramme → Histogramm mit Normalverteilungskurve
→ Statistik: Kennzahlen auswählen *[Median, Modalwert, Mittelwert, Schiefe, Kurtosis]* → OK

Statistiken[a]

f22_a fühle mich körperlich wohl

N	Gültig	323
	Fehlend	9
Mittelwert		4,25
Median		4,00
Modus		5
Schiefe		-,474
Standardfehler der Schiefe		,136
Kurtosis		-,457
Standardfehler der Kurtosis		,271

a. f14_di Dauer AL = 1 unter 2 Jahre



Schiefe zwischen 0 und ±1?

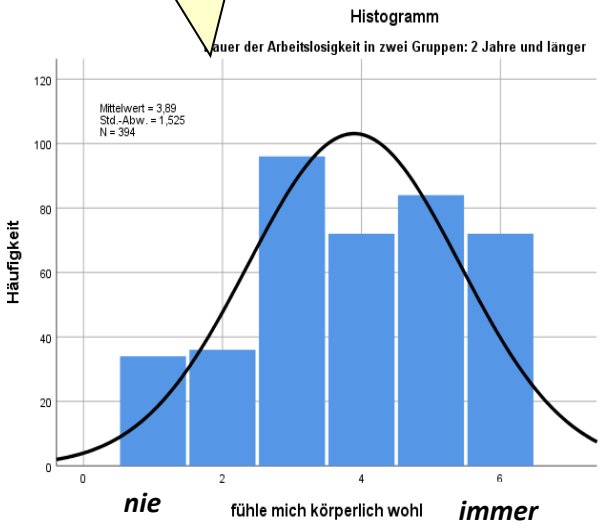
Gipfel annähernd in der Mitte?

Statistiken[a]

f22_a fühle mich körperlich wohl

N	Gültig	394
	Fehlend	21
Mittelwert		3,89
Median		4,00
Modus		3
Schiefe		-,252
Standardfehler der Schiefe		,123
Kurtosis		-,895
Standardfehler der Kurtosis		,245

a. f14_di Dauer AL = 2 2 Jahre und länger

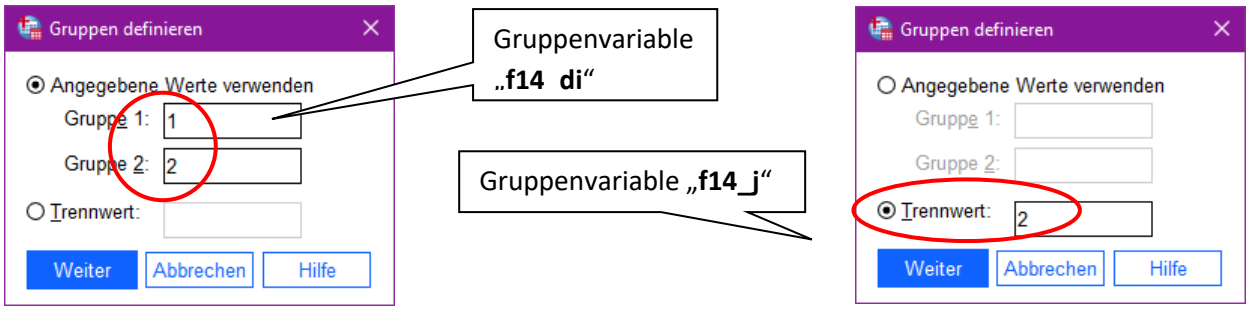


Interpretation: In beiden Vergleichsgruppen befindet sich der Gipfel der Verteilung annähernd in der Mitte der Verteilung, Mittelwert und Median sind annähernd gleich und Schiefe (und Kurtosis) sind zwischen 0 und ±1 (nahe bei 0). Daher nehmen wir für beide Gruppen eine Normalverteilung an und führen den t-Test durch.

Daten → Datei aufteilen → Alle Fälle analysieren, keine Gruppen bilden → Ok

Durchführung des t-Tests

Analysieren → Mittelwerte vergleichen→ t-Test bei unabhängigen Stichproben
→ **Testvariable(n):** z.B. f22_a (deren Mittelwert verglichen wird);
→ **Gruppenvariable**: f14_di (diese besteht nur aus 2 Ausprägungen, daher:
→ **Gruppen definieren** anklicken und bei Gruppe 1: „1" eintragen, Gruppe 2: „2" eintragen
Es können hier auch aus einer mehrkategorialen Variable zwei einzelne Gruppen zum Vergleich ausgewählt werden (z.B. Bildung, Vergleich der Gruppen mit der Ausprägung 1 und 4)
Oder: es kann auch ein **Trennwert** eingetragen werden, um zwei Gruppen zu bilden
– bei einer metrischen Variable, z.B. f14_j „Dauer AL in Jahren", Trennwert 2
((→ optional: ☑ Effektgrößen anklicken – liefert Cohen's d unter „Punktschätzung"))
→„Ok" oder „Einfügen" um den Befehl zunächst in eine Syntax-Datei (.sps) zu schreiben



Gruppenstatistiken

	f14_di Dauer AL	N	Mittelwert	Standard-abweichung	Standardfehler des Mittelwertes
f22_a fühle mich körperlich wohl	1 unter 2 Jahre	323	4,25	1,301	,072
	2 2 Jahre und länger	394	3,89	1,525	,077

Test bei unabhängigen Stichproben

		Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit			
		F	Signifikanz	T	df	Sig. (2-seitig)	Mittlere Differenz
f22_a fühle mich körperlich wohl	Varianzen sind gleich	10,145	,002	3,334	715	,001	,357
	Varianzen sind nicht gleich			3,386	713,86	,001	,357

Interpretation des t-Test

1. Schritt: F- Test (Levene-Test) auf Varianzgleichheit (rosa markiert) lesen:
→ wenn der F-Test nicht signifikant ist (>0,05), dann sind die „Varianzen gleich"
→ t- Testergebnis wird in der 1. Zeile abgelesen
→ wenn dieser signifikant ist (≤ 0,05), dann sind die „Varianzen nicht gleich"
→ t-Testergebnis wird in der 2. Zeile abgelesen

2. Schritt: Signifikanz des t-Tests überprüfen (1. oder 2. Zeile) (grün markiert)
→ signifikant (≤ 0,05),: die beiden Gruppen unterscheiden sich signifikant (H1 gilt)
→ nicht signifikant (>0,05),: die beiden Gruppen unterscheiden sich nicht signifikant (H0 gilt)

Interpretation:
Befragte mit einer kürzeren Dauer der Arbeitslosigkeit (unter 2 Jahre) weisen durchschnittlich ein höheres körperliches Wohlbefinden auf (4,25 Punkte versus 3,89 Punkte - beachte: je höher der Wert, desto höher das körperliche Wohlbefinden). Die Varianzen sind laut Levene-Test heterogen (nicht gleich), da der Levene-Test signifikant ist (p=0,002). Die Teststatistik für den t-Test wird daher in der zweiten Zeile abgelesen.
Es zeigt sich, dass dieser Mittelwertunterschied zwischen den beiden Gruppen signifikant ist (p=0,001). Das durchschnittliche körperliche Wohlbefinden unterscheidet sich somit signifikant nach der Dauer der Arbeitslosigkeit: Befragte, die unter 2 Jahre arbeitslos sind, fühlen sich durchschnittlich körperlich wohler als Befragte, die bereits länger arbeitslos sind.

Exkurs: Mit der Funktion "Datei aufteilen" können wir überprüfen, ob dieses signifikante Ergebnis für Frauen und Männer gleichermaßen gilt.

Daten → Datei aufteilen → Gruppen vergleichen: sex → OK

Wir teilen die Datei auf und führen den gleichen t-Test nochmals durch:

Gruppenstatistiken

sex Geschlecht		f14_di Dauer AL	N	Mittelwert	Standard-abweichung	Standardfehler des Mittelwertes
1 männlich	f22_a fühle mich körperlich wohl	1 unter 2 Jahre	167	4,34	1,321	,102
		2 2 Jahre und länger	230	4,06	1,482	,098
2 weiblich	f22_a fühle mich körperlich wohl	1 unter 2 Jahre	151	4,15	1,272	,104
		2 2 Jahre und länger	160	3,68	1,544	,122

Test bei unabhängigen Stichproben

			Levene-Test der Varianzgleichheit		T-Test für die Mittelwertgleichheit		
sex Geschlecht			F	Signi-fikanz	T	df	Sig. (2-seitig)
1 männlich	f22_a fühle mich körperlich wohl	Varianzen sind gleich	2,804	,095	1,95	395	,052
		Varianzen sind nicht gleich			1,98	378,8	,048
2 weiblich	f22_a fühle mich körperlich wohl	Varianzen sind gleich	7,561	,006	2,92	309	,004
		Varianzen sind nicht gleich			2,94	303,5	,004

Interpretation:
Sowohl bei den Männern (4,06 versus 4,34) als auch bei den Frauen (3,68 versus 4,15) weisen Befragte mit einer längeren Dauer der Arbeitslosigkeit (mehr als 2 Jahre) durchschnittlich ein schlechteres körperliches Wohlbefinden auf (erkennbar an den niedrigeren Mittelwerten auf der Punkteskala).

Die Varianzen der beiden Untergruppen der Männer sind laut Levene-Test homogen (gleich, p=0,095), die Teststatistik für den t-Test wird daher in der ersten Zeile abgelesen. Es zeigt sich ein knapp nicht signifikanter Unterschied im durchschnittlichen körperlichen Wohlbefinden nach der Dauer der Arbeitslosigkeit (p=0,052): Der Unterschied im körperlichen Wohlbefinden ist nach der Dauer der Arbeitslosigkeit für die Männer knapp nicht signifikant.

Die Varianzen der beiden Untergruppen der Frauen sind laut Levene-Test heterogen (nicht gleich, p=0,006)), die Teststatistik für den t-Test wird daher in der zweiten Zeile abgelesen. Es zeigt sich ein hoch signifikanter Unterschied im durchschnittlichen körperlichen Wohlbefinden nach der Dauer der Arbeitslosigkeit (p=0,004): Befragte Frauen, die bereits länger arbeitslos sind fühlen sich körperlich signifikant weniger wohl als Befragte, die erst unter zwei Jahre arbeitslos sind.

6.2 Hausübung 7: t-Test für unabhängige Stichproben

Pflichtaufgabe 7:
Wählen Sie einen der Datensätze für diese Aufgabe aus (Arbeitslose oder Darmmanagement).
Formulieren Sie eine relevante Fragestellung, die mit einem t-Test überprüfbar ist. Welches Ergebnis vermuten Sie? Begründen Sie Ihre Vermutung.
Im Vorfeld: Gegebenenfalls rekodieren Sie die Gruppenvariable Ihrer Wahl hierfür auf 2 Kategorien.
Untersuchen Sie vorher, ob Ihre Testvariable in beiden Untergruppen annähernd normalverteilt ist.
Führen Sie nun den **t-Test** durch. Beschreiben und interpretieren Sie die Ergebnisse.
Hat sich Ihre Vermutung bestätigt? Wenn nicht, welche Erklärung haben Sie dafür?

Fleißaufgabe für Lernwütige: Welche dichotome Gruppenvariable könnte hier interferieren? Begründen Sie, warum Sie das glauben. Fügen Sie diese "Schichtvariable" zu Ihrer Fragestellung ein. Gegebenenfalls rekodieren Sie die Variable Ihrer Wahl hierfür auf 2 Kategorien.
Führen Sie den t-Test nochmals durch. Beschreiben und interpretieren Sie die Ergebnisse.
Datei aufteilen wieder ausschalten nicht vergessen!

Fleißaufgabe für Lernwütige: Erstellen Sie dieselbe Aufgabe mit Ihrem eigenen fiktiven Datensatz!

6.3 U-Test (unabhängige Stichproben)

Beim U-Test wird untersucht, ob sich die mittleren Ränge einer Variable von zwei Gruppen (Gruppe A und Gruppe B) signifikant unterscheiden. Der U-Test ist also das "ordinale" Äquivalent für den t-Test. Immer dann, wenn der t-Test nicht anwendbar ist (entweder weil ordinales Datenniveau vorliegt oder die metrische Variable nicht annähernd normalverteilt ist), ist der U-Test passend.

Voraussetzungen für den U-Test für unabhängige Stichproben

1.) Eine **abhängige, ordinale ODER metrische, nicht annähernd normalverteilte Testvariable** und
2.) Eine **dichotome Gruppenvariable** (Gruppe A und Gruppe B werden verglichen)

Berechnung des U-Test für unabhängige Stichproben

Die **Nullhypothese**: $H_0: \bar{R}_A = \bar{R}_B$

Der mittlere Rang von Gruppe A ist gleich dem mittleren Rang von Gruppe B

Die **Alternativhypothese**: $H_A: \bar{R}_A \neq \bar{R}_B$

Der mittlere Rang von Gruppe A ist nicht gleich dem mittleren Rang von Gruppe B

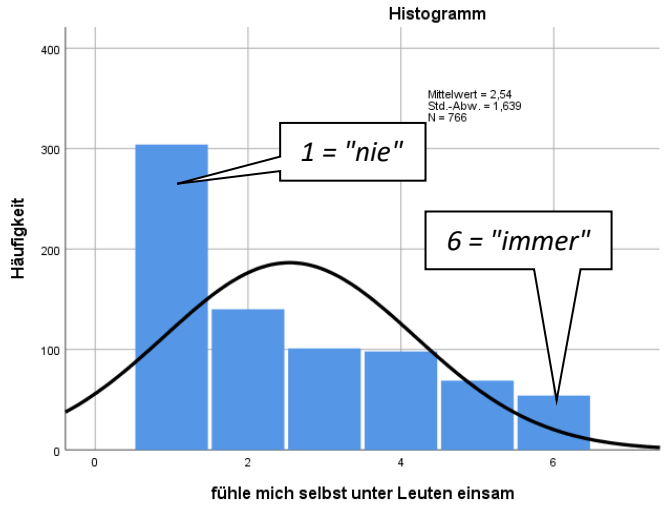
Die **Rechenlogik**:

$$z = \frac{\bar{R}_A - \bar{R}_B}{SE_{\bar{R}}}$$

Getestet wird der **Unterschied** (die Differenz) **von mittleren Rängen**.

Dieser Unterschied wird ins Verhältnis gesetzt zum **Standardfehler der Rangplatzdifferenz**.

Fragestellung: ***Unterscheidet sich die Verteilung der Variable "Einsamkeit unter Leuten", zwischen Personen, die unter 2 bzw. über 2 Jahre arbeitslos sind?***



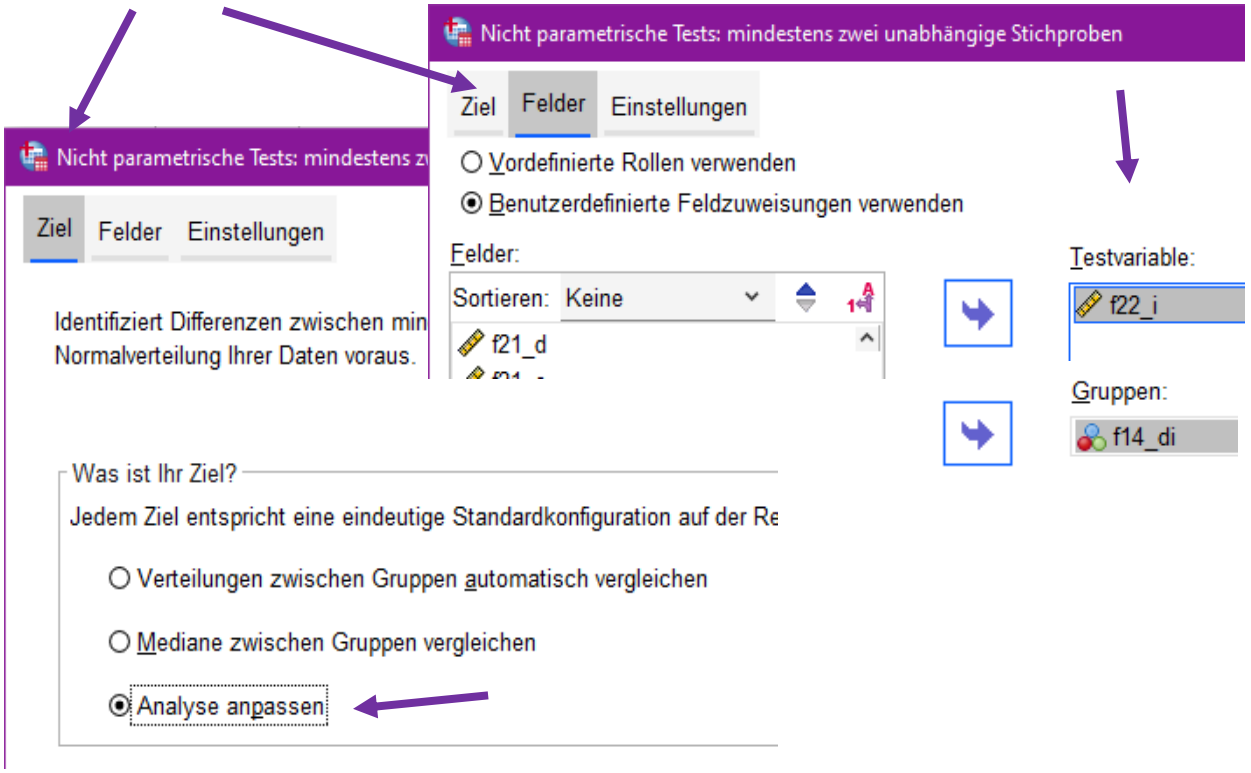
Testvariable **f22_i**
„Fühle mich selbst unter Leuten einsam“

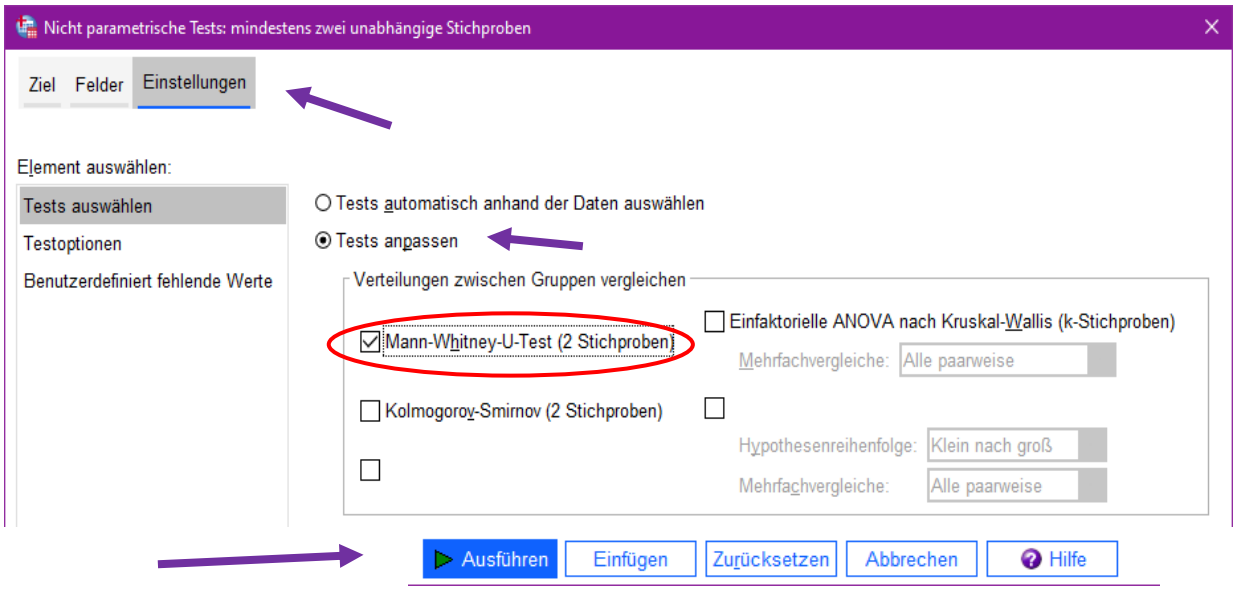
Der Gipfel ist eindeutig nicht in der Mitte der Verteilung.

Diese Variable ist nicht normalverteilt, daher führen wir den U-Test durch.

Durchführung des U-Test

Analysieren → Nichtparametrische Tests → unabhängige Stichproben
Reiter "**Ziel**": ⊙ Analyse anpassen
Reiter „**Felder**“: abhängige Variable in *Testfelder* **f22_i**
Gruppen: Gruppenvariable auswählen **f14_di**
Reiter "**Einstellungen**" ⊙ Tests anpassen → ☑ Mann-Whitney-U-Test (2 Stichproben →Ausführen
Beachte: Hier ist ein Trennwert nicht möglich, daher Variablen vorher in **zwei Gruppen rekodieren**!





Hypothesentestübersicht

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von f22_i fühle mich selbst unter Leuten einsam ist über die Kategorien von f14_di Dauer der Arbeitslosigkeit in zwei Gruppen identisch.	Mann-Whitney-U-Test bei unabhängigen Stichproben	,546	Nullhypothese beibehalten

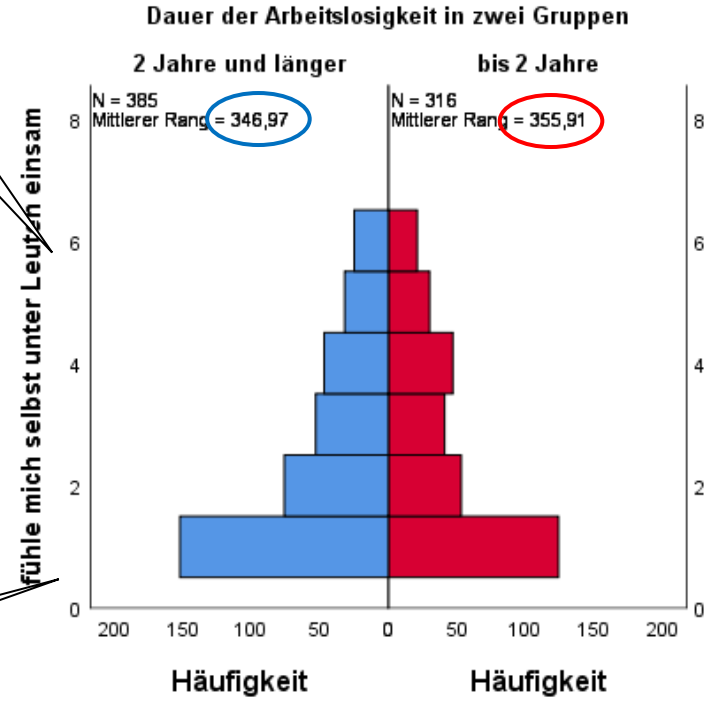
Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist ,050.

6 = "immer"

Zusammenfassung des Mann-Whitney-U-Tests bei unabhängigen Stichproben

Gesamtzahl	701
Mann-Whitney-U	59277,500
Wilcoxon-W	133582,500
Teststatistik	59277,500
Standardfehler	2568,487
Standardisierte Teststatistik	-,604
Asymptotische Sig. (zweiseitiger Test)	,546

1 = "nie"



Interpretation:

Befragte mit einer längeren Dauer der Arbeitslosigkeit (mehr als 2 Jahre) fühlen sich laut Testergebnis nicht einsamer als Befragte mit einer kürzeren Arbeitslosigkeitsdauer (z=-0,604; p=0,546).
Dies veranschaulicht auch die übereinstimmende Verteilung im Balkendiagramm.
Auch die mittleren Ränge liegen relativ nahe: Der mittlere Rang der Befragten mit längerer AL-Dauer beträgt 347 und jener der Befragten mit kürzerer AL-Dauer beträgt 356.

Fragestellung: *Gilt das Ergebnis, dass sich kein Unterschied in der Einsamkeit nach der Dauer der Arbeitslosigkeit zeigt, für Männer und Frauen?*

Wir teilen die Datei nach Geschlecht auf und führen den gleichen U-Test nochmals durch:

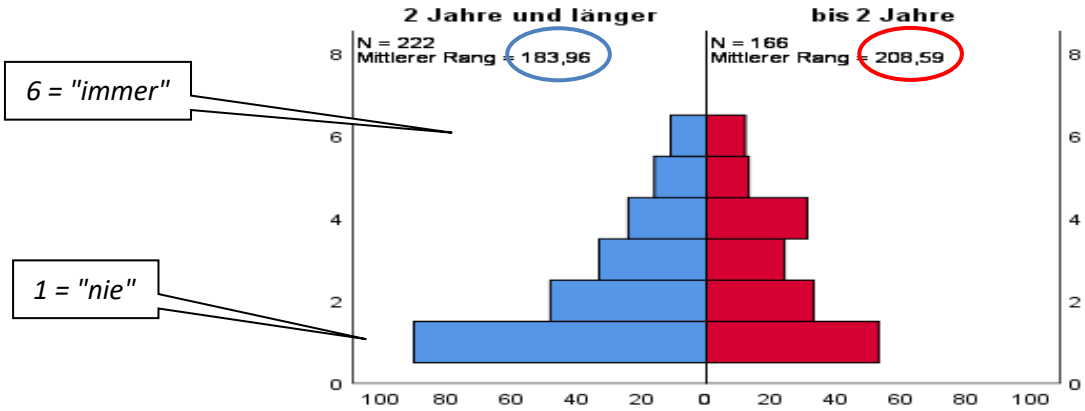
Daten → Datei aufteilen → ⊙ Ausgabe nach Gruppen aufteilen: sex → OK

sex Geschlecht = 1 männlich:

Hypothesentestübersicht

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von f22_i fühle mich selbst unter Leuten einsam ist über die Kategorien von f14_di Dauer der Arbeitslosigkeit in zwei Gruppen identisch.	Mann-Whitney-U-Test bei unabhängigen Stichproben	,027	Nullhypothese ablehnen

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist ,050.

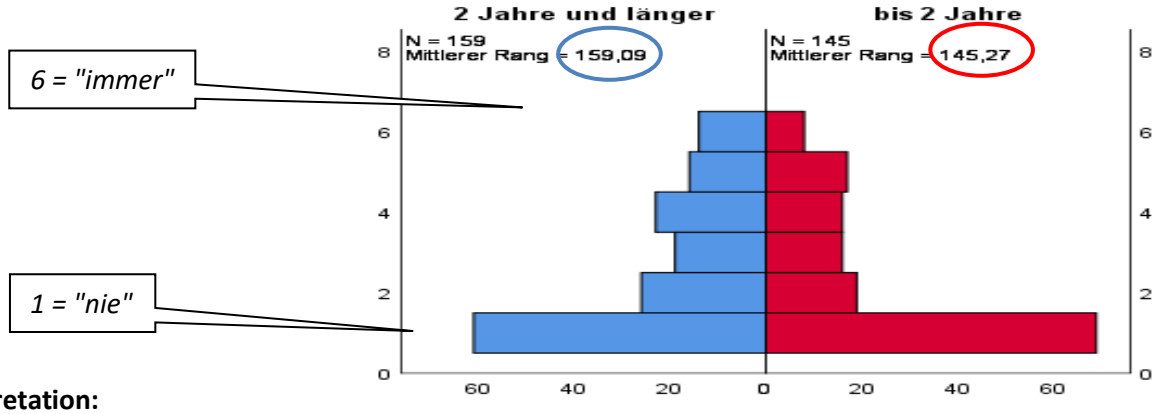


sex Geschlecht = 2weiblich

Hypothesentestübersicht

	Nullhypothese	Test	Sig.	Entscheidung
1	Die Verteilung von f22_i fühle mich selbst unter Leuten einsam ist über die Kategorien von f14_di Dauer der Arbeitslosigkeit in zwei Gruppen identisch.	Mann-Whitney-U-Test bei unabhängigen Stichproben	,152	Nullhypothese beibehalten

Asymptotische Signifikanzen werden angezeigt. Das Signifikanzniveau ist ,050.



Interpretation:

Bei den Männern zeigt sich, dass Befragte, die bereits länger als 2 Jahre arbeitslos sind, signifikant weniger einsam sind als jene mit kürzerer Arbeitslosigkeitsdauer (p= 0,027). Auch die mittleren Ränge liegen mit 209 (länger arbeitslos) versus 184 (kürzer arbeitslos) relativ deutlich auseinander.

Bei den Frauen zeigt sich ein ähnliches Ergebnis wie insgesamt. Frauen, die unter 2 Jahre arbeitslos sind fühlen sich zwar häufiger weniger einsam, und die länger arbeitslosen Frauen haben einen höheren mittleren Rang als die kürzer Arbeitslosen (159 versus 145), doch ist dieser Unterschied bei gegebener Fallzahl nicht signifikant (p= 0,152).

6.4 Hausübung 8: U-Test (unabhängige Stichproben)

Pflichtaufgabe 8:
Wählen Sie einen der Datensätze für diese Aufgabe aus (Arbeitslose oder Darmmanagement).
Formulieren Sie eine relevante Fragestellung, die mit einem U-Test überprüfbar ist.
Welches Ergebnis vermuten Sie? Begründen Sie Ihre Vermutung.
Im Vorfeld: Gegebenenfalls rekodieren Sie die Gruppenvariable Ihrer Wahl hierfür auf 2 Kategorien.
Führen Sie nun den **U-Test** durch. Beschreiben und interpretieren Sie die Ergebnisse.
Hat sich Ihre Vermutung bestätigt? Wenn nicht, welche Erklärung haben Sie dafür?

Welche weitere Gruppenvariable könnte im Zusammenhang mit der von Ihnen gewählten Testvariable stehen? Dichotomisieren Sie die weitere Gruppenvariable, und führen Sie einen weiteren U-Test durch. Interpretieren Sie das Ergebnis!

Fleißaufgabe für Lernwütige: Welche dichotome Gruppenvariable könnte hier interferieren? Begründen Sie, warum Sie das glauben. Fügen Sie diese "Schichtvariable" zu Ihrer Fragestellung ein. Gegebenenfalls rekodieren Sie die Variable Ihrer Wahl hierfür auf 2 Kategorien.
Führen Sie den U-Test nochmals durch. Beschreiben und interpretieren Sie die Ergebnisse.
Datei aufteilen wieder ausschalten nicht vergessen!

Fleißaufgabe für Lernwütige: Erstellen Sie dieselbe Aufgabe mit Ihrem eigenen fiktiven Datensatz!

7 Mehrfachantworten

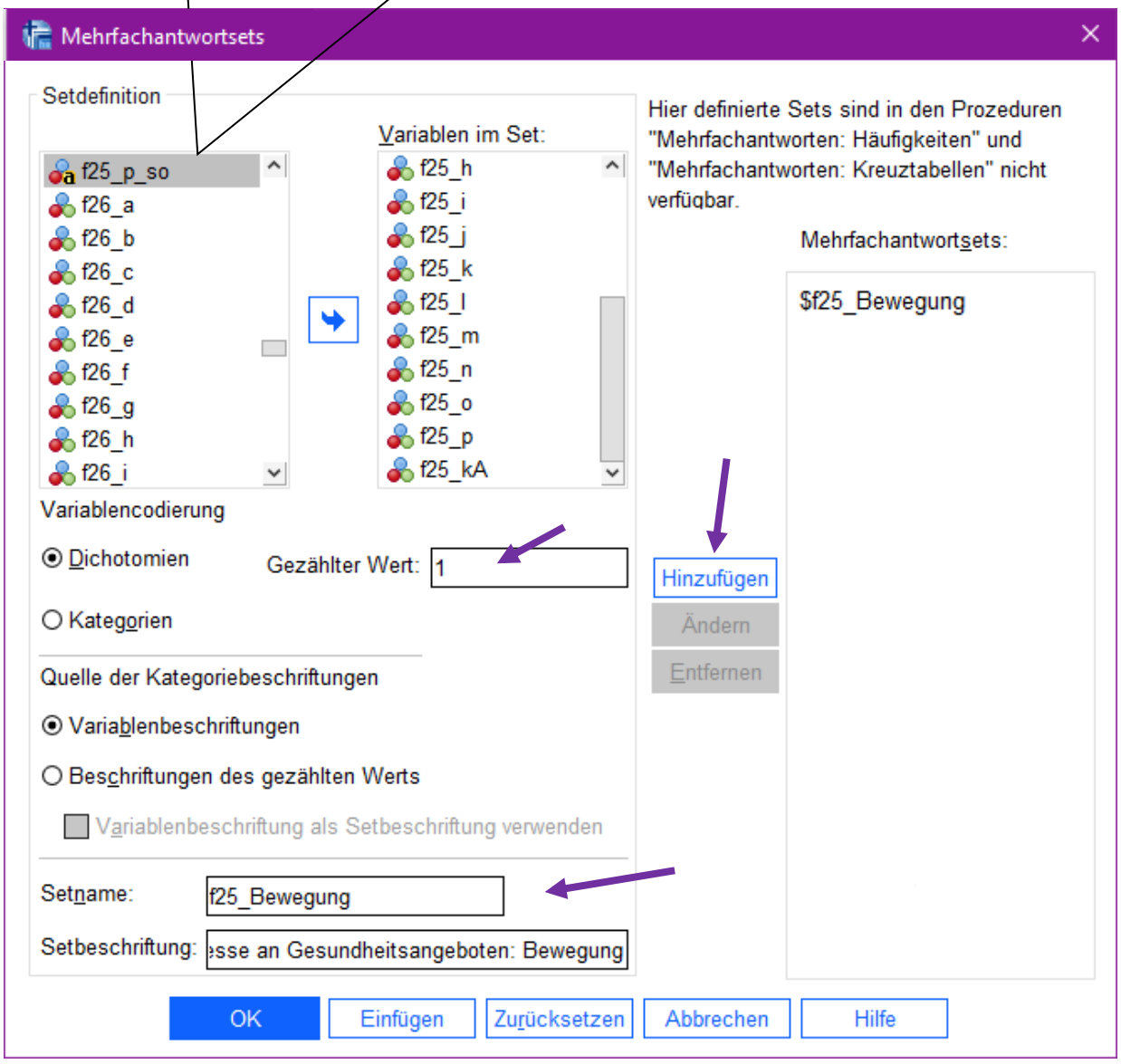
Sehr häufig finden sich in einem Fragebogen Mehrfachantworten. Eine Mehrfachantwort ist eine **Fragebatterie mit dem minimalsten Informationsgehalt**, nämlich „*genannt*“ und nicht „*genannt*“. Dies Erhebungsmöglichkeit ist sehr beliebt und geeignet, wenn keine feinere Abstufung (etwa ordinal) notwendig ist (z.B. bei der Frage: „*Wie haben Sie von unserer Beratungsstelle erfahren?“ – Internet / Zeitung / Fernsehen / Bekanntenkreis / Broschüre / Flyer / anderes, und zwar….*) oder nicht möglich bzw. schwierig ist (z.B. bei der Frage nach Liveevents).
Bei der Verwendung von Mehrfachantworten ist zu beachten, dass hier lediglich eine grobe Informationsqualität vorliegt, diese kann jedoch zu einem „Zähl-Index“ zusammengefasst werden, wobei die Anzahl der Nennungen gezählt wird, beispielsweise der Index „Anzahl der genannten Gründe, Alkohol zu trinken“.
Mehrfachantworten müssen für die Befragten durch einen Hinweis explizit gekennzeichnet werden, werden meistens mit **0 = *„nicht genannt“* / 1 = *„genannt“*** codiert und müssen mit den Items „anderes“ und/oder „keines“ abgeschlossen werden. Zum Item „anders“ wird meist noch ein Textfeld hintangestellt.

7.1 Definieren eins Mehrfachantwortsets

Wir untersuchen die Frage 25: "*Welche der folgenden Gesundheitsangebote sind für Sie besonders interessant? - Bewegungsangebote*". Sehen sich zuerst an, wie diese Frage im Fragebogen (Seite 9) dargestellt ist. Im Datenfile definieren wir, welche Variablen in dieses Mehrfachantwort-Set gehören.

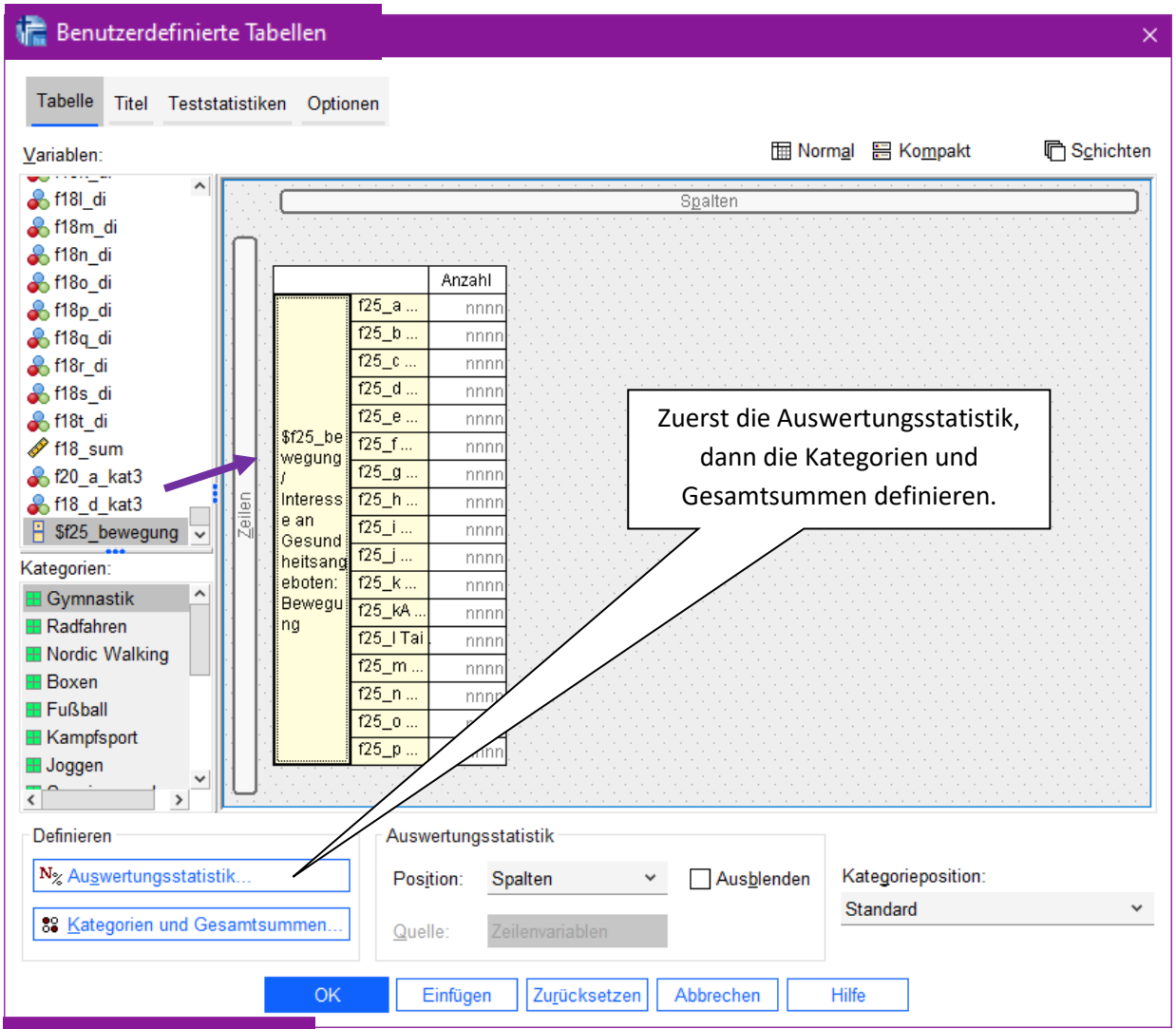
Analysieren → Tabellen → **Mehrfachantwortsets…..**
Alle Variablen (f25_a, f25_b, f25_c …… bis f25_kA) in ***Variablen im Set*** geben
Variablenkodierung: ⊙ Dichotomien → Gezählter Wert: 1
Quelle der Kategorienbeschriftungen: Variablenlabels→ Set-Name: **f25_bewegung**
Set-Label: *"Interesse an Gesundheitsangeboten Thema Bewegung"*
→ Hinzufügen → OK

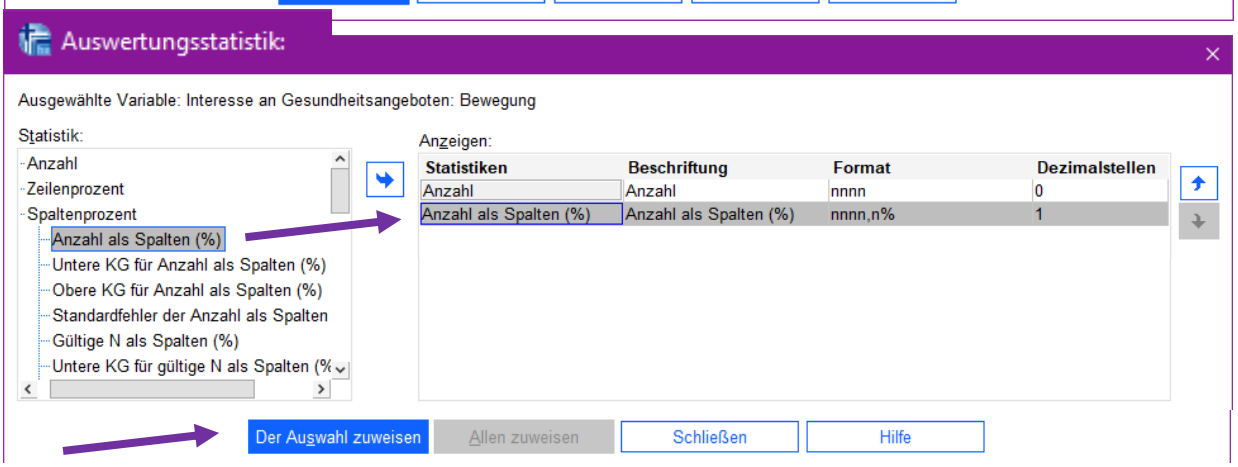
Beachte: Die Variable f25_p_so in der Variablenliste ist eine Textvariable „Sonstiges“, die nicht zum Mehrfachset gehört. Die muss ausgeschlossen sein.

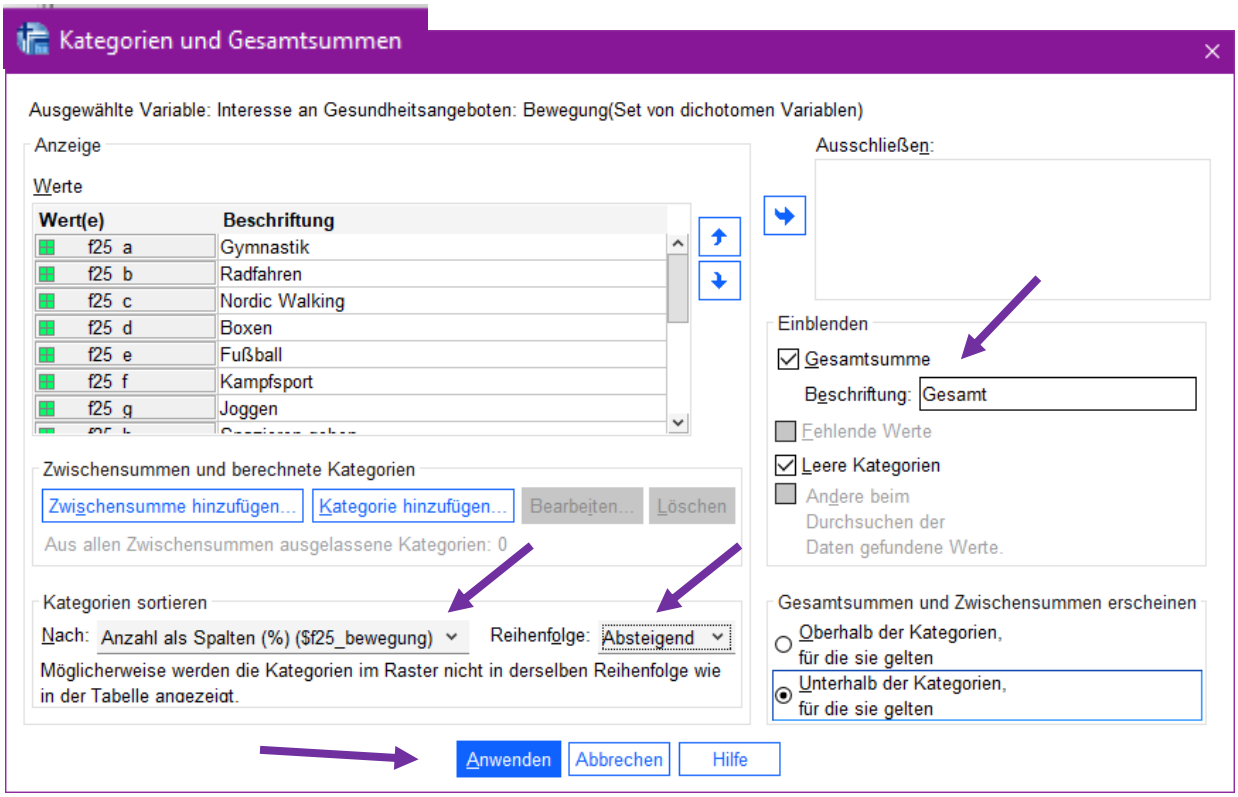


7.2 Erstellen einer Mehrfachantworttabelle

Analysieren → Tabellen → Benutzerdefinierte Tabellen
Das zuvor gebildete Mehrfachantwortset in der Variablenliste suchen
(steht ganz unten mit $ - Zeichen vorangestellt) und in die ***Zeilen*** ziehen: $f25_bewegung
Auswertungsstatistik → Spaltenprozent → *"Anzahl als Spalten %"*auswählen
→ der Auswahl zuweisen → Schließen
Kategorien und Gesamtsummen → Anzeigen, „*Gesamtergebnis*“ auswählen
Kategorien sortieren: nach Anzahl als Spalten% Reihenfolge: Absteigend → Anwenden → OK







		Anzahl	Anzahl als Spalten (%)
$f25_bewegung Interesse an Gesundheits-angeboten: Bewegung	f25_h Spazieren gehen	436	52,7%
	f25_b Radfahren	365	44,1%
	f25_a Gymnastik	230	27,8%
	f25_i Tanzen	227	27,4%
	f25_j Wirbelsäulentraining	174	21,0%
	f25_e Fußball	165	20,0%
	f25_g Joggen	127	15,4%
	f25_k Yoga	127	15,4%
	f25_f Kampfsport	111	13,4%
	f25_l Tai Chi	93	11,2%
	f25_kA Kein Interesse	92	11,1%
	f25_d Boxen	86	10,4%
	f25_m Qi Gong	86	10,4%
	f25_c Nordic Walking	76	9,2%
	f25_p Sonstiges	74	8,9%
	f25_n Pilates	64	7,7%
	f25_o Shiatsu	57	6,9%
	Gesamt	827	100,0%

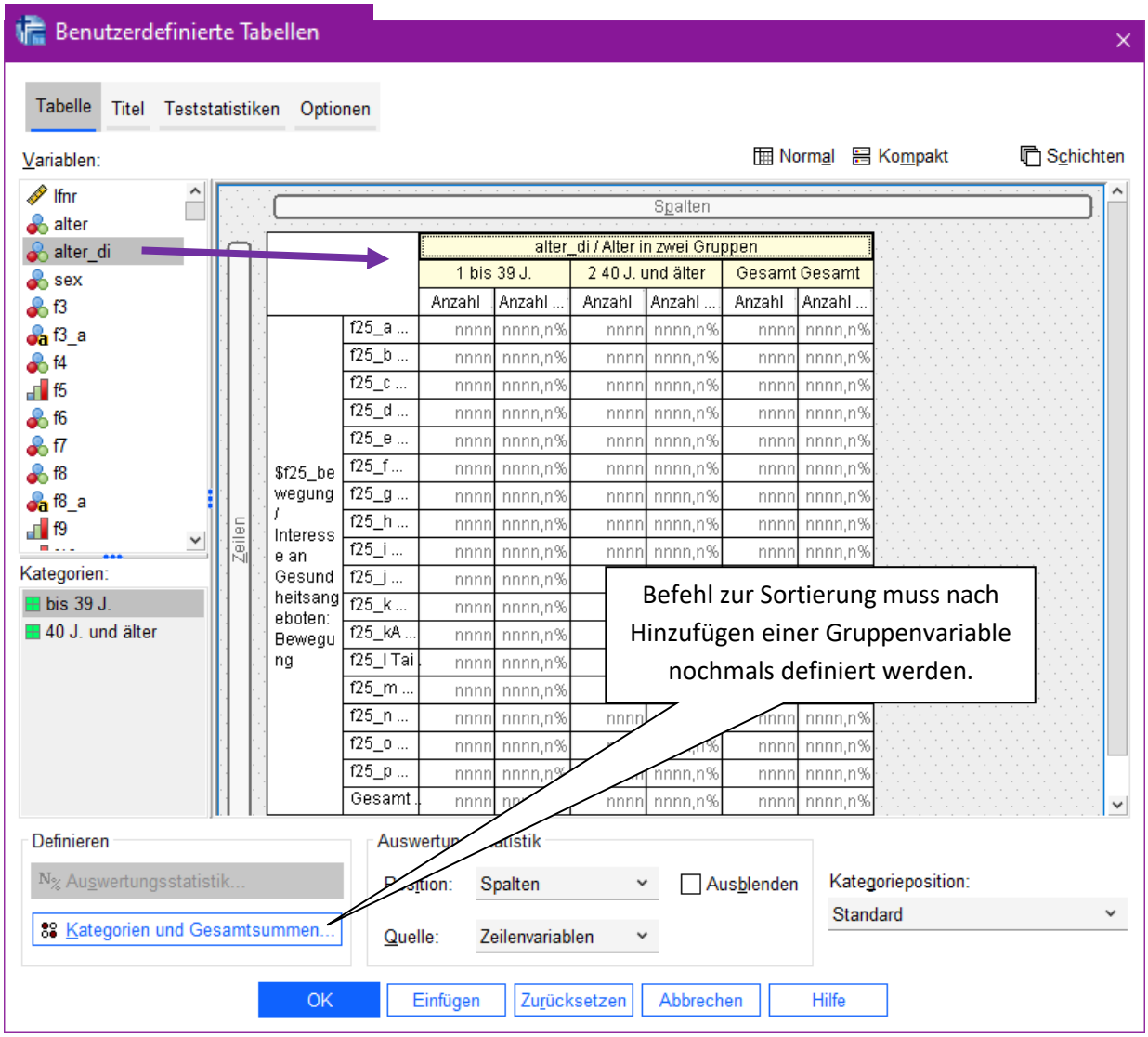
Die absteigende Sortierung der Prozentangaben erleichtert das Lesen und Interpretieren der Tabelle!

Interpretation:
In absteigender Reihenfolge sind die Häufigkeiten der Nennungen aufgelistet. Am häufigsten nennen die Befragten das Interesse am Angebot "Spazieren gehen", gefolgt von "Radfahren". Etwa die Hälfte der Befragten interessiert sich dafür. Etwas mehr als ein Viertel hat Interesse an Angeboten zu "Gymnastik" und "Tanzen", jeweils ein Fünftel an Angeboten zu "Wirbelsäulentraining" und "Fußball". Das Interesse an allen anderen Angeboten ist seltener gegeben. Etwas mehr als 10% haben an keinem der genannten Angebote Interesse.

7.3 Erstellen einer zweidimensionalen Mehrfachantworttabelle

Wir möchten nun untersuchen, ob sich das Interesse an Gesundheitsangeboten zu Bewegung nach Alter unterscheiden. Haben jüngere Befragte andere Interessen als Ältere?

Analysieren → Tabellen → Benutzerdefinierte Tabellen
Mehrfachantwortset **$f25_bewegung** belassen
Zusätzlich die Variable alter_di (Alter dichotom) in die „Spalten" ziehen
Auswertungsstatistik → wie gehabt belassen
Variable alter_di markieren (gelb unterlegt) dann →
→ **Kategorien und Gesamtsummen** → „*Gesamtsumme*" auswählen → anwenden
Dann Mehrfachset f25_bewegung markieren (gelb unterlegt) dann →
→ **Kategorien und Gesamtsummen** → „*Gesamtsumme*" auswählen →
Kategorien sortieren: nach *Anzahl als Spalten%* Reihenfolge: Absteigend → Anwenden → OK



Beachte: Beim Erstellen von Tabellen im Tabellen-Menü ist es wichtig, dass bei den Variablen das entsprechende **Messniveau** definiert ist. Falls das nicht der Fall ist, kann im Tabellen-Menü-Dialogfenster auf die Variable geklickt werden und mit der rechten Maustaste das Messniveau eingestellt werden.

Nicht auf die Spalte für die Gesamtstichprobe vergessen!

$f25_bewegung	alter_di Alter in zwei Gruppen					
Interesse an Gesundheitsangeboten:	1 bis 39 J.		2 40 J. und älter		Gesamt	
Bewegung	Anzahl	Anzahl als Spalten (%)	Anzahl	Anzahl als Spalten (%)	Anzahl	Anzahl als Spalten (%)
f25_h Spazieren gehen	151	48,4%	276	56,2%	427	53,2%
f25_b Radfahren	133	42,6%	220	44,8%	353	44,0%
f25_a Gymnastik	83	26,6%	140	28,5%	223	27,8%
f25_i Tanzen	98	31,4%	124	25,3%	222	27,6%
f25_j Wirbelsäulentraining	60	19,2%	110	22,4%	170	21,2%
f25_e Fußball	85	27,2%	73	14,9%	158	19,7%
f25_g Joggen	61	19,6%	62	12,6%	123	15,3%
f25_k Yoga	50	16,0%	72	14,7%	122	15,2%
f25_f Kampfsport	79	25,3%	28	5,7%	107	13,3%
f25_l Tai Chi	41	13,1%	50	10,2%	91	11,3%
f25_kA Kein Interesse	28	9,0%	61	12,4%	89	11,1%
f25_d Boxen	54	17,3%	27	5,5%	81	10,1%
f25_m Qi Gong	32	10,3%	52	10,6%	84	10,5%
f25_c Nordic Walking	16	5,1%	57	11,6%	73	9,1%
f25_p Sonstiges	25	8,0%	46	9,4%	71	8,8%
f25_n Pilates	19	6,1%	42	8,6%	61	7,6%
f25_o Shiatsu	27	8,7%	28	5,7%	55	6,8%
Gesamt	312	100,0%	491	100,0%	803	100,0%

Interpretation: In absteigender Reihenfolge werden nun die Häufigkeiten der Nennungen getrennt für Befragte unter bzw. über 40 Jahre dargestellt. Die größten Unterschiede (mehr als 5%-Punkte) zeigen sich bei den gekennzeichneten Angeboten:

Ältere Befragte über 40 Jahre interessieren sich häufiger für Spazierengehen (56% versus 48%) und Nordic Walking (12% versus 5%).

Jüngere Befragte unter 40 Jahre nennen deutlich häufiger Interesse an Fußball (27% versus 15%), Kampfsport (25% versus 6%), Boxen (17% versus 5%). Etwas häufiger nennen sie auch Tanzen (31% versus 25%) und Joggen (20% versus 13%). Insgesamt fällt auf, dass es viel mehr Angebote gibt, die für die Jüngeren interessant sind.

Die älteren Befragten haben auch insgesamt etwas häufiger an keinem der aufgelisteten Angebote Interesse genannt (12% versus 9%).

Faustregel: Beschreiben Sie einen Gruppenunterschied nur dann, wenn sich die Gruppen um **zumindest 5%-Punkte** unterscheiden.

7.4 Hausübung 9: Mehrfachantworten

Pflichtaufgabe 9

Erstellen Sie eine **Mehrfachantworttabelle** mit einem Set Ihrer Wahl und beschreiben Sie die Ergebnisse. Erstellen Sie weiters eine **zweidimensionale Mehrfachantworttabelle** mit diesem Set und einer Gruppenvariable, von der Sie meinen, dass diese einen Unterschied im Antwortverhalten zeigt. (Tipp: Wählen oder erstellen Sie eine Variable mit nur zwei Kategorien.) Zeigen sich die von Ihnen erwarteten Unterschiede? Interpretieren Sie die Ergebnisse.

8 Richtlinien zur Hausarbeit

Die in diesem Skriptum zusammengestellten Aufgaben stellen die Grundlage zur Note dar.
Die Aufgaben können alleine oder zu zweit bearbeitet werden.

Kopieren Sie die Tabellen und Grafiken, die Sie im SPSS erstellen, ins Word, und arbeiten Sie dort die Aufgaben weiter aus. Besonders wichtig sind die Interpretationen. Am Ende des Semesters laden Sie die perfekt layoutierte **Hausarbeit im PDF-Format** in moodle hoch.

Folgende formale Anforderungen an die Hausarbeit müssen erfüllt sein:
→ Standard-Deckblatt (von HP) mit Lehrveranstaltungstitel, Name und Matrikelnummer, Datum
→ Inhaltsverzeichnis mit Seitenzahlen
→ Vorbildliches Layout - das heißt: keine Tabellen über zwei Seiten, übersichtliche Darstellungen, leser:innenfreundliche Gestaltung
→ Ausführliche und verständliche Interpretationen in EIGENEN Worten und GANZEN Sätzen
→ Erstellen eines pdf und hochladen in moodle.

Punktevergabe:

Pflichtaufgabe 1: Erstellen Datenfile	10
Pflichtaufgabe 2: Häufigkeitstabelle	10
Pflichtaufgabe 3: Kennzahlen	12
Pflichtaufgabe 4: Rekodieren	10
Pflichtaufgabe 5: Kreuztabelle	12
Pflichtaufgabe 6: Chi-Quadrat	12
Pflichtaufgabe 7: t-Test	12
Pflichtaufgabe 8: U-Test	12
Pflichtaufgabe 9: Mehrfachantwort	10
Gesamt	**100**

Notenschlüssel:

91 bis 100 Punkte	Sehr Gut
81 bis 90 Punkte	Gut
71 bis 80 Punkte	Befriedigend
61 bis 70 Punkte	Genügend
0 bis 60 Punkte	Nicht Genügend