

How to SPSS und R für die Auswertung von Daten aus dem Sprachverständnistest (SpraveST) von Levumi

Jost, L. & Jungjohann, J.

Unter Mitarbeit von Laura Schmitz

Version 2.0

September 2025

Luisa Jost

<https://orcid.org/0000-0003-2561-9101>

Technische Universität Dortmund

Jun.-Prof. Dr. Jana Jungjohann

<https://orcid.org/0000-0002-9985-4780>

Technische Universität Dortmund

Veröffentlicht unter der Lizenz:

CC-BY-SA 4.0



Zusammenfassung

Das vorliegende How to erläutert den Umgang mit dem Statistikprogramm SPSS und der Programmiersprache R für die Datenauswertung mit Daten aus dem Sprachverständnis (SpraveT; Althelmig & Jungjohann, [2025](#)) aus Levumi.de (Jungjohann, Diehl et al., [2018](#)). Das How to bietet Vorschläge zum Vorgehen zur Datenauswertung und dient als Hilfestellung mit den generierten Daten aus dem SpraveT. Die Zielgruppe sind vorrangig Studierende, die ihre Abschlussarbeiten am [Fachgebiet Digitale Förderung und Inklusive Bildung](#) der TU Dortmund schreiben. Es werden grundlegende Hinweise zur Datenaufbereitung sowie konkrete Vorgehensweisen zur Auswertung deskriptiver und inferenzstatistischer Analysen bereitgestellt.

Schlagwörter

Datenauswertung, Statistik, Sprachverständnis

Zitation

Jost, L. & Jungjohann, J. (2025). *How to SPSS und R für die Auswertung von Daten aus Levumi Version 2.0*.

Version

Die vorliegende Version 2.0 basiert auf der Anleitung „How to SPSS für die Auswertung von Daten aus Levumi Version 1.0“ von Jana Jungjohann, Hanna Rütter und Jeffrey DeVries (unveröffentlicht).

Inhaltsverzeichnis

1 Software zur Auswertung von Daten aus Levumi	1
2 Vorbereitung des Datensatzes mit R	3
2.1 Datensätze aus Levumi	3
2.2 Umwandlung des Datensatzes in R	6
2.3 Datenaufbereitung in R	10
3 Arbeit mit SPSS	13
3.1 Datenimport in SPSS	13
3.2 Die drei Fenster von SPSS	14
3.3 Die Ansichten im Dateneditor von SPSS	15
3.4 Umgang mit Variablen in SPSS	17
3.5 Variablen berechnen in SPSS	24
3.6 SPSS-Datensatz speichern	25
4 Statistische Analysen in SPSS mit Levumi-Daten	26
4.1 Deskriptive Statistik in SPSS	26
4.1.1 Lagemaße in SPSS	27
4.1.2 Streuungsmaße in SPSS	33
4.1.3 Häufigkeitsverteilung in SPSS	35
4.1.4 Verteilungsmaße in SPSS	37
4.2 Inferenzstatistik in SPSS	38
4.2.1 Signifikanz	38
4.2.2 Voraussetzungsprüfung für inferenzstatistische Analysen	39
4.2.3 T-Test in SPSS	47
4.2.4 Einfaktorielle Varianzanalyse in SPSS	50
4.2.5 Korrelationen in SPSS	57
5 Abschluss der Datenauswertung	60

Abbildungsverzeichnis

Abbildung 1 <i>Levumi-Daten exportieren</i>	3
Abbildung 2 <i>Levumi-Datenexport im long-Format in Excel</i>	5
Abbildung 3 <i>Levumi-Datenexport wide-Format in Excel</i>	6
Abbildung 4 <i>Aufrufen des R-Skriptes</i>	7
Abbildung 5a <i>Arbeitsverzeichnis bestimmen</i>	8
Abbildung 5b <i>Arbeitsverzeichnis bestimmen</i>	8
Abbildung 6 <i>R Environment mit importiertem Datensatz SpraVesT1</i>	9
Abbildung 7 <i>R Konsole mit Fehlermeldungen in rot</i>	10
Abbildung 8 <i>SPSS Textimport-Assistent</i>	13
Abbildung 9 <i>SPSS Dateneditor</i>	14
Abbildung 10 <i>SPSS Syntaxfenster</i>	14
Abbildung 11 <i>SPSS Ausgabefenster</i>	15
Abbildung 12 <i>SPSS Ansichten</i>	15
Abbildung 13 <i>SPSS Übersichtsansicht</i>	16
Abbildung 14 <i>SPSS Datenansicht</i>	16
Abbildung 15 <i>SPSS Variablenansicht</i>	17
Abbildung 16 <i>SPSS Variablentyp definieren</i>	18
Abbildung 17 <i>SPSS Fehlende Werte festlegen</i>	19
Abbildung 18 <i>SPSS-Syntaxfenster mit Code zur Transformation von NA-Werten</i>	20
Abbildung 19 <i>SPSS Breite festlegen</i>	21
Abbildung 20 <i>SPSS Beschriftungen festlegen</i>	21
Abbildung 21 <i>SPSS Wertbeschriftungen</i>	22
Abbildung 22 <i>SPSS Messniveau festlegen</i>	23
Abbildung 23 <i>SPSS Variable berechnen</i>	24
Abbildung 24 <i>SPSS-Output Lagemaße</i>	27
Abbildung 25 <i>SPSS-Output zur Itemschwierigkeit</i>	29
Abbildung 26 <i>SPSS neuen Datensatz erstellen</i>	29
Abbildung 27a <i>SPSS Eingabe neuer Variablen (wide-Format)</i>	30
Abbildung 27b <i>SPSS Eingabe neuer Variablen (long-Format)</i>	30

Abbildung 28a SPSS Eingabe Ausprägungen einzelner Werte (wide-Format)	30
Abbildung 28b SPSS Eingabe Ausprägungen einzelner Werte (long-Format)	31
Abbildung 29 SPSS Eingabe Transformation (Variable berechnen).....	32
Abbildung 30 SPSS Datenansicht mit der neuen Variable Summenscore	32
Abbildung 31 SPSS-Output Streuungsmaße	34
Abbildung 32 SPSS-Output Häufigkeiten	35
Abbildung 33a SPSS-Output zur Häufigkeitsverteilung > Statistiken.....	36
Abbildung 33b SPSS-Output zur Häufigkeitsverteilung > Häufigkeitstabellen	36
Abbildung 33c SPSS-Output zur Häufigkeitsverteilung > Balkendiagramm	37
Abbildung 34 SPSS Eingabefenster Streudiagramm	40
Abbildung 35 SPSS-Output Streudiagramm (ohne linearen Zusammenhang).....	40
Abbildung 36 SPSS Eingabefenster zur Explorativen Datenanalyse	42
Abbildung 37 SPSS Eingabefenster zur Explorativen Datenanalyse > Diagramme	42
Abbildung 38 SPSS-Output Tests auf Normalitätsverteilung.....	43
Abbildung 39 SPSS-Output Stengel-Blatt-Diagramm	44
Abbildung 40 SPSS Eingabefenster zur Explorativen Datenanalyse.....	45
Abbildung 41 SPSS Eingabefenster zur Explorativen Datenanalyse	45
Abbildung 42 SPSS Eingabefenster zur Explorativen Datenanalyse	46
Abbildung 43 SPSS Eingabefenster zu t-Tests bei Stichproben mit paarigen Werten	47
Abbildung 44 SPSS-Output zu t-Tests bei Stichproben mit paarigen Werten	48
Abbildung 45 SPSS Eingabefenster zu t-Tests bei unabhängigen Stichproben	49
Abbildung 46 SPSS-Output zu t-Tests bei unabhängigen Stichproben inkl. Gruppenstatistiken	49
Abbildung 46b SPSS-Output zu t-Tests bei unabhängigen Stichproben > Effektgrößen	50
Abbildung 47 SPSS Eingabefenster zur einfaktoriellen ANOVA.....	52
Abbildung 48 SPSS Eingabefenster zur einfaktoriellen ANOVA > Optionen.....	52
Abbildung 49a SPSS-Output zur einfaktoriellen ANOVA > Deskriptive Statistiken	53
Abbildung 49b SPSS-Output zur einfaktoriellen ANOVA > Tests der Varianzhomogenität.....	53
Abbildung 50 SPSS-Output zur einfaktoriellen ANOVA > ANOVA-Tabelle und Effektgrößen.....	54
Abbildung 51a SPSS-Output zur einfaktoriellen ANOVA > Post-Hoc Analysen	54
Abbildung 51b SPSS-Output zur einfaktoriellen ANOVA > Homogene Gruppen	55

Abbildung 52 SPSS Eingabefenster zur einfaktoriellen ANOVA mit Messwiederholung	56
Abbildung 53 SPSS-Output zur einfaktoriellen ANOVA mit Messwiederholung > Innersubjektkon- traste.....	56
Abbildung 54 SPSS Eingabefenster zu Korrelationen	58
Abbildung 55a SPSS-Output zu Korrelationen nach Pearson.....	58
Abbildung 55b SPSS-Output zu Korrelationen nach Spearman	59

1 Software zur Auswertung von Daten aus Levumi

Für die statistische Auswertung der Daten aus Levumi eignet sich eine Kombination aus der Programmiersprache R und der Statistik Software SPSS. Es gibt eine Vielzahl von möglicher Software und Programmiersprachen zur Auswertung von statistischen Daten, die individuelle Vor- und Nachteile je nach Struktur der Daten sowie der Fragestellungen haben. Gängig sind zum Beispiel Statistical Analysis System (SAS; SAS Institute Inc., 2025), Statistical Package for the Social Sciences (SPSS; IBM Corp., 2025), die Programmiersprache R sowie die Programmiersprache Python. Alternativ kann auch Jamovi (The Jamovi project, 2025), eine Open Source Software zur statistischen Datenauswertung genutzt werden. Jamovi basiert auf der Programmiersprache R und ermöglicht die Analyse kleiner Datenmengen mit wenig komplexen Verfahren. Jamovi ist zur Auswertung komplexer Daten weniger geeignet als SAS und SPSS. Zudem ist SPSS gegenüber SAS kostengünstiger und wird in den Sozialwissenschaften und der Psychologie primär zur Datenanalyse eingesetzt. Für die Aufbereitung der Levumi-Daten wird daher das Verfahren für SPSS beschrieben.

Python (Python Software Foundation, 2024) und R (R Core Team, 2025) sind beide Open Source Programmiersprachen, allerdings ist Python als Programmiersprache entwickelt worden, während R mit dem Ziel der statistischen Auswertung von Daten entwickelt wurde (de Vries & Meys, 2012). Beide Softwares eignen sich gut, R ist dabei die primär genutzte Sprache für die Aufbereitung von Levumi-Daten, weshalb die Sprache auch hier eingesetzt wird.

R ist eine Open Source Programmiersprache, entwickelt für die statistische Datenanalyse, die neben dem Basispaket eine Vielzahl an zusätzlichen Paketen für spezifische Analysen aufweist (Sedlmeier & Burkhardt, 2021). Diese werden von einem umfangreichen Netzwerk an Wissenschaftler:innen entwickelt. R wird kostenfrei ohne Einschränkungen oder Vorbedingungen vom [Comprehensive R Archive Network](#) (CRAN) bereitgestellt. Die hier bereitgestellte Anleitung zur Aufbereitung der Levumi-Daten für das Statistikprogramm SPSS basiert auf der R Version 4.4.3 (2025-02-28; R core Team, 2025). R wird in dieser Anleitung zur Aufbereitung des Levumi Datenexports eingesetzt, um anschließend in SPSS die statistische Datenanalyse durchzuführen. Zum einfachen Einsatz der Programmiersprache empfiehlt es sich, [R Studio](#) herunterzuladen.

SPSS ist eine Software für die statistische Datenanalyse, die von der International Business Machines Corporation (IBM, 2025) bereitgestellt wird. Studierende der TU Dortmund können sich beim [Medienarbeitsraum für Studierende](#) ein Endgerät mit der Software ausleihen. Die hier bereitgestellte Anleitung zu SPSS bezieht sich auf die SPSS Version 29.

Download von R, RStudio & SPSS

2 Vorbereitung des Datensatzes mit R

Das hier geschilderte Vorgehen ist nur eine der Möglichkeiten zur statistischen Auswertung der aus Levumi.de (Jungjohann, deVries et al., [2018](#)) exportierten Daten. Bei der Auswertung gibt es vielfältige weitere Möglichkeiten, sodass das hier geschilderte Vorgehen nur eine Orientierung zur Auswertung der Daten bietet und erweitert oder abgewandelt werden kann.

2.1 Datensätze aus Levumi

Die **Exportdaten** von Levumi sind je nach Test verschieden aufgebaut. Diese Anleitung bezieht sich auf die Daten aus dem Sprachverständnistest (SpraVesT; Althelmig & Jungjohann, [2025](#), Althelmig & Jungjohann, [2024](#)). Diese Daten werden, wie alle Daten in Levumi über „Testungen exportieren“ (Abbildung 1) als .csv Datei in einem .zip-Ordner gespeichert.

Abbildung 1

Levumi-Daten exportieren



Der Datenexport wird nach folgender Logik benannt: SpraVesT_[Variante 1 oder 2]_[Datum]. In diesem Beispiel wird ein Beispiel-Datensatz (hier benannt als „SpraVesT_1_2025_01_16.csv“) verwendet. Der Export für den Test SpraVesT enthält beim Export 18 **Variablen** (Spalten; Tabelle 1).

Tabelle 1

Variablen aus dem Levumi-Datenexport

Variable	Erläuterung
Ergebnis_ID	Eindeutige Ergebnis-ID pro durchgeführte Testung pro Kind
Kind_ID	Anonymisierter ID-Code des Kindes
Kind_Login	Levumispezifischer Login, nicht anonym
Klassen_ID	Eindeutige ID für die Zuordnung zur Klasse des getesteten Kindes, jede Klasse hat eine eigene ID
Geschlecht	0 = weiblich, 1 = männlich, 2 = divers
Geburtsdatum	Angabe der Lehrkräfte zum Geburtsdatum der Kinder. Der Tag wird automatisch auf 1 gesetzt. NA = nicht angegeben
SPF	Sonderpädagogischer Förderbedarf 0 = Keiner, 1= Lernen, 2=Geistige Entwicklung, 3=Emotional-soziale Entwicklung, 4=Sprache, 5=Körperlich-motorische Entwicklung, 6=Sehen, 7=Hören, 8=Autismus, NA= nicht angegeben
Test_ID	ID des durchgeführten Tests = 360 (SpraVest Variante 1, in Hauptstudie verwendet), ID = 359 (SpraVest Variante 2, nur in Teilstichprobe verwendet)
Tags	Angabe der Lehrkräfte zu weiteren Merkmalen der Kinder (Mehrfachauswahl möglich, in Textform)
Testwoche	Immer auf den Montag der Woche, in welcher der Test durchgeführt wurde, datiert
Item	Angabe des Items mit folgenden Codes: I2 – I81. I1 kommt nicht vor, da es das Demo-Item ist.
Time	Reaktionszeit in ms
Group	„Group“ bezeichnet die Dimensionen der Items 1= Demo-Item, kommt daher nicht im Export vor, 2= Dimension (Sätze mit Variation im Numerus, externe Zählweise Dim2), 3= Dimension (Relativsätze, externe Zählweise Dim4), 4= Dimension (Sätze mit Steigerung, externe Zählweise Dim3), 5= Dimension (Präpositionalsätze, externe Zählweise Dim1)
Answer	Bezeichnung des korrekten Bildes, als webp. angegeben
Result	0=falsche Antwort, 1= korrekte Antwort, „Nicht beantwortet“ (= es wurde auf „?“ geklickt; kann durch nicht planmäßige Randomisierung im Datensatz enthalten sein)
Question	Beschreibt den Zielsatz, welcher mit der Audioausgabe vorgespielt wird.
AdditionalData	Häufigkeit des Abspielens der Audioausgabe, (1x standard default für alle): (1x) usw.

Die **Beobachtungen** (Zeilen) sind jeweils pro Antwort zu einem bearbeiteten Item eines Kindes abgespeichert. Jedem Kind ist demnach nicht nur eine Zeile zugeordnet, sondern zwischen 1 und 80 Zeilen, abhängig von der Anzahl der beantworteten Items. Der

Datensatz liegt also im **long-Format** vor. Das long-Format ist für den Menschen weniger intuitiv, hat aber den Vorteil, dass die Daten leichter zu filtern sind und dieses Format flexiblere Analysen zulässt (Wickham, 2014).

Neben dem long-Format gibt es auch das **wide-Format**, bei dem zu jeder Person nur eine Zeile zugewiesen ist, die dann mehrere Spalten für die einzelnen Messwerte (hier: Items) hat. Das wide-Format ist für deskriptive Statistiken und grundlegende statistische Analysen besser geeignet, weil es für den Menschen eingängiger ist und die Analyse in Statistikprogrammen so erleichtert (ebd.).

Der Unterschied der Datensätze wird bei der Ansicht in Excel sichtbar. Der gewählte Datensatz hat im Export von Levumi (long-Format) 18 Spalten und 2305 Zeilen (Abbildung 2). Umgewandelt in das wide-Format hat derselbe Datensatz ohne weitere Veränderungen 491 Spalten und 60 Zeilen (Abbildung 3).

Der Unterschied wird beim Vergleich der Variable Ergebnis_ID sichtbar. Im long-Format ist diese über alle Zeilen des Ausschnittes gleich („238597“). Im wide-Format wiederholen sich die Ergebnis-IDs nicht. Für die Datenanalyse werden beide Formate (wide und long) benötigt, weshalb der Datensatz zunächst vom long ins wide-Format umgewandelt wird. Hierfür eignet sich die Programmiersprache R.

Abbildung 2
Levumi-Datenexport im long-Format in Excel

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Ergebnis_ID	Kind_ID	Kind_Login	Klassen_ID	Geschlecht	Geburtsdatum	SPF	Tags	Test_ID	Testdatum	Testwoche	Item	time	group
2	238597	77325		12452	1	01.12.14	1		360	10.12.24 07:50	09.12.24	12	13060	2
3	238597	77325		12452	1	01.12.14	1		360	10.12.24 07:50	09.12.24	13	11055	3
4	238597	77325		12452	1	01.12.14	1		360	10.12.24 07:50	09.12.24	14	16727	4
5	238597	77325		12452	1	01.12.14	1		360	10.12.24 07:50	09.12.24	15	4775	5
6	238597	77325		12452	1	01.12.14	1		360	10.12.24 07:50	09.12.24	16	23939	2
7	238597	77325		12452	1	01.12.14	1		360	10.12.24 07:50	09.12.24	17	5942	4
8	238597	77325		12452	1	01.12.14	1		360	10.12.24 07:50	09.12.24	18	5164	3
9	238597	77325		12452	1	01.12.14	1		360	10.12.24 07:50	09.12.24	19	7832	5

Anmerkung. Die Ausprägungen der Variable Kind_Login sind aus Datenschutzgründen geschwärzt.

Abbildung 3
Levumi-Datenexport wide-Format in Excel

	A	B	C	D	E	F	G	H	I	J	K
1	Ergebnis_ID	Kind_ID	Kind_Login	Klassen_ID	Geschlecht	Geburtsdatum	SPF	Tags	Test_ID	Testdatum	Testwoche
2	243148	77220		18887	1	01.02.17	3	AD(H)S,Erstsprache: Deutsch,Klasse: SE2	359	18.12.24 11:07	16.12.24
3	243158	77221		18887	1	01.06.17	3	2. Förderbedarf: Lernen,Klasse: SE2	359	18.12.24 11:09	16.12.24
4	243159	77222		18887	0	01.02.16	3	Erstsprache: Deutsch,Klasse: SE3,Laufendes Prüfverfahren: 2. FB: GG / LE	359	18.12.24 11:09	16.12.24
5	243173	77223		18887	1	01.12.16	3	2. Förderbedarf: Lernen,ASS,Erstsprache: Deutsch,Klasse: SE2	359	18.12.24 11:11	16.12.24
6	243174	77224		18887	1	01.01.14	3	2. Förderbedarf: Lernen,AD(H)S,Deutsch als Zweitsprache,Erstsprache: Türkisch,Fluchtgeschichte,K	359	18.12.24 11:11	16.12.24
7	243180	77226		18887	1	01.09.22	3	Erstsprache: Deutsch,Klasse: SE2,weiterer Förderbedarf: SQ, LE	359	18.12.24 11:12	16.12.24
8	238739	77228		18888	1	01.01.15	3	2. FS: Lernen,Erstsprache: Deutsch,Klasse: 3	359	10.12.24 08:40	09.12.24
9	238741	77231		18888	1	01.03.15	3	Erstsprache: Deutsch,Klasse: 3	359	10.12.24 08:40	09.12.24

Anmerkung. Die Ausprägungen der Variable Kind_Login sind aus Datenschutzgründen geschwärzt.

2.2 Umwandlung des Datensatzes in R

Die Arbeit mit R erfolgt durch die Kombination von einzelnen **Funktionen** in **Auswertungsskripten**, die eine wiederholbare und nachvollziehbare Arbeit mit Daten ermöglichen. Eine Funktion ist immer gleich aufgebaut und besteht aus einer Funktionsbezeichnung (z. B. `read.csv()` für das Einlesen von csv-Dateien) und einem Ausdruck in Klammern, der spezifiziert, auf welche Datenrahmen, Variablen oder Vektoren die Funktion angewandt werden sollen sowie spezifische Argumente, die bestimmen, wie eine Funktion konkret ausgeführt werden soll.

Ein Beispiel hierfür wäre:

```
read.csv(file="SpravesT_1_2025_01_16.csv", header= TRUE, sep= ",", stringsAsFactors = FALSE, encoding = "UTF-8").
```

In dieser Funktion bestimmt die **Funktionsbezeichnung** `read.csv`, welche Funktion ausgeführt wird. Die einzelnen **Argumente** spezifizieren, wie die Funktion konkret ausgeführt wird.

Für die Umwandlung und Aufbereitung des SpravesT-Datensatzes steht ein vorbereitetes **R-Skript** zur Verfügung, das entwickelt wurde, um den Datensatz in das wide-Format zu transformieren. Dieses Skript enthält alle erforderlichen Schritte, wie das Laden der Daten, das Bereinigen von Fehlern und das Umstrukturieren der Daten. Das R-Skript muss demnach nur an wenigen Stellen abgewandelt werden, um es für den eigenen Datensatz zu nutzen. Es ist über folgenden Link abrufbar: https://osf.io/b6wzp/?view_only=933e7afe2b6346d6bfa6c6815e813bb5.

Zur Arbeit mit dem R-Skript und dem eigenen Datensatz sollte beides in einen gemeinsamen Ordner gespeichert werden. Das vereinfacht später den Zugriff auf den gesuchten Datensatz.

Speichern von Datensatz und Syntax in einem gemeinsamen Ordner

Beim Öffnen des R-Skriptes in **RStudio** wird das R-Skript zunächst nur aufgerufen aber noch nicht ausgeführt (Abbildung 4). Oben finden sich Informationen zum Skript und den benötigten Paketen. Diese müssen zuvor installiert werden. Dazu kann der Befehl `install.packages(„[Name des Paketes]“)` verwendet werden oder es wird beim Öffnen des Skriptes bereits durch RStudio darauf verwiesen, dass diese Pakete erforderlich sind und installiert werden müssen. Um eine Funktion auszuführen, muss der Cursor in die jeweilige Zeile geklickt werden und der Code mit `strg + Enter` (bei Mac: `cmd + Enter`) ausgeführt werden.

Öffnen des R-Skriptes, installieren und laden der benötigten Pakete in RStudio

Hinweis zur Arbeit mit diesem How To:

Durch das Ergänzen oder Verändern des R-Skriptes verschiebt sich unter Umständen die Nummerierung der Zeilen. Die Zeilenangaben beziehen sich daher auf den unbearbeiteten Syntax.

Abbildung 4
Aufrufen des R-Skriptes

```

1 ## Data Transformation and Preparation for Sprachverständnistest (SpravesT)
2 ## Last changed: 20-03-2025
3 ## Contact: Luisa Jost; luisa2.jost@tu-dortmund.de
4 ## data: Export data from www.levumi.de (here: "SpravesT_1_2025_01_16.csv")
5
6 # This syntax transforms the export created in levumi.de from long to wide format
7 # Moreover, it cleans and prepares the data by handling missing values, renaming variables,
8 # and restructuring the dataset for further analysis.
9
10 ### Needed Packages
11 library(dplyr)
12 library(readr)
13 library(tidyr)

```

Um das R-Skript zu nutzen, muss noch festgelegt werden, wo die Daten liegen, also auf welches Arbeitsverzeichnis (engl. **working directory**) zurückgegriffen wird. Hierzu werden auf dem PC im **Dateimanager** (bei Mac: Finder) über Rechtsklick auf den Ordner, der Daten und R-Skript die **Eigenschaften** (bei Mac: Informationen) abgerufen (Abbildung 5a und Abbildung 5b).

Abbildung 5a
Arbeitsverzeichnis bestimmen

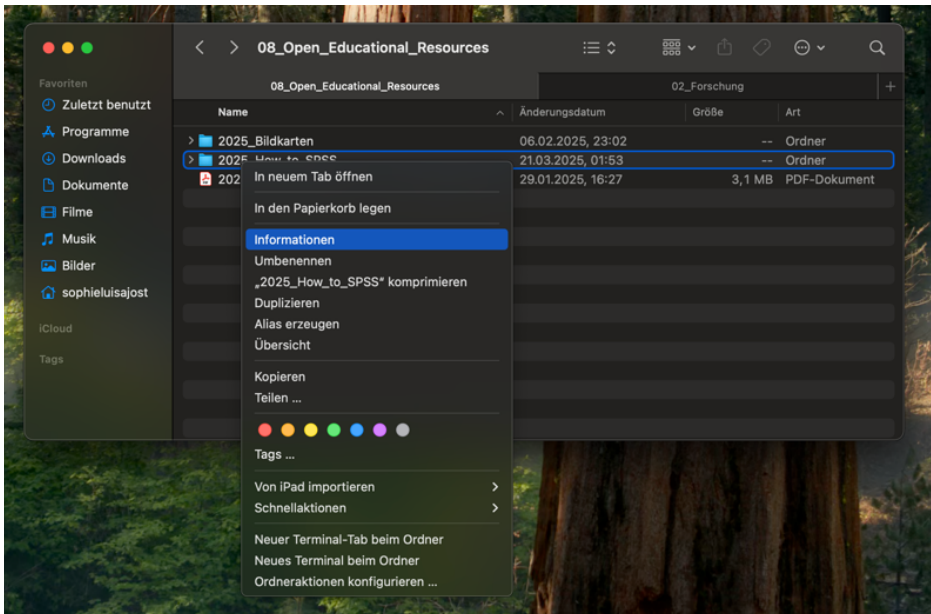
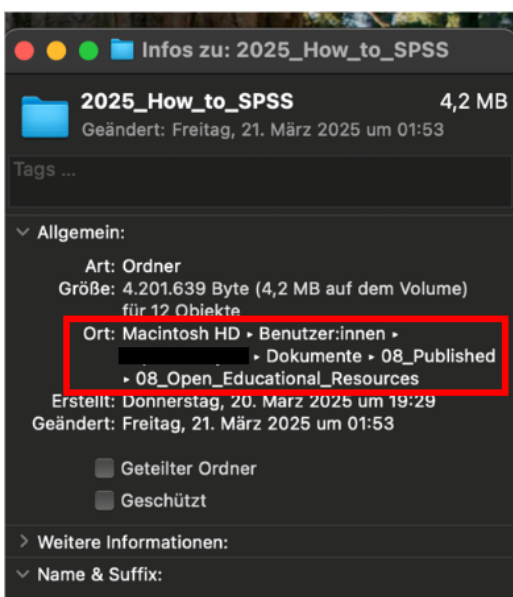


Abbildung 5b
Arbeitsverzeichnis bestimmen



Anmerkung. Ein Teil des Ortes ist aus Datenschutzgründen geschwärzt.

Dort ist der **Speicherort** (bei Mac: Ort) einzusehen. Im Befehl `setwd()` (Zeile 18 des R-Skriptes) muss dieser Speicherort festgelegt werden. Dabei ist zu beachten, dass bei den Systemordnern (z. B. Benutzer:innen oder Dokumente) immer der Englische Begriff genutzt wird.

Ergänzen des richtigen working directory im R-Skript
`setwd(„/Users/[Name des Nutzenden]/[Ordner: z.B. Dokumente]/[ggf. weitere Ordner]/[Name des Ordners mit den Daten“]`

Zum Import der Daten muss nun noch im Skript (Zeile 22) der Titel der Exportdatei aus Levumi (hier: „SpraVesT_1_2025_01_16.csv“) eingefügt werden. Wenn zwei Varianten des Tests bearbeitet wurden (SpraVesT 1 und 2) werden beide Datensätze einzeln exportiert. In diesem Fall muss der Titel der Exportdatei für SpraVesT 2 ebenfalls angepasst werden. Sonst kann dieser gelöscht oder ignoriert werden. Zum praktikablen Umgang können auch `#` vor die entsprechenden Zeilen gesetzt werden, um sie nicht auszuführen. R interpretiert die Raute `#` und alle folgenden Zeichen in der Zeile nicht als Befehl. So wird das R-Skript auch kommentiert. Hier kann es helfen, damit die Funktionen für die nicht benötigte Variante nicht gelöscht werden müssen, sondern einfach nicht ausgeführt werden.

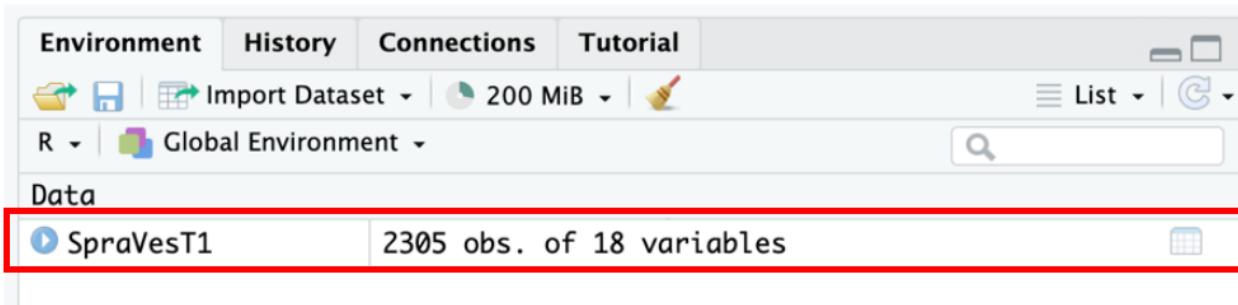
Ergänzen des richtigen Dateititels im R-Skript

```
SpraVesT1 <- read.csv(file= "SpraVesT_1_2025_01_16.csv", header= TRUE, sep= ",",
stringsAsFactors = FALSE, encoding = "UTF-8") # import the data from version 1
```

Der Datensatz erscheint im **Environment** (Arbeitsbereich mit Datenobjekten, die während einer R-Sitzung erstellt, bzw. geladen werden; Abbildung 6). Danach kann der Code unverändert bis Zeile 135 ausgeführt werden.

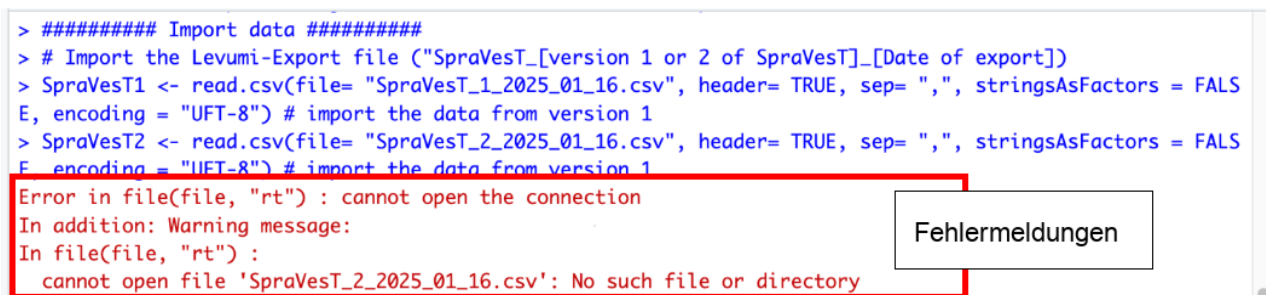
Abbildung 6

R Environment mit importiertem Datensatz SpraVesT1



Wenn SpravesT1 und 2 geladen wurden, gibt es in der Regel keine Fehlermeldungen. Wenn nur eine der beiden Varianten geladen wurde, kommt es zu einigen Fehlermeldungen, die in der Konsole (Abbildung 7) angezeigt werden. Das liegt am Fehlen der entsprechenden Daten der anderen Variante. Diese Fehlermeldungen können ignoriert werden.

Abbildung 7
R Konsole mit Fehlermeldungen in rot



Nach erfolgreichem Durchlaufen des R-Skriptes bis Zeile 135 liegen 5 Datenrahmen (ein Objekt zum Sammeln von Vektoren in R, vorstellbar als Tabelle) und 1 Funktion im Environment. Der Datensatz *SpravesT1_wide* liegt im wide-Format und ohne Duplikate vor. Um diesen in SPSS zu nutzen, wird er im Skript in eine .csv-Datei geschrieben. Diese Datei wird im entsprechenden Ordner abgelegt, der zuvor als working directory festgelegt wurde.

2.3 Datenkontrolle in R

Im R-Skript wird anschließend überprüft, ob es **Duplikate** (also Dopplungen in den Daten) gibt (ab Zeile 146). Wenn es keine Duplikate gibt, kann das Skript einfach weiter durchlaufen werden. Gibt es Duplikate, muss eine Entscheidung getroffen werden, welche behalten werden. Eine ungeplante zweite Messung am Testtag kann zum Beispiel

durch einen technischen Fehler entstanden sein und wird gelöscht oder weitere Messungen, die nicht für die Analysen relevant sind. Hierzu muss die Kind_ID des entsprechenden Duplikates ersetzt werden.

Kind_ID ersetzen

```
rowSums(SpraVesT1_clean[SpraVesT1_clean$Kind_ID == [Kind_ID einfügen],
      grep("_result", names(SpraVesT1_clean), value = TRUE)], na.rm = TRUE)
```

```
View(SpraVesT1_clean[SpraVesT1_clean$Kind_ID == [Kind_ID einfügen],])
```

Hier muss händisch entschieden werden, welcher Wert beibehalten wird:

```
SpraVesT1_clean <- SpraVesT1_clean[SpraVesT1_clean$Ergebnis_ID != [Ergebnis_ID
einfügen, die gelöscht werden soll], ]
```

Um in SPSS mit den **(demografischen) Informationen** über die Proband:innen rechnen zu können, müssen diese zunächst aufbereitet werden. Die Variable *Tags* beinhaltet verschiedene Informationen (z. B. Migrationshintergrund). Damit diese ausgewertet werden können, müssen alle gesetzten Tags und die eingegebenen Zusatzinformationen in einzelnen Variablen gespeichert werden. Im R-Skript wird das für Migrationshintergrund, Fluchtgeschichte, Deutsch als Zweitsprache, LRS, Dyskalkulie, Hochbegabung, Integrationskraft, ADHS, ADS gemacht (ab Zeile 177). Wenn darüberhinausgehende Informationen eingegeben wurden, müssen neue Variablen ergänzt werden.

Evtl. Ergänzen von Variablen zu Proband:innen-Informationen

```
SpraVesT1_clean_info$[Neue Variable] <- grep(„[Begriff]“, SpraveT1_clean_info$Tags)
```

Die Klassenstufe wird nach dem gleichen Prinzip ergänzt (ab Zeile 202). Hier müssen die Klassenbezeichnungen angepasst werden. Im Beispiel wurden die Klassen SE2, 2.1 und 3 getestet. Das bedeutet, es muss nach diesen Tags gesucht werden.

Anpassen der Klassenbezeichnungen

```
SpraVesT1_clean_info$[Bezeichnung Klasse] <-grep("Klasse\\: [Bezeichnung]",
SpraVesT1_clean_info$Tags)
```

Anschließend muss die richtige Klassenstufe zugewiesen werden. Das muss ebenfalls händisch angepasst werden.

Zuweisen der Klassenstufe

```
SpraVesT1_clean_info$grade[SpraVesT1_clean_info$KlasseSE2 == TRUE] <- 2
```

Die nun neu generierten Variablen sind im Datensatz *SpraVesT1_clean_info* gespeichert und werden wieder in eine .csv-Datei geschrieben.

Dieser Datensatz kann, falls erforderlich auch wieder in das long-Format überführt werden (ab Zeile 308).

Ausführen des R-Skriptes

Der Abschnitt „return to long format“ im R-Skript muss ausgeführt werden.

Als Ergebnis des R-Skriptes liegen pro Variante fünf Datensätze vor (Tabelle 2).

Tabelle 2

Datensätze aus dem R-Skript zur Weiternutzung in SPSS

Datensatz	Ursprung des Datensatzes
SpraVesT_1_[Datum].csv	Levumi-Datenexport
SpraVesT1_wide.csv	Datensatz umgewandelt ins wide Format
SpraVesT1_wide_clean.csv	Datensatz umgewandelt ins wide Format und von Duplikaten bereinigt
Spra- VesT1_wide_clean_info	Datensatz umgewandelt ins wide Format, von Duplikaten bereinigt und mit den verarbeiteten Zusatzinformationen
SpraVesT1_long_clean_info	Datensatz von Duplikaten bereinigt, mit den verarbeiteten Zusatzinformationen, umgewandelt ins long-Format

3 Arbeit mit SPSS

Die in R generierten Datensätze werden für die Datenauswertung in SPSS eingesetzt. Die Entscheidung, welcher Datensatz die benötigten Informationen für die Analyse enthält, basiert auf den jeweiligen statistischen Analysen.

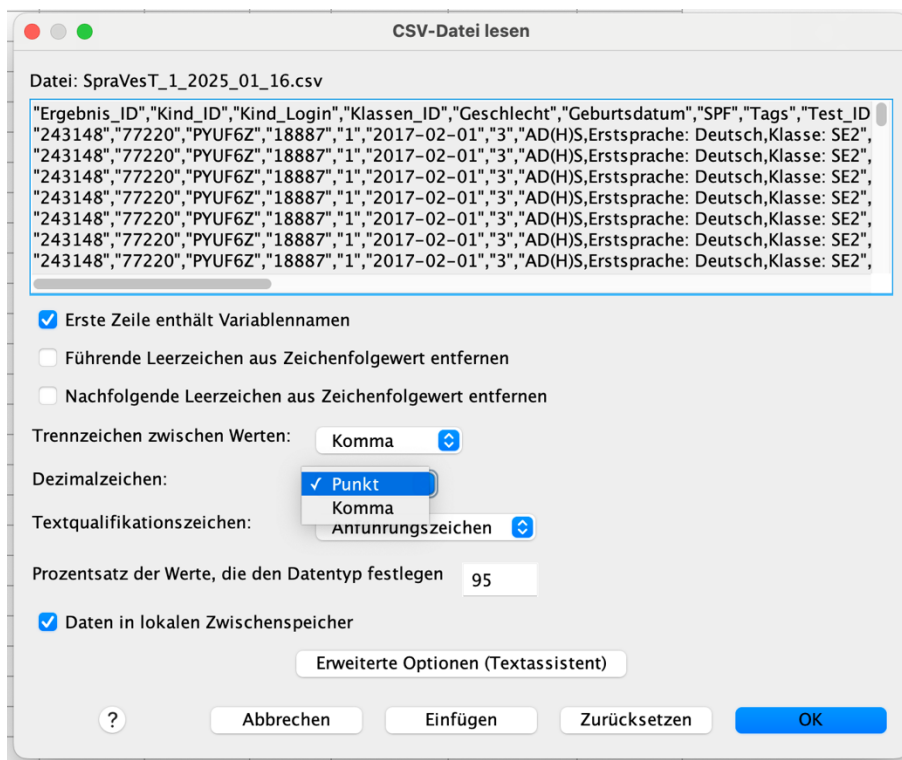
3.1 Datenimport in SPSS

Um einen .csv-Datensatz in SPSS zu importieren, gibt es im Menü die Option

Daten importieren > CSV-Daten.

Nach Auswahl der entsprechenden Datei erscheint der **Textimport-Assistent**, der die Anpassung der Import-Optionen ermöglicht (Abbildung 8). Die Werte in einer .csv-Datei sind üblicherweise durch Kommata getrennt. Die Option *Getrennt* sowie das Trennzeichen *Komma* werden festgelegt. Wenn die erste Zeile der .csv-Datei mit Variablennamen versehen ist, muss die Option *Erste Zeile enthält Variablennamen* aktiviert sein. Nach der Überprüfung und Anpassung der Einstellungen wird der Import der Daten in SPSS bestätigt.

Abbildung 8
SPSS Textimport-Assistent

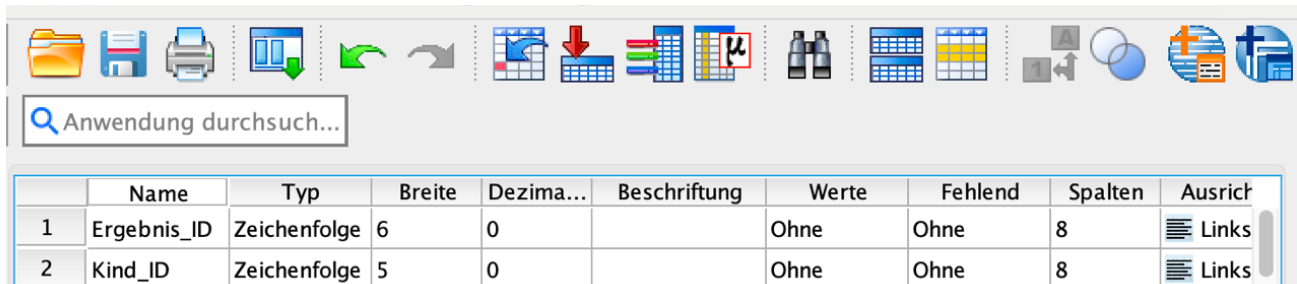


3.2 Die drei Fenster von SPSS

Die drei Hauptfenster in SPSS – der **Dateneditor**, das **Syntaxfenster** und das **Ausgabefenster** – spielen eine zentrale Rolle bei der Arbeit mit SPSS und ermöglichen eine effiziente Datenanalyse.

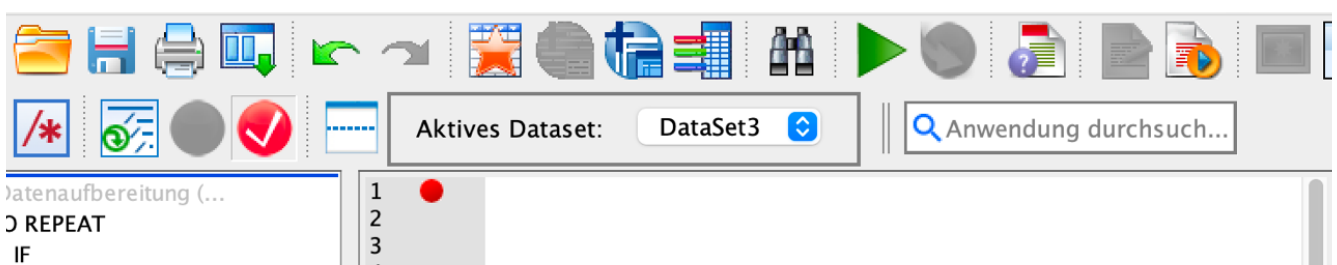
Der **Dateneditor** dient der Eingabe und Bearbeitung der Daten (Abbildung 9). Im Dateneditor sind die Daten in verschiedenen Ansichten verfügbar (siehe Kapitel 3.3).

Abbildung 9
SPSS Dateneditor



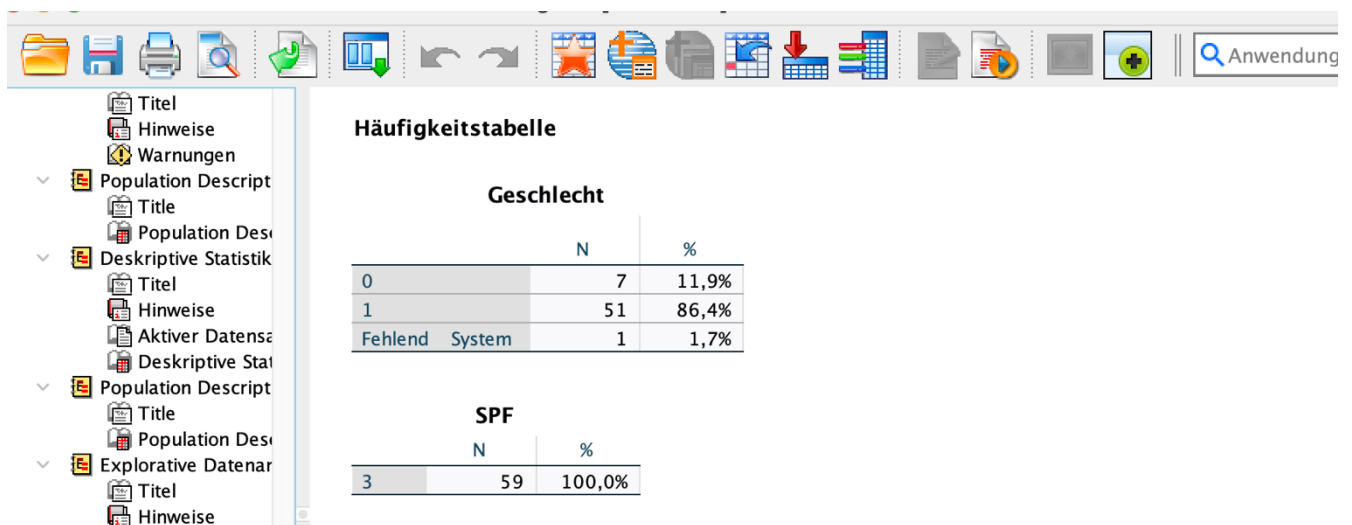
Das **Syntaxfenster** ist der Bereich, in dem SPSS-Befehle (Syntax) manuell geschrieben und ausgeführt werden (Abbildung 10). Hier werden Befehle und Transformationen eingegeben, um die Daten zu analysieren oder zu modifizieren. Der Vorteil der Verwendung von Syntax liegt in der Möglichkeit, wiederholbare Aufgaben zu automatisieren und zu speichern, was besonders dann von Nutzen ist, wenn häufig dieselben Analysen durchgeführt werden müssen oder wenn der Arbeitsprozess für die spätere Verwendung dokumentiert werden soll. Zudem erlaubt die Syntax eine präzisere Kontrolle über die Durchführung der Analysen, ohne auf das Menü zugreifen zu müssen.

Abbildung 10
SPSS Syntaxfenster



Im **Ausgabefenster** werden die Ergebnisse der ausgeführten Analysen angezeigt (Abbildung 11). Nachdem ein Befehl ausgeführt wurde, erscheinen die Ergebnisse (z. B. Statistiken, Tabellen oder Diagramme) in diesem Fenster. Es ermöglicht die Weiterbearbeitung, das Speichern oder den Export der Ergebnisse in verschiedene Formate, um sie später zu nutzen oder zu präsentieren. Zudem werden hier mögliche Warnungen oder Fehlermeldungen angezeigt, die bei der Fehlerbehebung und den Analysen helfen.

Abbildung 11
SPSS Ausgabefenster



3.3 Die Ansichten im Dateneditor von SPSS

Nach erfolgreichem Import des Datensatzes in SPSS gibt es drei zentrale Ansichten, die für die Arbeit mit Datensätzen von Bedeutung sind: die **Übersichtsansicht**, die **Datenansicht**, und die **Variablenansicht** (Abbildung 12).

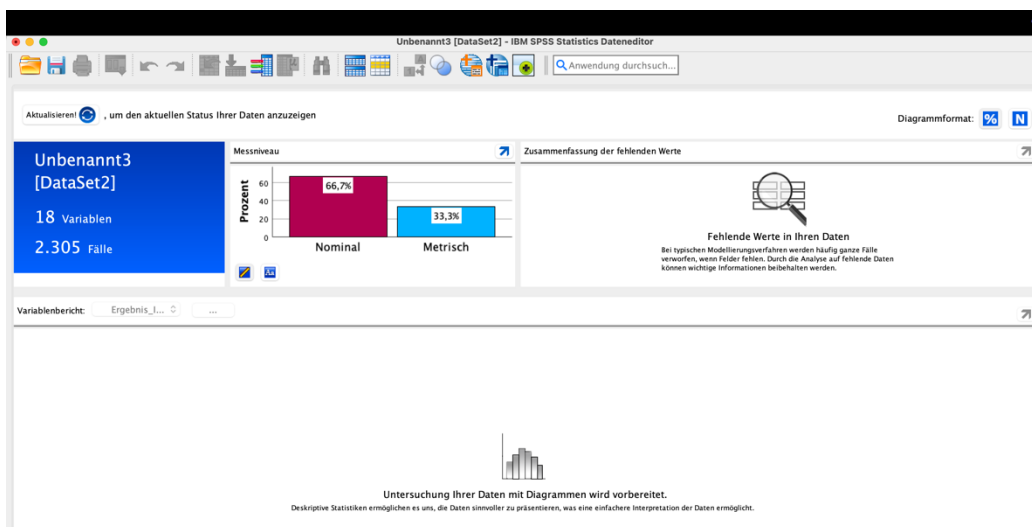
Abbildung 12
SPSS Ansichten



Die **Übersichtsansicht** (ab Version 29) bietet eine zusammenfassende Darstellung des gesamten Datensatzes (Abbildung 13). Hier lässt sich auf einen Blick die Anzahl der Variablen und Beobachtungen sowie allgemeine statistische Kennzahlen wie Mittelwerte,

Standardabweichungen und Verteilungen erkennen. Diese Ansicht eignet sich, um einen ersten Überblick über den Datensatz zu erhalten.

Abbildung 13
SPSS Übersichtsansicht



In der **Datenansicht** werden die untersuchten Objekte, wie beispielsweise getestete Kinder, in Zeilen dargestellt (Abbildung 14). Jede Zeile zeigt dabei die konkreten Werte für ein bestimmtes Objekt, und diese Werte sind den Spalten zugeordnet, die die verschiedenen Variablen repräsentieren, wie etwa Alter, Geschlecht oder Testscores. Diese Ansicht entspricht im Wesentlichen der Darstellung von Daten in Excel.

Abbildung 14
SPSS Datenansicht

	Ergebnis_ID	Kind_ID	Kind_Logi_n	Klassen_ID	Geschlecht	Geburtsdatum	SPF	Tags	Test_ID	Testdatum	Testwoche	ite
1	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I2
2	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I3
3	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I4
4	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I5
5	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I6
6	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I7
7	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I8
8	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I9
9	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I10
10	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I11
11	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I12
12	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I13
13	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I14
14	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I15
15	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I16
16	243148	77220	PYUF6Z	18887	1	2017-02-01	3	AD(H)S,Erstsprache: Deutsch,Klass...	359	2024-12-18 11:07:56	2024-12-16	I17

In der **Variablenansicht** hingegen werden die Variablen definiert (Abbildung 15). Hier kann der Benutzer die Eigenschaften jeder Variablen festlegen, zum Beispiel den Variablennamen, den Datentyp, sowie Wertebezeichner für kategoriale Variablen (z. B. für Geschlecht: 0 = divers, 1 = männlich, 2 = weiblich). Diese Ansicht ermöglicht es, die Struktur und den Aufbau der Variablen zu überprüfen und gegebenenfalls anzupassen, was für die korrekte Durchführung statistischer Analysen entscheidend ist. Das ist wichtig, da die Eigenschaften einer Variablen nicht automatisch aus dem Originaldatensatz übernommen werden. Sie müssen in der Variablenansicht manuell eingegeben werden.

Abbildung 15
SPSS Variablenansicht

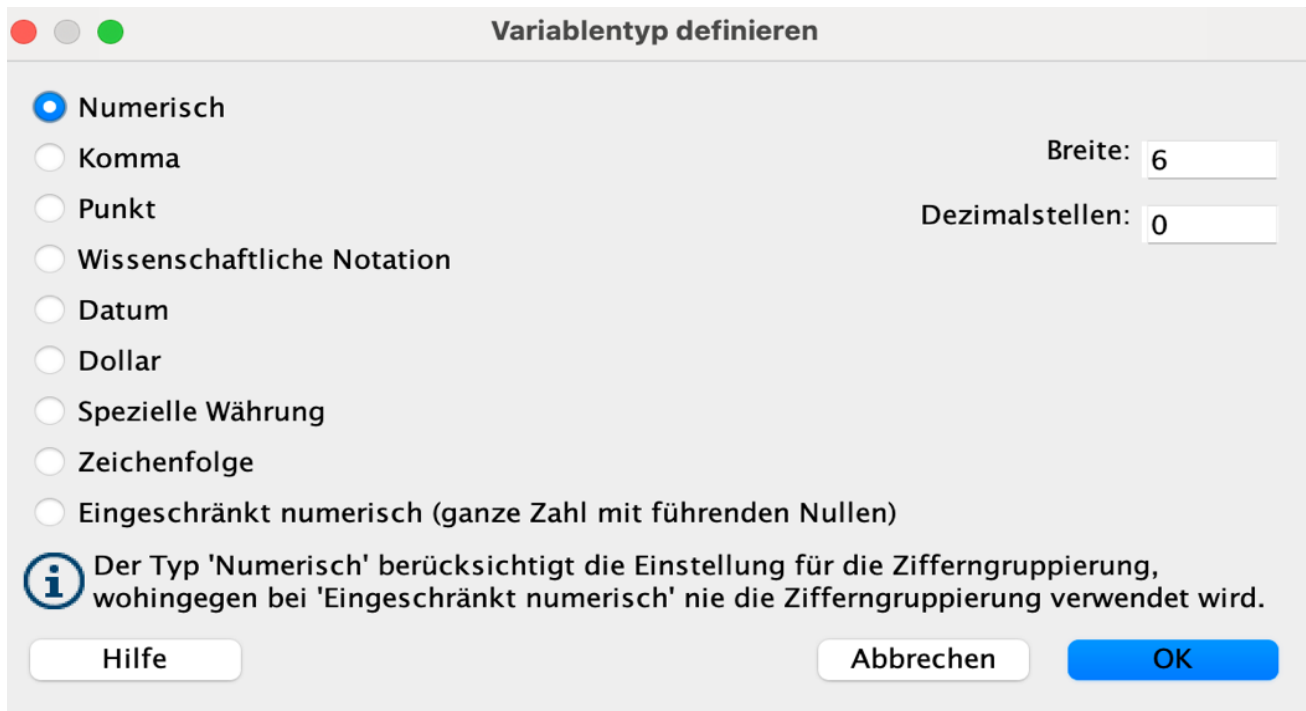
	Name	Typ	Breite	Dezima...	Beschriftung	Werte	Fehlend	Spalten	Ausrichtung	Messniveau	Rolle
1	Ergebnis_ID	Numerisch	6	0		Ohne	Ohne	8	Rechts	Metrisch	Eingabe
2	Kind_ID	Numerisch	5	0		Ohne	Ohne	8	Rechts	Metrisch	Eingabe
3	Kind_Login	Zeichenfolge	9	0		Ohne	Ohne	9	Links	Nominal	Eingabe
4	Klassen_ID	Numerisch	5	0		Ohne	Ohne	8	Rechts	Metrisch	Eingabe
5	Geschlecht	Numerisch	1	0		Ohne	Ohne	8	Rechts	Nominal	Eingabe
6	Geburtsda...	Zeichenfolge	10	0		Ohne	Ohne	10	Links	Nominal	Eingabe
7	SPF	Numerisch	1	0		Ohne	Ohne	8	Rechts	Nominal	Eingabe
8	Tags	Zeichenfolge	172	0		Ohne	Ohne	26	Links	Nominal	Eingabe
9	Test_ID	Numerisch	3	0		Ohne	Ohne	8	Rechts	Metrisch	Eingabe
10	Testdatum	Datum	19	0		Ohne	Ohne	21	Rechts	Metrisch	Eingabe
11	Testwoche	Zeichenfolge	10	0		Ohne	Ohne	10	Links	Nominal	Eingabe
12	item	Zeichenfolge	3	0		Ohne	Ohne	3	Links	Nominal	Eingabe
13	time	Numerisch	6	0		Ohne	Ohne	8	Rechts	Metrisch	Eingabe
14	group	Numerisch	1	0		Ohne	Ohne	8	Rechts	Nominal	Eingabe
15	answer	Zeichenfolge	17	0		Ohne	Ohne	17	Links	Nominal	Eingabe

3.4 Umgang mit Variablen in SPSS

Eine Variable in SPSS besitzt einen **Namen** sowie weitere Eigenschaften, die individuell festgelegt werden und in der Variablenansicht definiert werden. Der Variablenname sollte möglichst kurz und prägnant gewählt werden, um eine eindeutige Identifikation zu gewährleisten. Beispielsweise kann eine Variable zur Erfassung des Alters einfach *Alter* genannt werden. Hierzu wird in der entsprechenden Zelle der Name händisch eingegeben und so verändert. Auch neue Variablen können so ergänzt werden.

Neben dem Namen muss auch der **Typ der Variable** bestimmt werden (Abbildung 16). Dieser gibt an, welche Art von Daten die Variable enthält. Gängige Typen sind *Numerisch* für Zahlenwerte, *Datum* für Zeitangaben oder *Zeichenfolge* für Textdaten.

Abbildung 16
SPSS Variablentyp definieren



Ändern des Variablentyps

Hier ist es wichtig, inhaltlich zu denken: Welche Werte liegen vor und wie sollen diese interpretiert werden?

Beim Import der Daten werden sie durch SPSS automatisch einem Variablentyp zugewiesen. Bestimmt wird das durch die Werte, die in der jeweiligen Spalte identifiziert werden. Bei vollständig numerischen Werten, wird der Dateityp auf *Numerisch*, bei textbasierten Werten (z. B. NA) auf *Zeichenfolge* festgelegt.

In dem SpraVesT-Datensatz gibt es auf Grund des Speedtests eine hohe Anzahl nicht beantworteter Items. Zeilen mit den entsprechenden NA-Werten, werden so oft automatisch zu „Zeichenfolgen“, obwohl die Variable inhaltlich eine numerische Variable ist. Wenn der Dateityp anschließend angepasst wird, entstehen Schwierigkeiten mit

der Definition von **fehlenden Werten**. Als numerische Variable werden „NA“ Ausprägungen nicht mehr zugelassen. Stattdessen sind in den entsprechenden Spalten nur Punkte oder leere Felder zu sehen. Hier müssen **Werte** definiert werden, die numerisch sind, um sie in den Analysen als fehlende Werte berücksichtigen zu können. In SPSS müssen **fehlende Werte** explizit definiert werden, beispielsweise durch die Zahl „-999“ (Abbildung 17). Möglich ist das wieder in der Variablenansicht unter der Spalte Werte zu jeder Variablen.

Abbildung 17
SPSS Fehlende Werte festlegen

The image shows the 'Fehlende Werte' dialog box in SPSS. It has three radio button options: 'Keine fehlenden Werte' (selected), 'Einzelne fehlende Werte' (highlighted with a red box), and 'Spannweite plus ein optional diskreter fehlender Wert'. Under 'Einzelne fehlende Werte', there are three input fields, with the first one containing '-999'. Under the third option, there are input fields for 'Kleinster Wert', 'Größter Wert', and 'Diskreter Wert'. At the bottom, there are buttons for '?', 'Abbrechen', and 'OK'.

Dieses Vorgehen ermöglicht eine systematische Erfassung und spätere Behandlung dieser Werte, um Verzerrungen in der Analyse zu vermeiden. Allerdings müssen diese Werte dann auch im Datensatz zu finden sein. Das kann händisch für jeden einzelnen Eintrag erfolgen, ist aber fehleranfällig und zeitraubend. Dazu wird in der Datenansicht für jede:n einzelnen Proband:in der Wert eingegeben. Wichtig ist, dass die Variable zuvor bereits als numerische Variable festgelegt wurde.

Das gilt auch für die praktikablere Variante im Syntaxfenster. Zur automatisierten Umwandlung der NA Werte, werden im Syntaxfenster alle NA-Werte einer Variablen gesucht und durch -999 ersetzt (Abbildung 18). Um das nicht für jede Variable einzeln machen zu müssen, werden die einzelnen Variablen miteinander verknüpft. SPSS braucht

dabei die Info, welche (1) Variablen verändert werden sollen, (2) wonach gesucht werden soll, (3) und was dann mit den identifizierten Fällen passieren soll.

Transformation von NA-Werten

Ziel: alle leeren Felder (NA-Werte) der Variablen I2_result bis I81_result sollen mit -999 aufgefüllt werden.

Hierzu werden zunächst die entsprechenden Variablen ausgewählt

```
var = I2_result TO I81_result.
```

Da mehrere Variablen gleichzeitig verändert werden, wird diese Auswahl in einen REPEAT-Befehl eingebettet

```
DO REPEAT var = I2_result TO I81_result.
```

```
END REPEAT.
```

Nun fehlt noch eine Anweisung, was mit den Daten passieren soll: „Wenn ein Feld leer ist, soll der Wert -999 vergeben werden:“

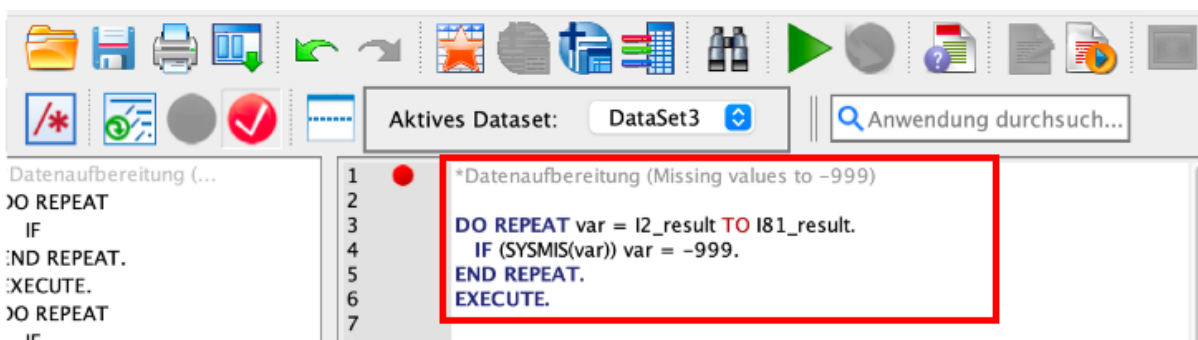
```
DO REPEAT var = I2_result TO I81_result.
```

```
IF (SYSMIS(var)) var = -999.
```

```
END REPEAT.
```

Abbildung 18

SPSS-Syntaxfenster mit Code zur Transformation von NA-Werten

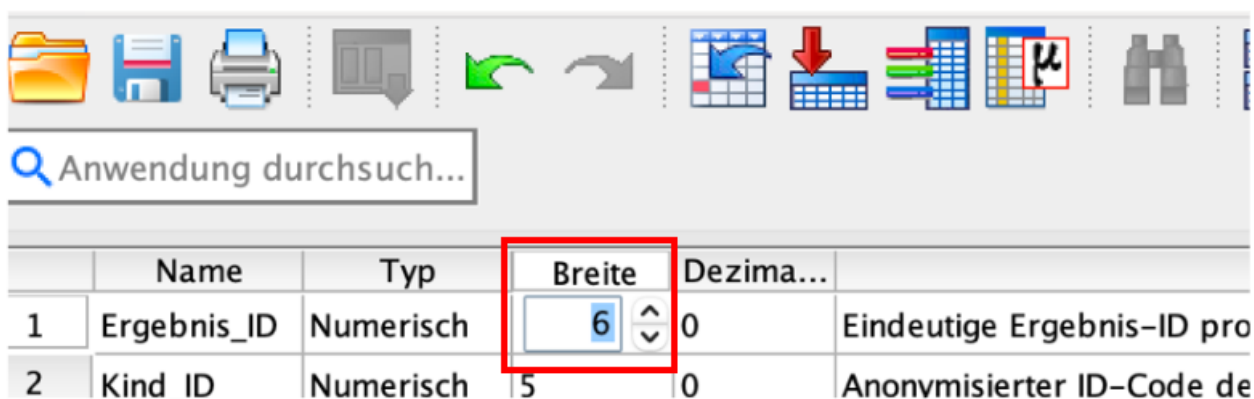


Zusätzlich zur Festlegung des Typs muss die **Breite der Spalte** definiert werden, die bestimmt, wie viele Zeichen oder Stellen ein Variablenwert maximal umfassen kann

(Abbildung 19). Es ist ratsam, eine ausreichend große Breite zu wählen, um Datenverlust zu vermeiden. Eng damit verbunden ist die Festlegung der Dezimalstellen, die insbesondere bei numerischen Variablen von Bedeutung ist, um die gewünschte Genauigkeit sicherzustellen.

Festlegung der Spaltenbreite

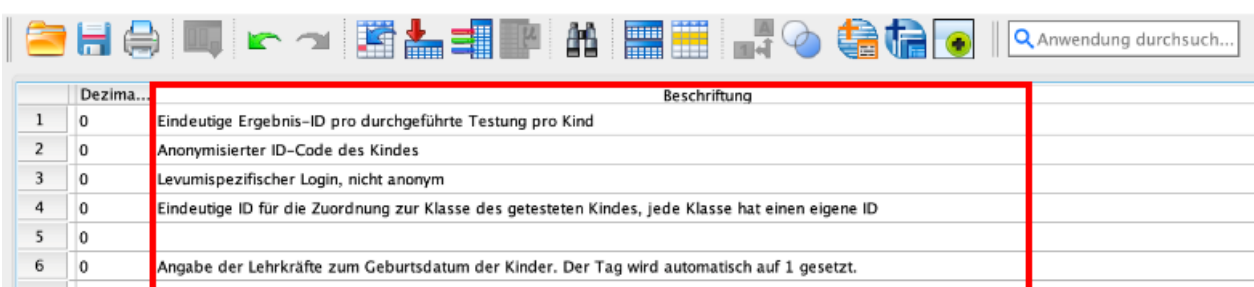
Abbildung 19
SPSS Breite festlegen



Neben diesen technischen Spezifikationen sollte eine aussagekräftige **Beschriftung** der Variablen erfolgen (Abbildung 20). Diese dient dazu, die Bedeutung der Variable nachvollziehbar darzustellen, sodass auch Personen ohne Vorkenntnisse die Daten korrekt interpretieren können. Beispiele für solche Beschriftungen sind „Kind-ID“, „Frage des Fragebogens“ oder „Itemname“.

Festlegung der Variablenbeschriftung

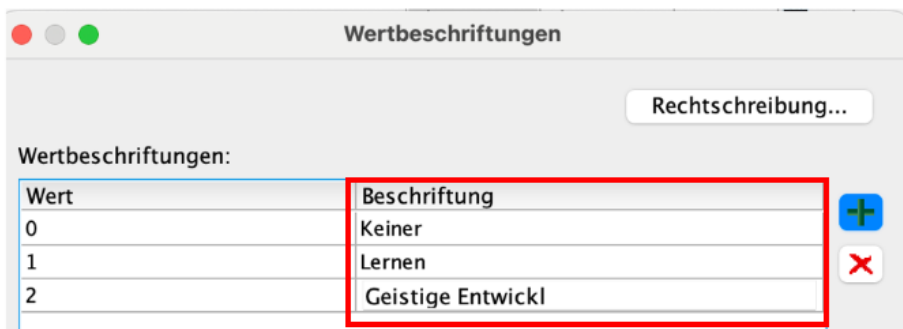
Abbildung 20
SPSS Beschriftungen festlegen



Für nominale und ordinale Variablen ist zudem die **Festlegung von Wertelabels** notwendig (Abbildung 21). Hierbei werden numerische Codes bestimmten Kategorien zugeordnet, etwa 0 = weiblich, 1 = männlich und 2 = divers. Dies erleichtert die spätere Datenanalyse und verbessert die Übersichtlichkeit der Datensätze.

Einfügen der Variablenwerte

Abbildung 21
SPSS Wertbeschriftungen



Die aus der Tags-Variable generierten Variablen mit den weiteren Informationen zu den Proband:innen sind ebenfalls nominale Variablen. In SPSS werden sie als Zeichenfolgen und mit den Ausprägungen TRUE oder FALSE importiert. So lässt sich damit allerdings nicht rechnen. Neben der Änderung des Variablentyps müssen also auch die Werteausprägungen und Beschriftungen verändert werden. Hier hilft wieder das Syntaxfenster.

Transformation von Text zu numerischen Werten

Ziel: Ausprägung der Variable Mig wenn TRUE in 1 und wenn FALSE in 0 umwandeln

„Wenn ein Feld die Ausprägung „TRUE“ ist, soll der Wert 1 vergeben werden“

```
IF (Mig = "TRUE") Mig = "1".
```

„Wenn ein Feld die Ausprägung „FALSE“ ist, soll der Wert 0 vergeben werden:“

```
IF (Mig = "FALSE") Mig = "0".
```

Ausführen der Befehle mit EXECUTE:

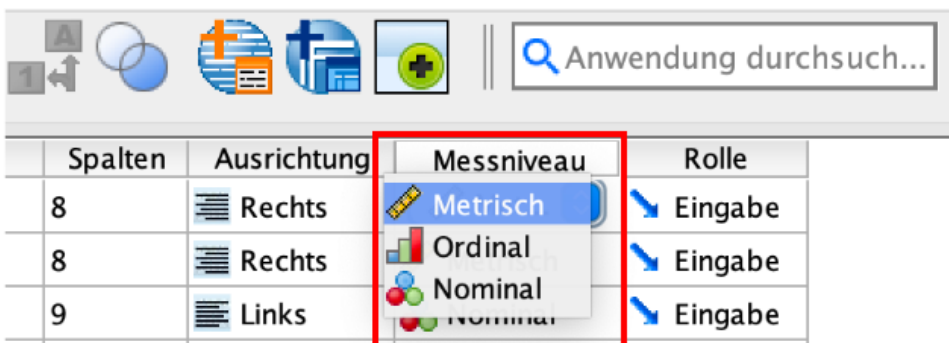
Schließlich muss für jede Variable das **Messniveau** bestimmt werden, welches die Art der statistischen Analyse beeinflusst. Es gibt drei grundlegende Messniveaus: nominal, ordinal und metrisch.

- **Nominalskalierte Variablen** dienen der Kategorisierung ohne eine inhaltliche Rangfolge, beispielsweise Geschlecht oder Nationalität.
- **Ordinalskalierte Variablen** hingegen ermöglichen eine Rangordnung, jedoch ohne eine gleichbleibende Differenz zwischen den Ausprägungen, wie etwa bei Schulnoten.
- **Metrische Variablen** zeichnen sich dadurch aus, dass Abstandsvergleiche zwischen den Werten möglich sind, etwa bei der Erfassung des Alters in Jahren.

Die korrekte Festlegung des Messniveaus ist essenziell, da sie die Wahl der geeigneten statistischen Verfahren bestimmt (Sedlmeier & Renkewitz, 2018). In SPSS ist dies per Drop-down Menü möglich (Abbildung 22).

Festlegung des Messniveaus

Abbildung 22
SPSS Messniveau festlegen



Die korrekte Aufbereitung der Daten ist eine notwendige Bedingung für die erfolgreiche Datenanalyse und sollte daher sorgfältig und vollständig erfolgen.

3.5 Variablen berechnen in SPSS

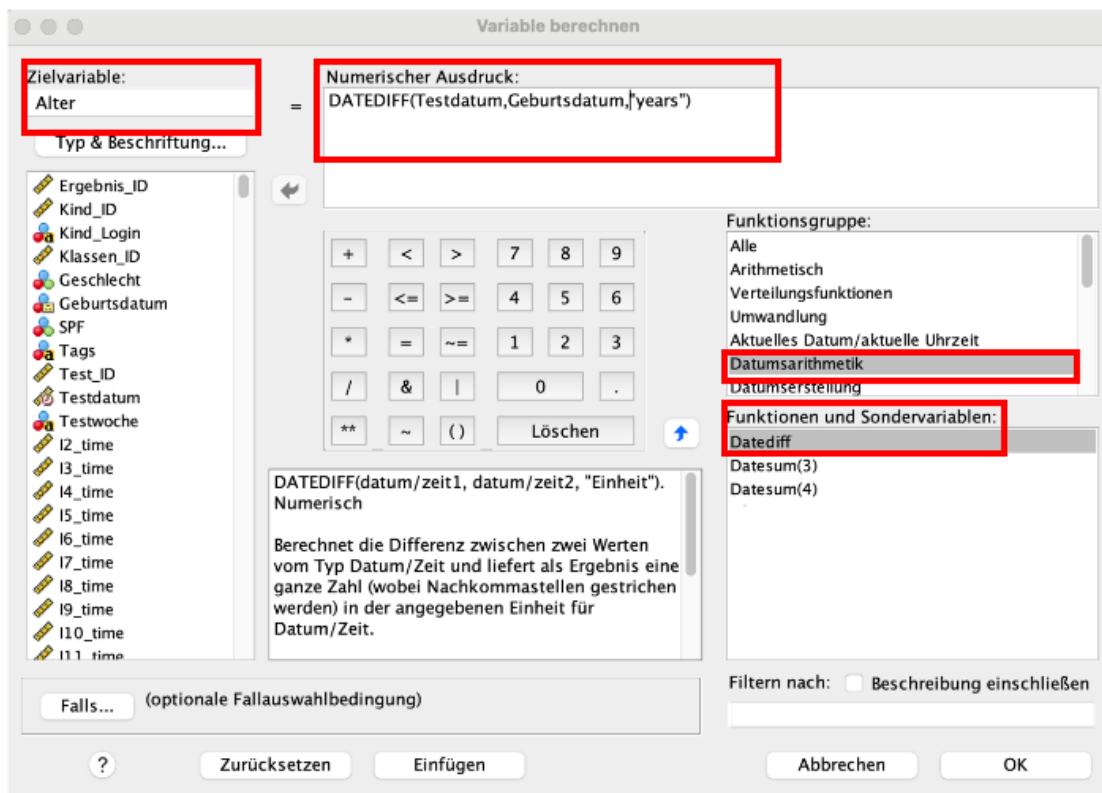
In manchen Fällen müssen Variablen völlig neu berechnet werden (z. B. das Alter aus der Differenz von Geburts- und Testdatum). Das ist unter

Transformieren > Variable berechnen

möglich.

Im Feld *Zielvariable* wird der neue Name der Variable festgelegt (hier: Alter). Anschließend wird unter *Numerischer Ausdruck* festgelegt, wie die entsprechende Variable berechnet werden soll. Für die Berechnung der Variable Alter werden zunächst zwei als Datumsvariablen definierten Variablen benötigt. Im SpraVesT werden das Geburtsdatum und das Testdatum erhoben. Das Testdatum wird durch SPSS als Datum erkannt, das Geburtsdatum wird als Zeichenfolge definiert und muss zunächst umgewandelt werden. Hierzu muss der Variablentyp auf Datum (z. B. jjjj/mm/tt) geändert werden. Anschließend kann die Funktion Datediff ausgeführt werden (Abbildung 23).

Abbildung 23
SPSS Variable berechnen



3.6 SPSS-Datensatz speichern

Um den SPSS-Output zu speichern, gibt es verschiedene Möglichkeiten, je nachdem, ob die Daten als SPSS-Datei oder in einem anderen Format, beispielsweise Excel, gesichert werden sollen.

Der SPSS-Datensatz kann als SPSS-Datei gespeichert werden, indem im Menü „Datei“ die Option „Speichern unter“ gewählt und anschließend die Datei gesichert wird. Es ist wichtig, nicht nur den Datensatz selbst, sondern auch die zugehörige Syntax und die Ausgabe zu speichern. Nur so ist gewährleistet, dass bei einer späteren Weiterbearbeitung alle relevanten Informationen erhalten bleiben. Beim erneuten Öffnen der Dateien sollten daher stets alle drei Komponenten – Datensatz, Syntax und Ausgabe – gleichzeitig geladen werden, da andernfalls eine neue Syntax und eine neue Ausgabe generiert werden, was zu Inkonsistenzen führen kann.

Speichern des Datensatzes als .sas-Datei

Soll der SPSS-Datensatz in einem anderen Format, beispielsweise als Excel-Datei, gespeichert werden, erfolgt dies über den Export. Dazu wird im Menü „Datei“ die Option „Exportieren“ gewählt, anschließend das Format „Excel“ ausgewählt und die Datei gespeichert. Hierbei ist darauf zu achten, dass der Originaldatensatz nicht überschrieben wird. Daher sollte beim Speichern ein eindeutiger neuer Name vergeben werden, um eine Verwechslung oder den Verlust der ursprünglichen SPSS-Daten zu vermeiden.

Speichern des Datensatzes als .xlsx-Datei

4 Statistische Analysen in SPSS mit Levumi-Daten

Für die statistische Analyse von Datensätzen stehen zahlreiche Methoden zur Verfügung, die sich grundsätzlich in zwei Hauptbereiche unterteilen lassen: die **deskriptive Statistik** und die **inferenzstatistische Analyse**.

Die **deskriptive Statistik** umfasst Verfahren zur Zusammenfassung, Darstellung und Beschreibung von Daten. Hierbei werden Kennwerte wie Mittelwert, Median, Standardabweichung oder Häufigkeiten berechnet, um zentrale Tendenzen und Variabilität innerhalb eines Datensatzes zu erfassen. Zudem kommen visuelle Darstellungen wie Histogramme, Boxplots oder Streudiagramme zum Einsatz, um Muster und Strukturen in den Daten zu identifizieren (Döring & Bortz, 2016).

Die **Inferenzstatistik** hingegen ermöglicht es, auf Basis einer Stichprobe Rückschlüsse auf eine zugrunde liegende Population zu ziehen. Mithilfe statistischer Tests, Konfidenzintervalle oder Regressionsanalysen wird geprüft, ob beobachtete Effekte oder Zusammenhänge zufällig entstanden sind oder auf allgemeine Gesetzmäßigkeiten hindeuten. Typische Verfahren der Inferenzstatistik sind der t-Test, die Varianzanalyse (ANOVA) oder die lineare Regression. Während die deskriptive Statistik sich darauf konzentriert, bestehende Daten zu beschreiben, geht die inferenzstatistische Analyse darüber hinaus und erlaubt es, Hypothesen zu testen und kausale Zusammenhänge zu untersuchen (ebd.).

4.1 Deskriptive Statistik in SPSS

Die deskriptive Statistik in SPSS ermöglicht eine umfassende Analyse und Darstellung von Datensätzen, um grundlegende Merkmale und Verteilungen sichtbar zu machen. Dazu stehen verschiedene Methoden zur Verfügung, mit denen zentrale Kennwerte berechnet und grafische Darstellungen erstellt werden können.

Zur Berechnung grundlegender statistischer Kennwerte wie Mittelwert, Median, Standardabweichung sowie Minimal- und Maximalwert bietet SPSS mehrere Optionen im Menü

Analysieren > Deskriptive Statistiken.

Hierzu eignet sich der Datensatz `SpraVesT1_wide_clean_info.csv`, da er durch das `wide`-Format leichtere Rückschlüsse erlaubt.

Importieren des richtigen Datensatzes

`SpraVesT1_wide_clean_info.csv`

4.1.1 Lagemaße in SPSS

Die **Lagemaße** sind statistische Kennzahlen, die dazu dienen, den Mittelpunkt oder den Durchschnitt einer Verteilung zu beschreiben. Sie dienen dazu, die Daten zusammenzufassen und eine Vorstellung von der typischen oder durchschnittlichen Ausprägung einer Variablen zu erhalten. Zu den wichtigsten Kennzahlen gehören der Modus, der Median und der Mittelwert (Sedlmeier & Renkewitz, 2018).

Der **Modus** ist der Wert, der in einem Datensatz am häufigsten vorkommt. Er eignet sich besonders gut für kategoriale oder nominalskalierte Daten (ebd.). Der **Median** ist der Wert, der in der Mitte einer geordneten Reihe von Daten liegt, wenn die Daten der Größe nach sortiert sind (ebd.). Der **Mittelwert** stellt den Durchschnittswert der Daten dar und wird berechnet, indem alle Werte summiert und durch die Anzahl der Werte geteilt werden (ebd.).

In SPSS können diese Maße in wenigen Schritten mit

Analysieren > Deskriptive Statistiken > Häufigkeiten

berechnet werden. Nachdem die entsprechenden Variablen ausgewählt wurden, zeigt SPSS eine Tabelle an, die Modus, Median und Mittelwert enthält. Diese Tabelle bietet eine schnelle Übersicht über die zentralen Tendenzen der ausgewählten Variablen und ermöglicht eine erste statistische Einschätzung der Daten (Abbildung 24).

Abbildung 24
SPSS-Output Lagemaße

		Statistiken													
		I2_result	I3_result	I4_result	I5_result	I6_result	I7_result	I8_result	I9_result	I10_result	I11_result	I12_result	I13_result	I14_result	
N	Gültig	59	59	59	59	59	59	59	58	58	58	57	57	57	
	Fehlend	0	0	0	0	0	0	0	1	1	1	2	2	2	
Mittelwert		,61	,51	,58	,78	,47	,47	,71	,67	,55	,45	,72	,58	,68	
Median		1,00	1,00	1,00	1,00	,00	,00	1,00	1,00	1,00	,00	1,00	1,00	1,00	
Modus		1	1	1	1	0	0	1	1	1	0	1	1	1	

a. Mehrere Modi vorhanden. Der kleinste Wert wird angezeigt.

Lagemaße berechnen

Analysieren > Deskriptive Statistiken > Häufigkeiten

Ein spezifisches Lagemaß im SpraVesT ist die **Itemschwierigkeit**. Sie gibt an, wie schwer oder leicht ein Testitem für die Proband:innen im Durchschnitt zu lösen ist. Sie wird berechnet, indem der Anteil der richtigen Antworten für jedes Item ermittelt wird. Ein Wert nahe 0 zeigt an, dass das Item für die meisten Proband:innen sehr schwierig war, während ein Wert nahe 1 bedeutet, dass das Item für die Mehrheit der Proband:innen einfach war.

Zur **Berechnung der Itemschwierigkeit** in SPSS sind die Daten im wide-Format erforderlich. Wichtig ist, dass es pro Item eine eigene Spalte (Variable) gibt (z. B. I2_result), in der der Wert 1 für eine richtige Antwort und 0 für eine falsche Antwort steht. Jede Zeile stellt dabei eine Proband:in dar, der:die die Items beantwortet hat. Besonders wichtig ist, dass die fehlenden Werte als solche definiert sind.

Da die Berechnung der **Itemschwierigkeit** auf der Berechnung der Mittelwerte basiert, erfolgt diese erneut über

Analysieren > Deskriptive Statistiken > Häufigkeiten.

Als interessierende Variablen werden alle Item-Variablen ausgewählt (I2_result bis I81_result), da diese die Antworten der Teilnehmer enthalten. Unter Statistiken werden die Mittelwerte ausgewählt.

Der ausgegebene Mittelwert jedes Items stellt die Itemschwierigkeit dar. Ein Wert von beispielsweise 0,75 bedeutet, dass 75 % der Teilnehmer das jeweilige Item richtig beantwortet haben, was auf ein relativ leichtes Item hinweist. Ein Mittelwert von 0,25 deutet darauf hin, dass nur 25 % der Teilnehmer das Item richtig beantwortet haben, was das Item als schwieriger klassifizieren würde (Abbildung 25).

Abbildung 25
SPSS-Output zur Itemschwierigkeit

Häufigkeiten

		Statistiken												
		I2_result	I3_result	I4_result	I5_result	I6_result	I7_result	I8_result	I9_result	I10_result	I11_result	I12_result	I13_result	I14_result
N	Gültig	59	59	59	59	59	59	59	58	58	58	57	57	57
	Fehlend	0	0	0	0	0	0	0	1	1	1	2	2	2
Mittelwert		,61	,51	,58	,78	,47	,47	,71	,67	,55	,45	,72	,58	,68

Häufigkeitstabelle

I2_result

	N	%
0	23	39,0%
1	36	61,0%

I3_result

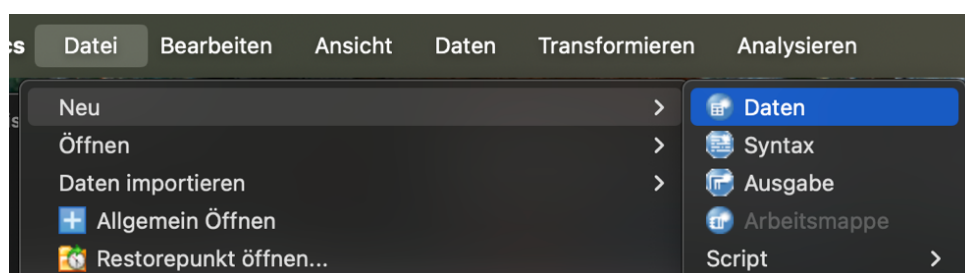
	N	%
0	29	49,2%
1	30	50,8%

Bei der Interpretation sind die fehlenden Werte, die durch die Testung als Speedtest entstehen, zu berücksichtigen. Diese werden im Output pro Item direkt mit angegeben.

Itemschwierigkeit berechnen
Analysieren > Deskriptive Statistiken > Häufigkeiten

Für manche Analysen (z. B. den Vergleich der Itemschwierigkeiten zweier Stichproben) kann es sinnvoll sein, die Itemschwierigkeit in einem neuen Datensatz zu speichern, um später darauf zurückgreifen zu können. Dieser kann über die Syntax oder händisch erstellt werden. Hier wird das händische Vorgehen geschildert und zunächst ein neuer leerer Datensatz in SPSS erstellt (Abbildung 26). Das Erstellen eines neuen Datensatzes ist hierbei notwendig, da die Mittelwerte nicht in die bestehende Datenstruktur (wide-Format & Personenbezug) hereinpasse.

Abbildung 26
SPSS neuen Datensatz erstellen



Die Daten können anschließend im wide- oder long-Format eingetragen werden. Hierzu werden in der Variablenansicht neue Variablen erstellt (Abbildung 27) und diese anschließend in der Datenansicht (Abbildung 28) mit Werten gefüllt.

Abbildung 27a
SPSS Eingabe neuer Variablen (wide-Format)

	Name	Typ	Breite	Dezima...	Beschriftung	Werte	Fehlend	Spalten	Ausrichtung
1	I_I2	Numerisch	8	2		Ohne	Ohne	8	Rechts
2	I_I3	Numerisch	8	2		Ohne	Ohne	8	Rechts
3	I_I4	Numerisch	8	2		Ohne	Ohne	8	Rechts
4	I_I5	Numerisch	8	2		Ohne	Ohne	8	Rechts
5									
6									

Abbildung 27b
SPSS Eingabe neuer Variablen (long-Format)

	Name	Typ	Breite	Dezima...	Beschriftung	Werte	Fehlend	Spalten	Ausrichtung
1	Item	Zeichenfolge	8	0		Ohne	Ohne	8	Links
2	Itemschwie...	Numerisch	8	2		Ohne	Ohne	8	Rechts
3									
4									

Abbildung 28a
SPSS Eingabe Ausprägungen einzelner Werte (wide-Format)

	I_I2	I_I3	I_I4	I_I5	Var	Var	Var	Var	Var
1	,61	0,	.	.					
2									
3									

Abbildung 28b
SPSS Eingabe Ausprägungen einzelner Werte (long-Format)

	Item	Itemschwierigkeit	Var	Var	Var	Var	Var	Var	Var
1	I2	,61							
2	I3	,51							
3	I4	,58							
4	I5	,78							
5	I6	,47							
6	I7	,47							
7	I8	,71							
8	I9	,67							
9	I10	,55							

Ein weiteres spezifisches Lagemaß für die Gesamtleistung einer:s Proband:in im SpraveT ist der **Summenscore**. Er gibt an, wie viele Items eine Proband:in insgesamt richtig beantwortet hat. Dieser Wert wird berechnet, indem die Anzahl der korrekt beantworteten Items für jede Person summiert wird. Ein hoher Summenscore zeigt an, dass die Proband:in viele Items richtig beantwortet hat, was auf eine gute Leistung im Test hinweist. Ein niedriger Summenscore deutet darauf hin, dass die Proband:in weniger Items korrekt beantwortet hat, was auf eine schlechtere Leistung im Test hinweist.

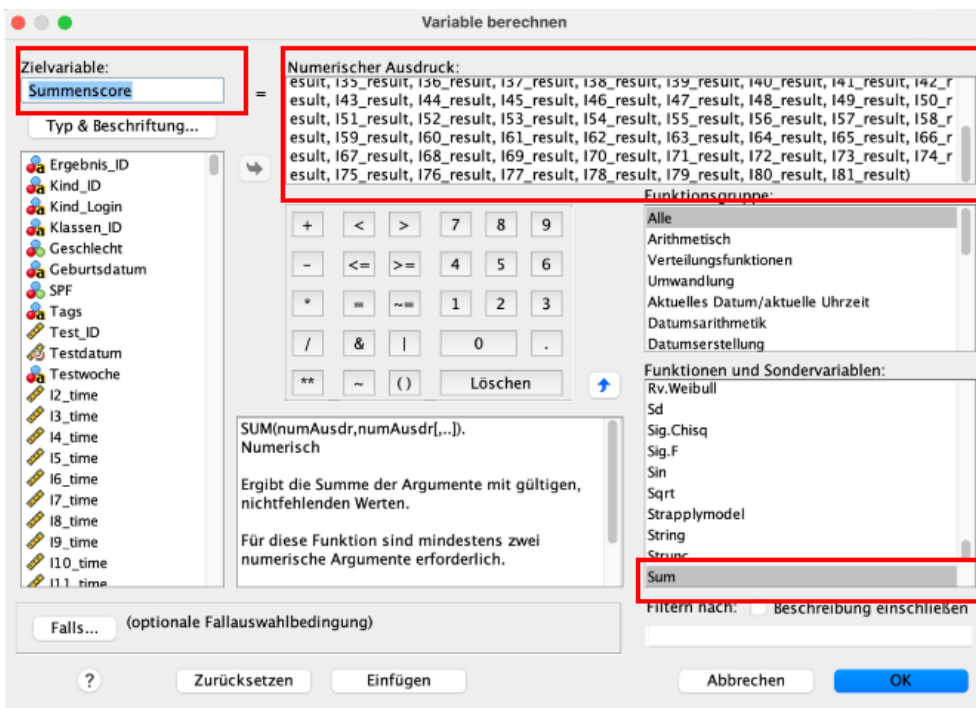
Zur Berechnung des Summenscores in SPSS sind die Daten im wide-Format erforderlich. Es ist wichtig, dass für jedes Item eine eigene Spalte (Variable) existiert, in der der Wert 1 für eine richtige Antwort und 0 für eine falsche Antwort steht (z. B. I2_result). Jede Zeile entspricht einer Proband:in, die die Items beantwortet hat. Zudem müssen fehlende Werte korrekt definiert werden.

Die Berechnung des Summenscores erfolgt durch die Summierung der Werte über alle Item-Variablen. Eine einfache Möglichkeit zur Berechnung des Summenscores in SPSS ist die Funktion

Transformieren > Variablen berechnen.

Dort wird die Summe der richtig gelösten Items pro Person über alle Items gebildet. Der Summenscore gibt also an, wie viele der getesteten Items eine Proband:in korrekt beantwortet hat. Hierzu wird bei der Transformation die Funktion Sum verwendet, deren Ausdruck kommasetrennt die interessierenden Variablen (hier: die _result Variablen) enthält (Abbildung 29).

Abbildung 29
SPSS Eingabe Transformation (Variable berechnen)



Der Summenscore wird für jede:n Proband:in berechnet und als neue Variable ans Ende des Datensatzes angehängen (Abbildung 30).

Abbildung 30
SPSS Datenansicht mit der neuen Variable Summenscore

	LRS	Dysk	Gifted	PerSup	ADHS	Autismus	grade	Summenscore	Var	Var
8	0	0	0	0	0	0	4	21		
9	0	0	0	0	0	0	3	5		

Wenn eine Proband:in 60 Items korrekt beantwortet hat, beträgt der Summenscore 60. Ein hoher Summenscore (z. B. 75 von 80) deutet darauf hin, dass die Proband:in insgesamt gut abgeschnitten hat, während ein niedriger Summenscore (z. B. 35 von 80) auf eine schlechtere Leistung hinweist. In der Interpretation des Summenscores muss einbezogen werden, wie viele der Items in der vorgegebenen Bearbeitungszeit beantwortet wurden (z. B. 25 von 30 hoher Summenscore, 5 von 30 niedriger Summenscore).

Entsprechend kann der Summenscore auch als Mittelwert pro Person ausgegeben werden, der dann angibt, wie viel Prozent der Items der:die Proband:in richtig gelöst hat.

4.1.2 Streuungsmaße in SPSS

Um die Abweichung der einzelnen Werte einer Verteilung von der Tendenz (z. B. dem Mittelwert) anzugeben, werden Streuungsmaße berechnet. Zu den wichtigsten Streuungsmaßen gehören die **Standardabweichung**, die **Varianz**, die **Spannweite** und die **Interquartilsdifferenz** (Sedlmeier & Renkewitz, 2018).

Eine zentrale Kennzahl in diesem Zusammenhang ist die **Standardabweichung**, die angibt, wie stark die einzelnen Werte im Durchschnitt vom Mittelwert abweichen. Eine hohe Standardabweichung weist darauf hin, dass die Werte weit gestreut sind, während eine niedrige Standardabweichung darauf hindeutet, dass die Werte nahe beieinander liegen (ebd.).

Eng mit der Standardabweichung verknüpft ist die **Varianz**, die als Quadrat der Standardabweichung ebenfalls die Streuung der Werte misst. Sie bietet eine alternative Möglichkeit zur Quantifizierung der Datenvariabilität und wird insbesondere in inferenzstatistischen Berechnungen häufig verwendet (ebd.).

Ein weiteres wichtiges Maß zur Beschreibung der Streuung ist die **Spannweite**, die als Differenz zwischen dem höchsten und dem niedrigsten Wert eines Datensatzes berechnet wird. Sie gibt an, über welchen Wertebereich sich die Daten erstrecken, reagiert jedoch empfindlich auf Ausreißer, da sie ausschließlich auf den Extremwerten basiert (ebd.).

Um eine robustere Einschätzung der Streuung zu erhalten, wird oft die **Interquartilsdifferenz (IQR)** herangezogen. Diese beschreibt die Spannweite der mittleren 50 % der

Daten, indem sie die Differenz zwischen dem oberen und unteren Quartil berechnet. Da der IQR weniger anfällig für Ausreißer ist als die Spannweite, eignet er sich besonders gut zur Analyse von Daten mit möglichen Extremwerten (ebd.).

In SPSS werden Standardabweichung, Varianz und Spannweite über

Analysieren > Deskriptive Statistiken > Deskriptive Statistiken für Grundgesamtheit

abgerufen. Nachdem die entsprechenden Variablen ausgewählt wurden, zeigt SPSS eine Tabelle an, die neben dem Mittelwert auch Standardabweichung, Varianz und Spannweite, enthält (Abbildung 31).

Abbildung 31
SPSS-Output Streuungsmaße

			Statistik	Standard Fehler
I2_result	Mittelwert		,62	,081
	95% Konfidenzintervall des Mittelwerts	Untergrenze	,46	
		Obergrenze	,79	
	5% getrimmtes Mittel		,64	
	Median		1,00	
	Varianz		,242	
	Standard Abweichung		,492	
	Minimum		0	
	Maximum		1	
	Spannweite		1	
	Interquartilbereich		1	
	Schiefe		-,523	,388
	Kurtosis		-1,828	,759

Streuungsmaße berechnen

Analysieren > Deskriptive Statistiken > Deskriptive Statistiken für Grundgesamtheit

4.1.3 Häufigkeitsverteilung in SPSS

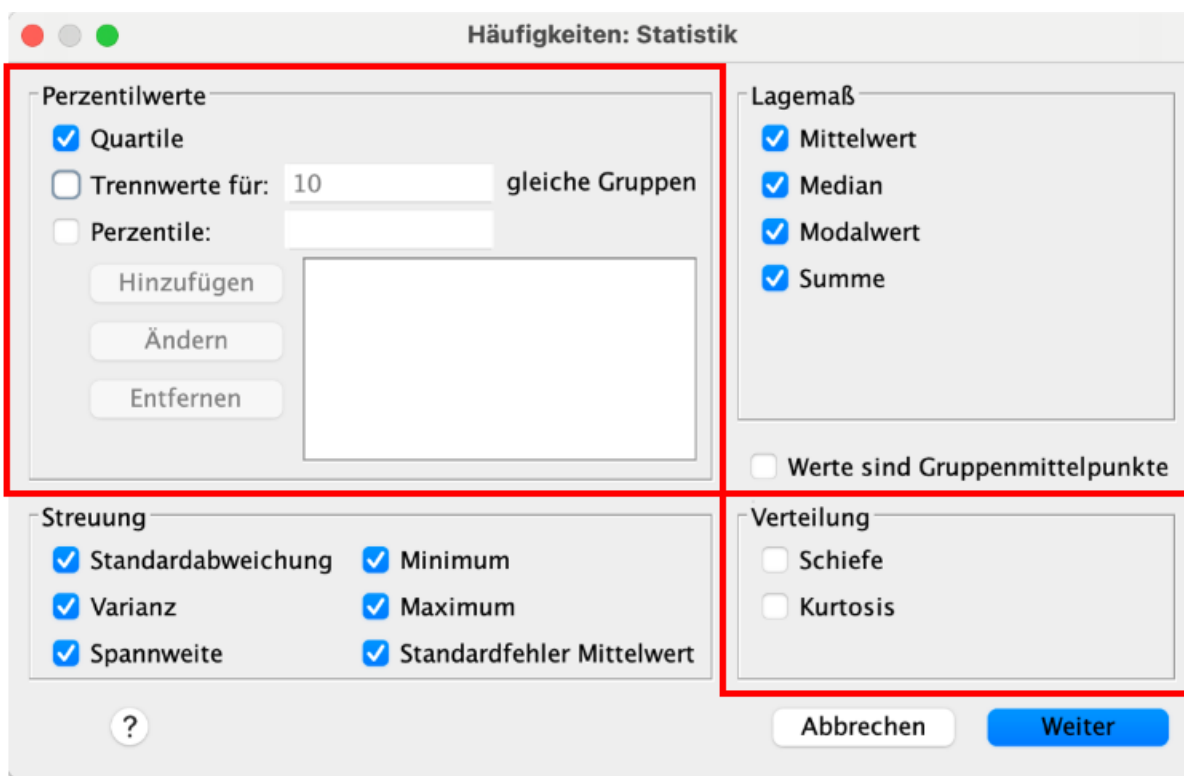
Die **Häufigkeitsverteilung** beschreibt, wie häufig bestimmte Werte in einem Datensatz auftreten. Eine Häufigkeitsverteilung kann als Tabelle oder Diagramm dargestellt werden und gibt Aufschluss darüber, wie die Werte auf die verschiedenen Kategorien oder Intervalle verteilt sind. Sie ist besonders nützlich, um die Form der Verteilung zu visualisieren und zentrale Tendenzen in den Daten zu erkennen (Sedlmeier & Renkewitz, 2018).

In SPSS wird sie über

Analysieren > Deskriptive Statistiken > Häufigkeiten

abgerufen. Es werden die interessierenden Variablen ausgewählt und die Einstellungen für Statistik und Diagramme angepasst. Für die Statistik können die Lage- und Streuungsmaße aber auch die Verteilung und die Perzentile angewählt werden (Abbildung 32).

Abbildung 32
SPSS-Output Häufigkeiten



Neben der Statistik gibt die Häufigkeitsverteilung auch Diagramme aus, wenn gewollt. Hier besteht die Auswahl zwischen Balken- und Kreisdiagrammen sowie Histogrammen. Die Ausgabe der Häufigkeitsverteilung ist dann dreigeteilt. Der Output besteht aus (1) den Lage- und Streuungsmaßen für alle ausgewählten Variablen, (2) Häufigkeitstabellen und (3) Diagrammen zur Häufigkeitsverteilung (Abbildung 32).

Häufigkeiten berechnen

Analysieren > Deskriptive Statistiken > Häufigkeiten

Abbildung 33a

SPSS-Output zur Häufigkeitsverteilung > Statistiken

		Statistiken	
		Geschlecht	SPF
N	Gültig	58	59
	Fehlend	1	0
Mittelwert		,88	3,00
Standardfehler des Mittelwerts		,043	,000
Median		1,00	3,00
Modus		1	3
Std.-Abweichung		,329	,000
Varianz		,108	,000
Spannweite		1	0
Minimum		0	3
Maximum		1	3

Abbildung 33b

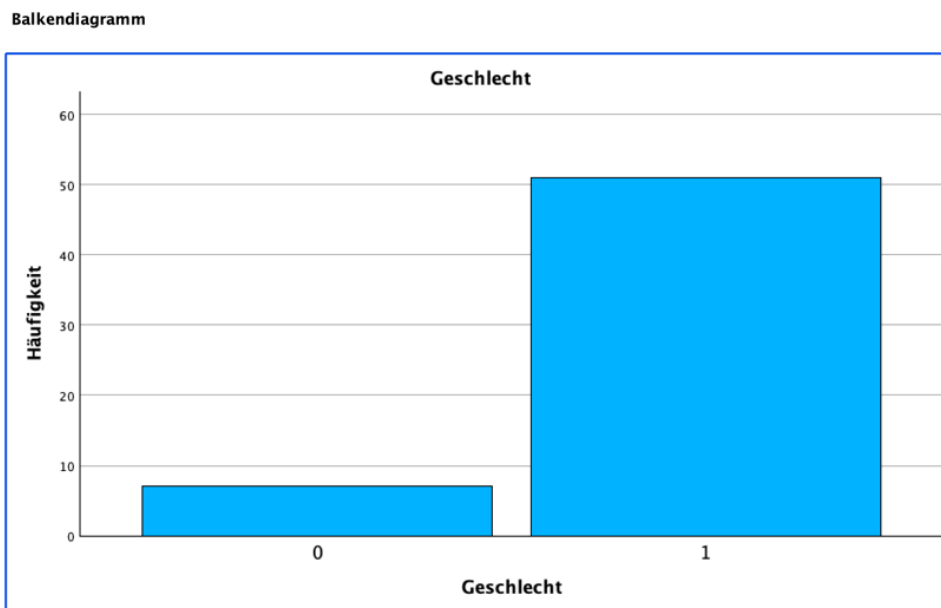
SPSS-Output zur Häufigkeitsverteilung > Häufigkeitstabellen

Häufigkeitstabelle

		Geschlecht	
		N	%
0		7	11,9%
1		51	86,4%
Fehlend	System	1	1,7%

		SPF	
		N	%
3		59	100,0%

Abbildung 33c
SPSS-Output zur Häufigkeitsverteilung > Balkendiagramm



4.1.4 Verteilungsmaße in SPSS

Verteilungsmaße wie **Schiefe** und **Kurtosis** liefern zusätzliche Informationen über die Form einer Verteilung. Die **Schiefe** (Skewness) misst, ob eine Verteilung asymmetrisch ist. Eine positive Schiefe bedeutet, dass der rechte Teil der Verteilung länger ist, während eine negative Schiefe darauf hinweist, dass der linke Teil länger ist. Die **Kurtosis** gibt hingegen an, wie stark die Verteilung im Vergleich zu einer normalen Verteilung ausgeprägt ist. Eine hohe Kurtosis bedeutet, dass die Daten einen steilen Gipfel und schwerere Ausreißer haben.

In SPSS können diese Maße ebenfalls über die Funktion

Analysieren > Deskriptive Statistiken > Deskriptive Statistik

abgerufen werden oder bei der Häufigkeit mit angewählt werden.

Die Ergebnisse der deskriptiven Statistik ermöglichen eine erste Einschätzung der Datenstruktur und liefern wertvolle Hinweise für die weitere Analyse. Dabei sind Lage- und Streuungsmaße wichtig, um mögliche Besonderheiten wie Ausreißer oder Schiefe in

der Verteilung zu erkennen. Diese Erkenntnisse dienen als Grundlage für weiterführende inferenzstatistische Verfahren, mit denen über die Stichprobe hinausgehende Schlussfolgerungen gezogen werden können.

Verteilungsmaße ausgeben

Analysieren > Deskriptive Statistiken > Häufigkeiten

4.2 Inferenzstatistik in SPSS

Während die **deskriptive Statistik** dazu dient, Daten zu beschreiben und zusammenzufassen, geht die **inferenzstatistische Analyse** einen Schritt weiter. Sie ermöglicht es, aus den Daten einer Stichprobe Rückschlüsse auf eine größere Population zu ziehen. Das bedeutet, dass mit inferenzstatistischen Methoden allgemeine Muster und Zusammenhänge erkannt werden (Sedlmeier & Renkewitz, 2018).

4.2.1 Signifikanz

Ein zentrales Konzept der Inferenzstatistik ist dabei die **statistische Signifikanz**. In der Statistik werden häufig zwei oder mehrere Gruppen miteinander verglichen, um festzustellen, ob sich diese hinsichtlich einer bestimmten Eigenschaft unterscheiden. Dabei stellt sich die Frage, ob der beobachtete Unterschied **zufällig** ist oder ob er tatsächlich existiert und auf die gesamte Grundgesamtheit übertragbar ist.

Zufällige Unterschiede können zum Beispiel durch eine ungünstige Auswahl der Stichprobe entstehen. Die statistische Signifikanz gibt an, wie wahrscheinlich es ist, dass der beobachtete Unterschied nicht nur zufällig aufgetreten ist, sondern auch in der Population besteht (Döring & Bortz, 2016).

Um die Signifikanz eines Ergebnisses zu überprüfen, werden statistische Tests eingesetzt, beispielsweise der **t-Test** (siehe 4.2.3 t-Tests in SPSS). Diese Tests berechnen den **p-Wert**, der angibt, wie hoch die Wahrscheinlichkeit ist, dass der gefundene Unterschied zufällig zustande gekommen ist (ebd.).

Das **Signifikanzniveau** definiert die Grenze, ab wann ein Ergebnis als statistisch signifikant gilt. Häufig wird ein Signifikanzniveau von 5 % ($p \leq 0,05$) verwendet. Das bedeutet:

Wenn der p-Wert kleiner oder gleich 0,05 ist, kann der Unterschied als statistisch signifikant betrachtet werden (Sedlmeier & Renkewitz, 2018).

Wichtig zu beachten ist jedoch, dass die Größe des p-Werts keine Aussage über die Stärke des Effekts trifft. Ein kleiner p-Wert bedeutet lediglich, dass ein Zusammenhang besteht – nicht, wie groß dieser Zusammenhang tatsächlich ist (ebd.).

Allerdings gibt es dabei immer ein gewisses Risiko, Fehler zu machen. Man unterscheidet zwei Arten von Fehlern: den **α -Fehler** und den **β -Fehler**. Ein **α -Fehler (Fehler 1. Art)** tritt auf, wenn ein Effekt vermutet wird, obwohl es in Wirklichkeit gar keinen gibt. Das bedeutet, dass ein Unterschied als signifikant eingestuft wird, obwohl er nur durch Zufall entstanden ist. Ein **β -Fehler (Fehler 2. Art)** passiert, wenn es tatsächlich einen Effekt gibt, er aber nicht erkannt wird. In diesem Fall wird ein echter Unterschied übersehen, weil die Stichprobe oder der statistische Test nicht empfindlich genug war. Diese Fehlerquellen sind entscheidend für die Interpretation von Testergebnissen und sollten bei der Wahl des Signifikanzniveaus sowie der Stichprobengröße berücksichtigt werden (ebd.).

4.2.2 Voraussetzungsprüfung für inferenzstatistische Analysen

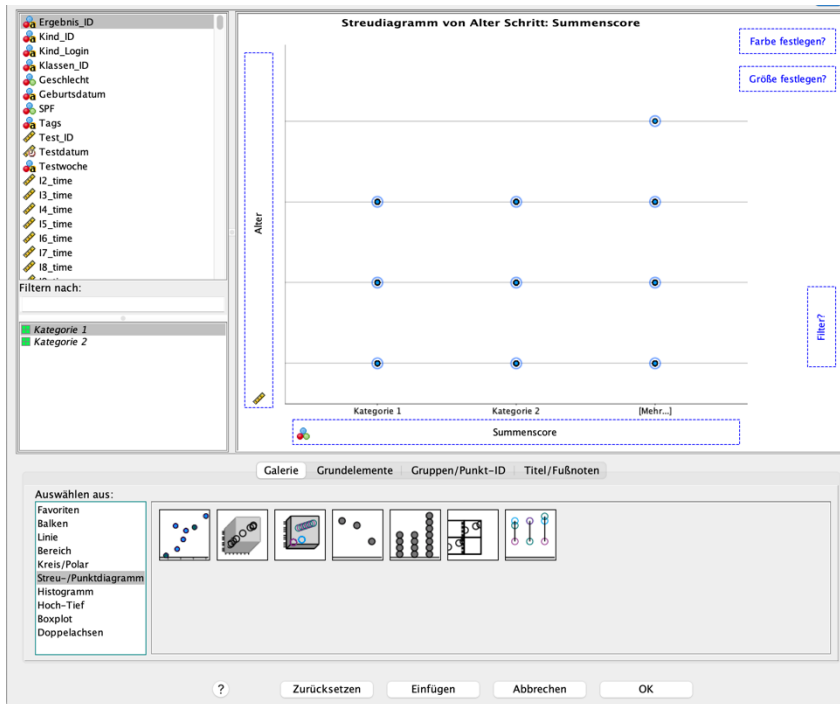
Damit inferenzstatistische Tests sinnvoll angewendet werden können, müssen bestimmte **Voraussetzungen** erfüllt sein. Typische Voraussetzungen sind Lineare Zusammenhänge, Normalitätsverteilungen und Varianzhomogenität.

Um zu prüfen, ob ein **linearer Zusammenhang** zwischen zwei Variablen vorliegt, kann ein Streudiagramm erstellt werden. Ein Streudiagramm visualisiert die Beziehung zwischen den Variablen und ermöglicht es, auf den ersten Blick zu erkennen, ob ein linearer oder nicht-linearer Zusammenhang besteht. In SPSS kann ein einfaches Streudiagramm unter

Grafik > Diagrammerstellung > Streu-/Punkt-Diagramm

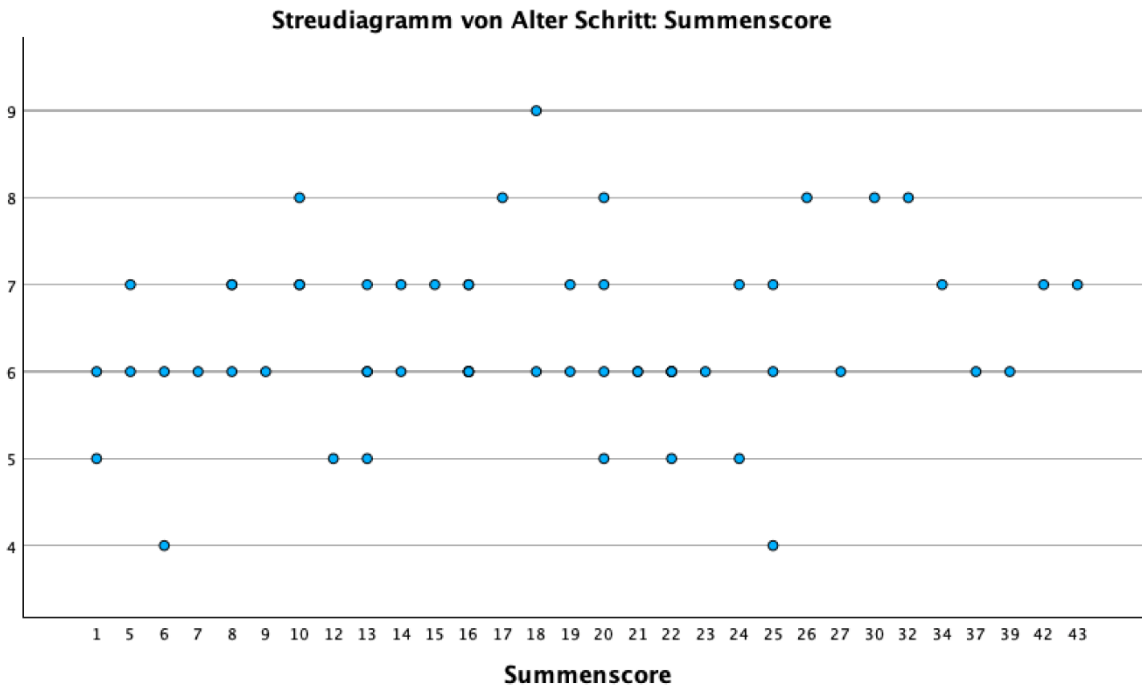
erstellt werden (Abbildung 34).

Abbildung 34
SPSS Eingabefenster Streudiagramm



In der Ausgabe erscheint das Streudiagramm (Abbildung 35). Ein Streudiagramm mit einem linearen Zusammenhang zeigt eine klare, gerade Linie, die entweder einen positiven oder negativen Trend widerspiegelt. Wenn die Datenpunkte willkürlich ohne erkennbare Richtung verteilt sind, deutet dies auf das Fehlen eines linearen Zusammenhangs hin.

Abbildung 35
SPSS-Output Streudiagramm (ohne linearen Zusammenhang)

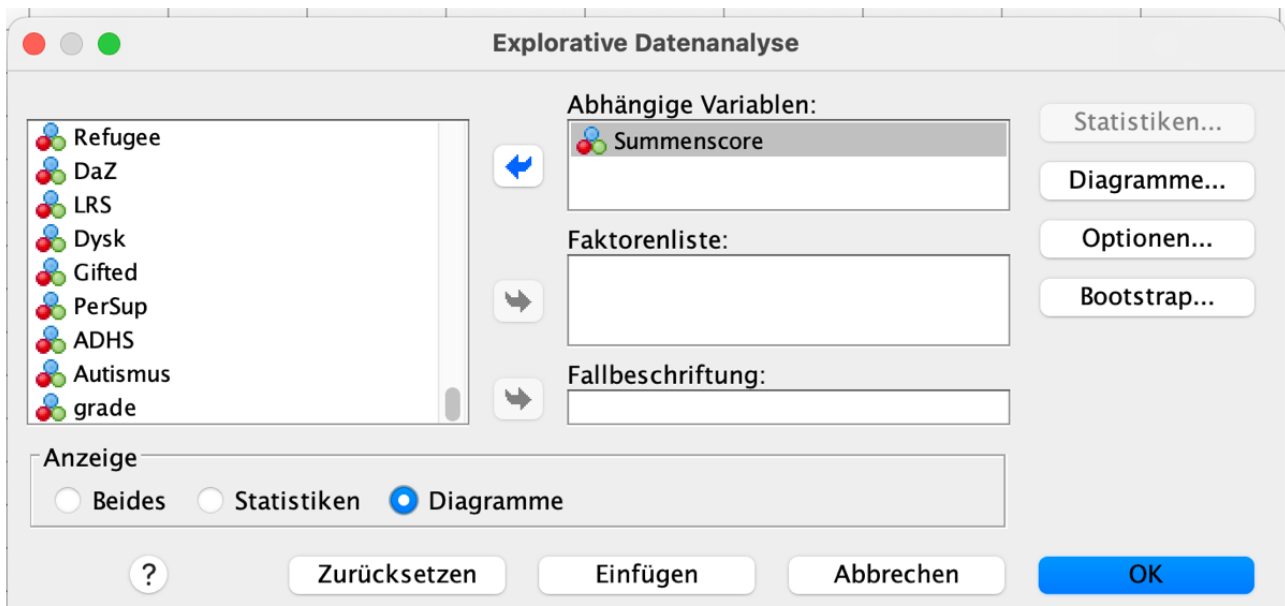


Eine weitere Voraussetzung ist die **Normalverteilung** der Daten, die vor allem für Tests wie den t-Test oder die ANOVA wichtig ist. In SPSS kann unter

Analysieren > Deskriptive Statistiken > Explorative Datenanalyse

überprüft werden, ob die Daten normalverteilt sind (Abbildung 36).

Abbildung 36
SPSS Eingabefenster zur Explorativen Datenanalyse

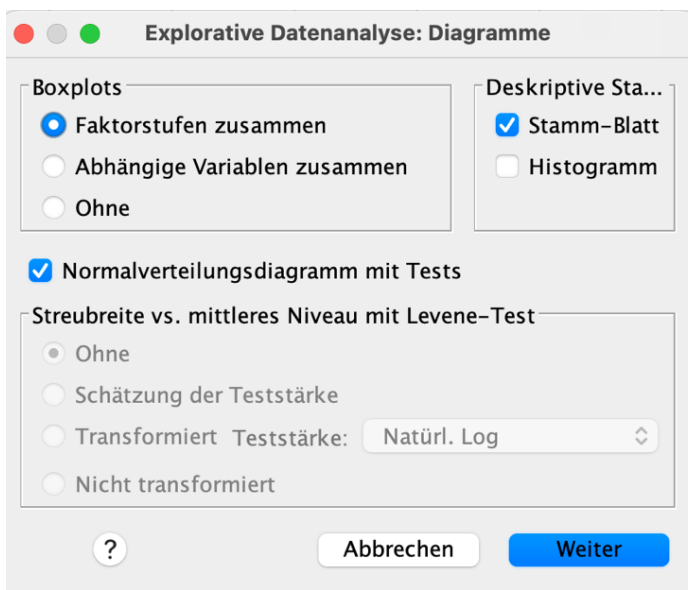


Die Funktion zur Überprüfung der Normalitätsverteilung ist unter

Diagramme > Normalverteilungsdiagramm mit Tests

zu finden (Abbildung 37).

Abbildung 37
SPSS Eingabefenster zur Explorativen Datenanalyse > Diagramme



Die Funktion berechnet den **Kolmogorov-Smirnov** sowie den **Shapiro-Wilk Test** (Abbildung 38). Dabei prüft der Kolmogorov-Smirnov Test, ob eine gegebene Variable einer bestimmten Verteilung folgt (z. B. der Normalverteilung), indem er die maximale Abweichung zwischen der beobachteten und der erwarteten Verteilung misst (Field, 2022).

Hingegen prüft der Shapiro-Wilk Test, ob eine Variable normalverteilt ist, indem er die Abweichung der beobachteten Werte von einer idealen Normalverteilung analysiert. Er ist besonders für kleine Stichproben ($n < 50$) geeignet und sensitiver als der Kolmogorov-Smirnov-Test (ebd.)

Abbildung 38
SPSS-Output Tests auf Normalitätsverteilung

Verarbeitete Fälle						
	Gültig		Fälle Fehlend		Gesamt	
	N	Prozent	N	Prozent	N	Prozent
Summenscore	59	100,0%	0	0,0%	59	100,0%

Tests auf Normalverteilung						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Summenscore	,091	59	,200 [*]	,965	59	,086

*. Dies ist eine untere Grenze der echten Signifikanz.
a. Signifikanzkorrektur nach Lilliefors

Zur Interpretation der Tests auf Normalitätsverteilung ist die Signifikanz (der p-Wert) entscheidend. Die Nullhypothese der Tests ist, dass die Daten normalverteilt sind. Ein p-Wert von $p < .50$ bedeutet also, dass die Nullhypothese abgelehnt wird, also keine Normalverteilung angenommen werden kann. In diesem Beispiel sind beide Tests $p > .50$ und damit kann Normalverteilung angenommen werden. Berichtet wird das Ergebnis zum Beispiel so: Der Summenscore war gemäß dem Shapiro-Wilk bzw. Kolmogorov-Smirnov Test normalverteilt, $p > .50$.

Darüber hinaus gibt sie verschiedene Plots zur Visualisierung aus. Dabei wird zunächst ein **Stengel-Blatt-Diagramm** (auch Stem-and-Leaf-Plot) ausgegeben (Abbildung 39). Das ist ein grafisches Verfahren zur Darstellung von Verteilungen, welches die Daten in einer sortierten, leicht lesbaren Form anzeigt. Im Kontext der Normalität hilft es, die Form der Verteilung visuell zu überprüfen.

Wenn die Daten normalverteilt sind, sollte das Stengel-Blatt-Diagramm eine symmetrische, glockenförmige Struktur aufweisen, mit den meisten Datenpunkten in der Mitte und einer gleichmäßigen Abnahme der Häufigkeit in Richtung der Ränder. Wenn die Verteilung verzerrt oder schief ist (positiv oder negativ), wird dies im Diagramm durch eine asymmetrische Form erkennbar.

Abbildung 39
SPSS-Output Stengel-Blatt-Diagramm

Summenscore		
Summenscore Stengel-Blatt-Diagramm		
Häufigkeit	Stem	Blatt
2,00	0	. 11
9,00	0	. 556678889
10,00	1	. 0002333344
12,00	1	. 566666678899
14,00	2	. 00001122222344
5,00	2	. 55567
3,00	3	. 024
1,00	3	. 7
3,00	Extremwerte (>=39)	
Stambbreite:	10	
Jedes Blatt:	1 Fälle	

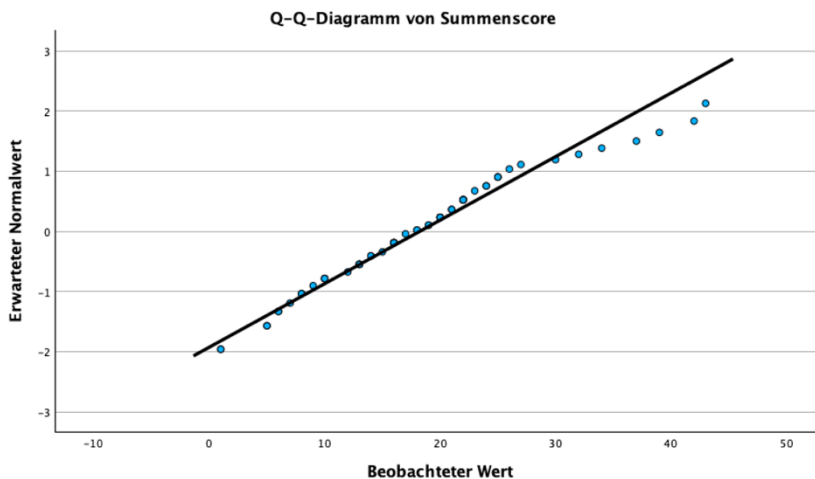
Weiter wird ein Quantil-Quantil-Diagramm (Q-Q-Diagramm) ausgegeben (Abbildung 40). Das ist ebenfalls grafische Methode zur Überprüfung der Normalität einer Verteilung. Es vergleicht die Quantile einer Stichprobe mit den Quantilen einer theoretischen Normalverteilung.

Im Q-Q-Diagramm werden die theoretischen Quantile (basierend auf der Normalverteilung) gegen die empirischen Quantile (die aus den tatsächlichen Daten berechnet werden) aufgetragen. Wenn die Daten normalverteilt sind, sollten die Punkte im Diagramm nahe einer geraden Linie liegen. Abweichungen von dieser Linie deuten auf Abweichungen von der Normalverteilung hin:

- Eine S-Form (mit einer positiven oder negativen Krümmung) deutet auf eine schiefe Verteilung hin.
- Ein deutliches Abweichen an den Rändern zeigt an, dass die Verteilung extreme Werte (z.B. Ausreißer) enthält.

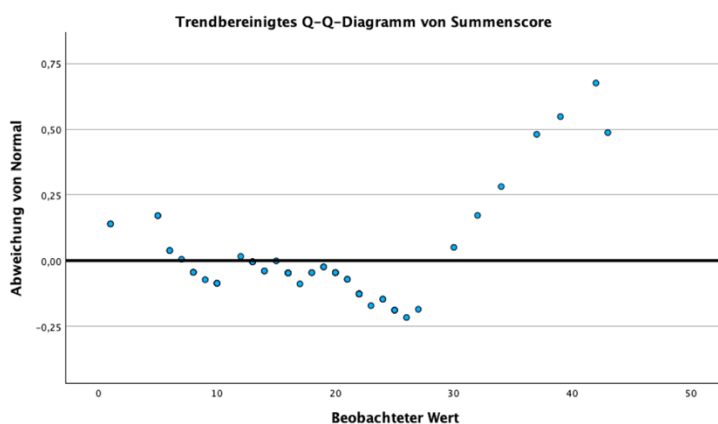
Das Q-Q-Diagramm ist also ein hilfreiches Werkzeug, um visuell die Übereinstimmung einer Verteilung mit einer Normalverteilung zu überprüfen.

Abbildung 40
SPSS Eingabefenster zur Explorativen Datenanalyse



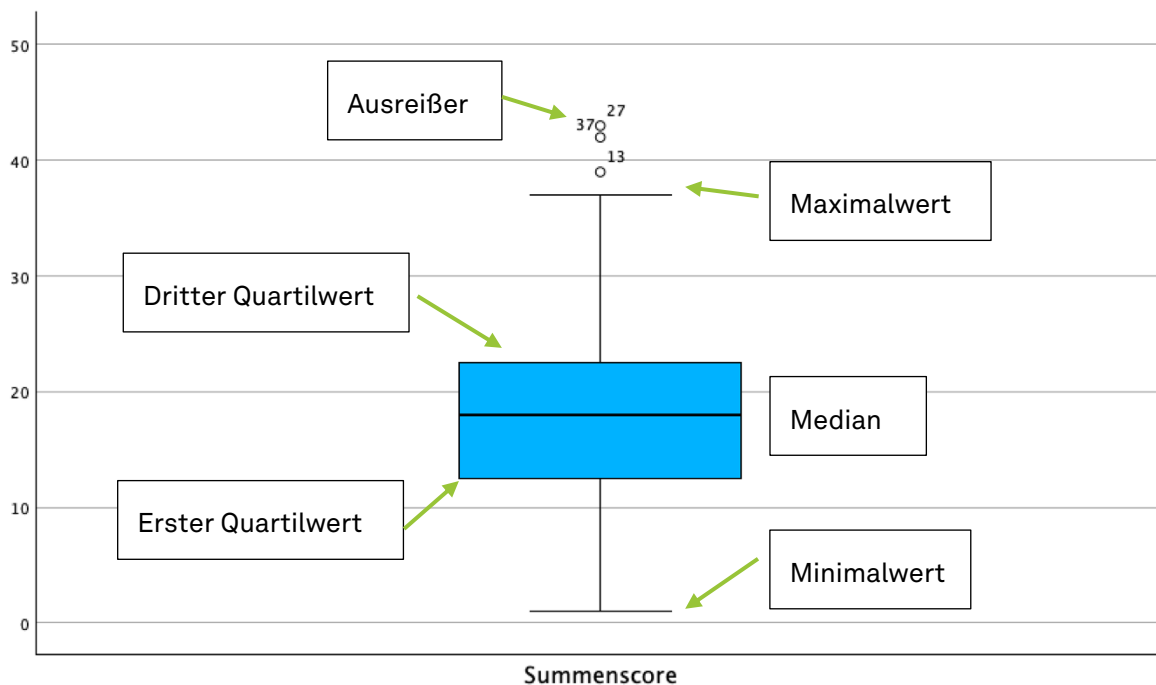
Außerdem ist ein trendbereinigtes Q-Q-Diagramm Teil der Ausgabe (Abbildung 41). Das ist eine erweiterte Version des klassischen Q-Q-Diagramms, bei dem zusätzlich eine Trendkomponente entfernt wird, um den Vergleich zwischen den empirischen Daten und einer theoretischen Verteilung (häufig der Normalverteilung) zu erleichtern. Dies wird durchgeführt, um den Effekt von systematischen Verzerrungen oder Trends in den Daten zu eliminieren und dadurch eine genauere Beurteilung der Verteilung der Daten zu ermöglichen.

Abbildung 41
SPSS Eingabefenster zur Explorativen Datenanalyse



Abschließend wird der **Boxplot** ausgegeben (Abbildung 42). Dieser wird verwendet, um die Verteilung und mögliche Ausreißer in den Daten visuell darzustellen, insbesondere im Kontext der Überprüfung der Normalität. Ein Boxplot zeigt die fünf wichtigsten Kennzahlen einer Verteilung: den **Minimalwert**, den **ersten Quartilwert (Q1)**, den **Median**, den **dritten Quartilwert (Q3)** und den **Maximalwert**. Zusätzlich werden potenzielle **Ausreißer** außerhalb der Whisker (den Linien, die die Box nach oben und unten verlängern) angezeigt.

Abbildung 42
SPSS Eingabefenster zur Explorativen Datenanalyse



Eine weitere wichtige Voraussetzung für parametrische Tests wie den t-Test oder die einfaktorielle Varianzanalyse (ANOVA) ist die **Varianzhomogenität**, also die Frage, ob verschiedene Gruppen in einer Analyse ungefähr gleich große Schwankungen in den Daten aufweisen. Die Varianzhomogenität kann in SPSS direkt für den t-Test oder die ANOVA mit ausgegeben werden. Verwendet wird dafür der **Levene Test**. Wie er in SPSS berechnet werden kann, wird im Weiteren für den t-Test und die ANOVA erläutert.

4.2.3 T-Test in SPSS

Ein t-Test wird verwendet, um die Mittelwerte von zwei Gruppen zu vergleichen und herauszufinden, ob es einen signifikanten Unterschied zwischen diesen gibt. In SPSS gibt es verschiedene Arten von t-Tests, die je nach Art der Gruppen und der Datenstruktur ausgewählt werden.

Der **t-Test für abhängige Stichproben** wird verwendet, um zu testen, ob die Mittelwerte zweier abhängiger Gruppen, also Gruppen, bei denen sich die Messwerte gegenseitig beeinflussen (wie bei wiederholten Messungen an denselben Proband:innen), unterschiedlich sind (Sedlmeier & Renkewitz, 2018). Ein Beispiel für abhängige Stichproben wäre der Vergleich der Leistungen von Schüler:innen zu zwei verschiedenen Zeitpunkten.

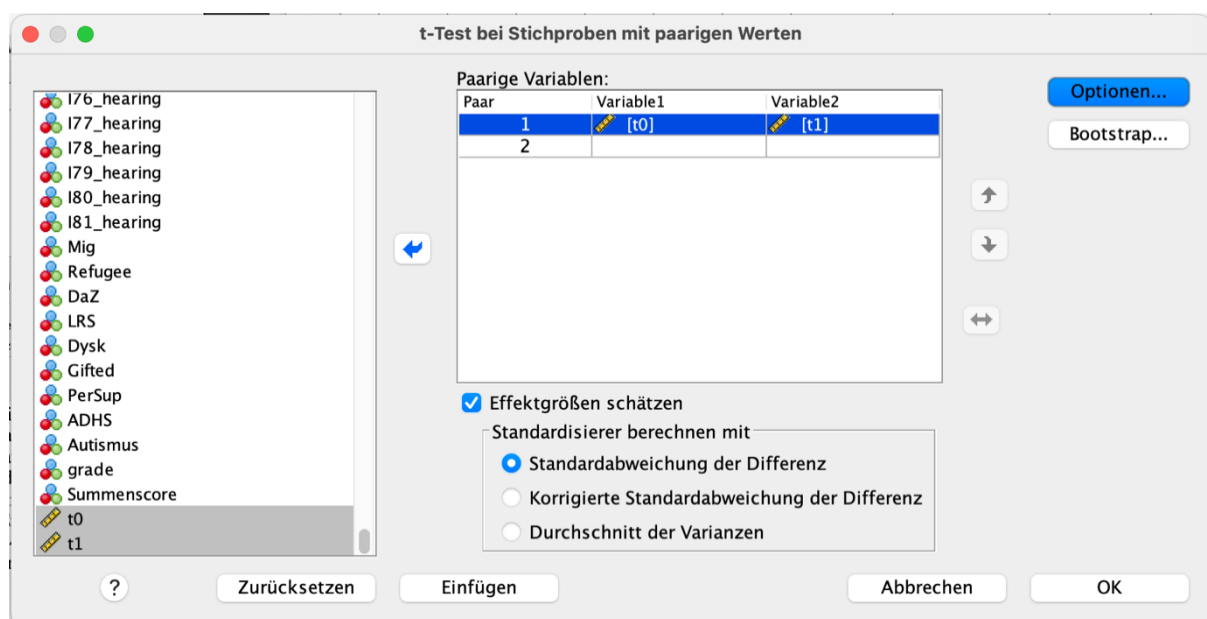
In SPSS ist der Test unter

Analysieren > Mittelwerte und Proportionen vergleichen > t-Test bei Stichproben mit paarigen Werten

zu finden. Die beiden Variablen, deren Mittelwerte verglichen werden sollen (z. B. die Messwerte zu den Zeitpunkten t0 und t1), werden als Variablenpaar ausgewählt (Abbildung 43).

Abbildung 43

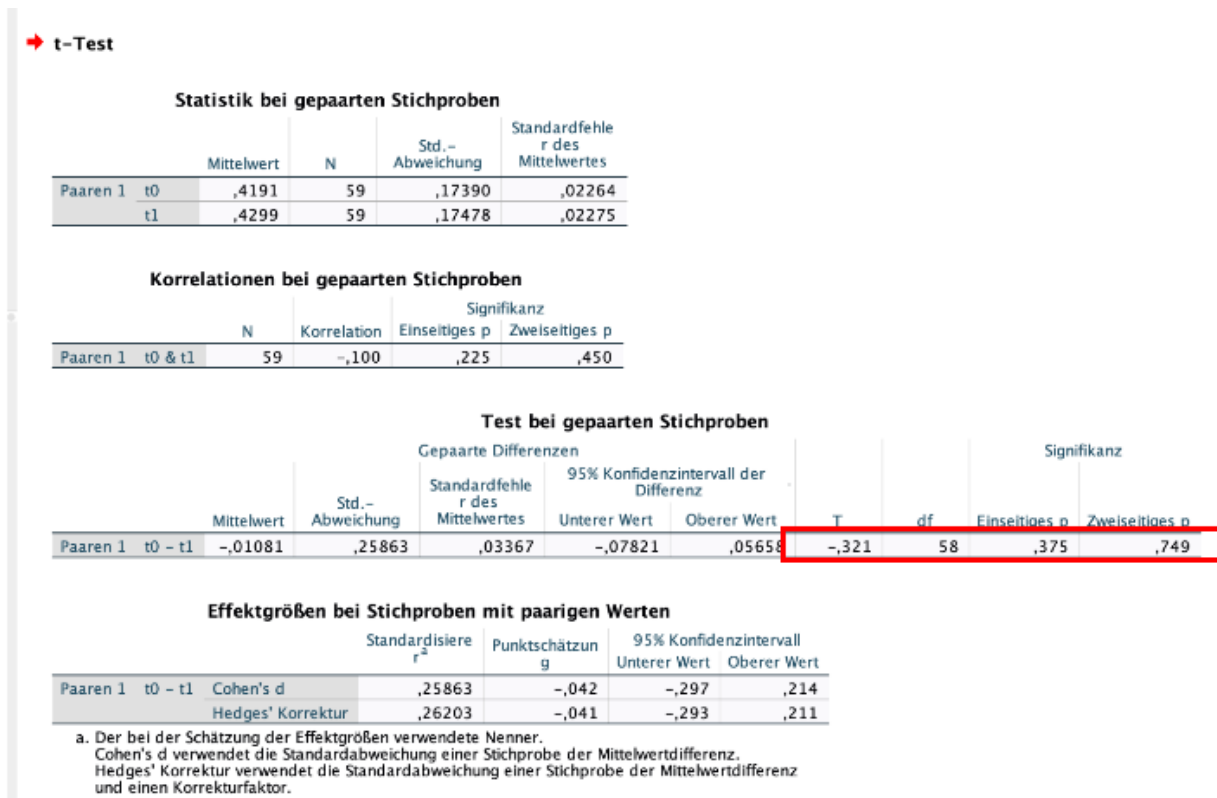
SPSS Eingabefenster zu t-Tests bei Stichproben mit paarigen Werten



SPSS berücksichtigt bei der Berechnung des t-Tests nur die Fälle, bei denen Daten zu beiden Messzeitpunkten vorliegen. In der Ausgabe erscheinen die Teststatistik, Freiheitsgrade (df) und p-Werte (Abbildung 44).

Abbildung 44

SPSS-Output zu t-Tests bei Stichproben mit paarigen Werten



Ein Beispiel für die Verschriftlichung dieser Ergebnisse im Fließtext lautet: „Zwischen dem ersten Messzeitpunkt (M = 0.42, SD = 0.17) und dem zweiten Messzeitpunkt (M = 0.43, SD = 0.17) ergab sich kein signifikanter Mittelwertunterschied (t(58) = -0.321, df = 58, p = 0.749).“

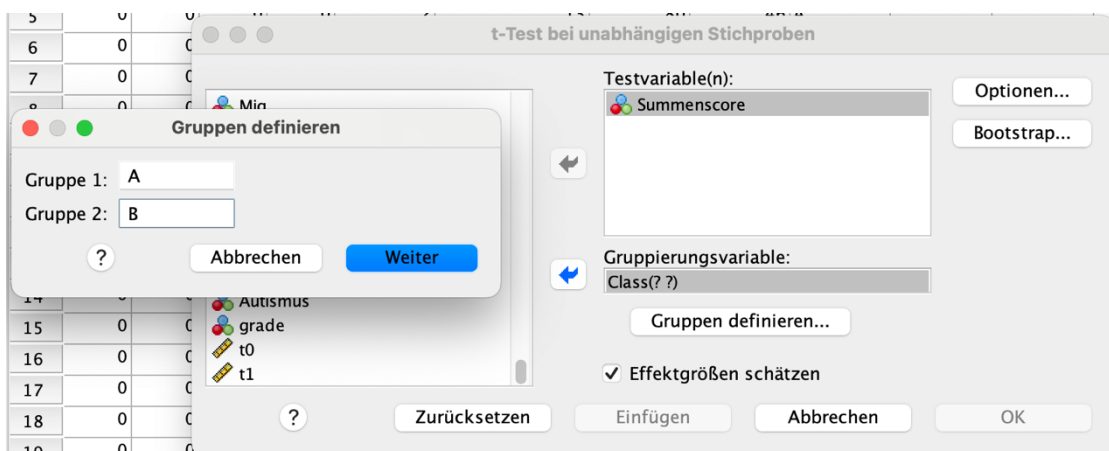
Der **t-Test für unabhängige Stichproben** wird verwendet, um die Mittelwerte von zwei unabhängigen Gruppen zu vergleichen, bei denen die Messwerte der einen Gruppe die Messwerte der anderen Gruppe nicht beeinflussen (Sedlmeier & Renkewitz, 2018). Dies ist beispielsweise der Fall, wenn zwei Klassen hinsichtlich eines Tests verglichen werden.

In SPSS ist der Test unter

Analysieren > Mittelwerte und Proportionen vergleichen > t-Test für unabhängige Stichproben

zu finden. Als Gruppierungsvariable wird die Variable festgelegt, die die Stichprobe in zwei Gruppen teilt (z. B. *Class* für die Zugehörigkeit zu einer Klasse, hier beispielhaft mit den Ausprägungen *Klasse A und B*). Zusätzlich müssen die beiden Gruppen definiert werden, die miteinander verglichen werden sollen (hier also A und B). Die getestete Variable (z. B. *Summenscore*) wird als Testvariable festgelegt (Abbildung 45).

Abbildung 45
SPSS Eingabefenster zu t-Tests bei unabhängigen Stichproben



In der Ausgabe werden die Mittelwerte beider Gruppen, die Teststatistik und der p-Wert angezeigt, aber auch der Levene-Test zur Varianzhomogenität (Abbildung 46).

Abbildung 46
SPSS-Output zu t-Tests bei unabhängigen Stichproben inkl. Gruppenstatistiken

t-Test

Gruppenstatistiken					
	Class	N	Mittelwert	Std.-Abweichung	Standardfehler des Mittelwertes
Summenscore	A	40	17,25	7,902	1,249
	B	19	20,32	12,120	2,781

Test bei unabhängigen Stichproben											
		Levene-Test der Varianzgleichheit			t-Test für die Mittelwertgleichheit						
		F	Sig.	T	df	Signifikanz		Mittlere Differenz	Differenz für Standardfehler	95% Konfidenzintervall der Differenz	
						Einseitiges p	Zweiseitiges p			Unterer Wert	Oberer Wert
Summenscore	Varianzen sind gleich	2,817	,099	-1,166	57	,124	,249	-3,066	2,630	-8,333	2,201
	Varianzen sind nicht gleich			-1,006	25,523	,162	,324	-3,066	3,048	-9,337	3,206

Abbildung 46b

SPSS-Output zu *t*-Tests bei unabhängigen Stichproben > Effektgrößen

Effektgrößen bei unabhängigen Stichproben

Summenscore	Cohen's d	Standardisiere	Punktschätzung	95% Konfidenzintervall	
		r ^a	g	Unterer Wert	Oberer Wert
		9,440	-,325	-,873	,226
	Hedges' Korrektur	9,567	-,320	-,861	,223
	Glass' Delta	12,120	-,253	-,802	,303

- a. Der bei der Schätzung der Effektgrößen verwendete Nenner.
 Für 'Cohen d' wird die zusammengefasste Standardabweichung verwendet.
 Für die Hedges-Korrektur wird die zusammengefasste Standardabweichung mit einem Korrekturfaktor verwendet.
 Für das Glass-Delta wird die Standardabweichung der Stichprobe der Kontrollgruppe (d. h. der zweiten Gruppe) verwendet.

Ein Beispiel für den Fließtext könnte lauten: „Zwischen dem Mittelwert der Klasse A (M = 17.25, SD = 7.902) und dem Mittelwert der Klasse B (M = 20.32, SD = 2.781) ergab sich kein signifikanter Mittelwertunterschied ($t(57) = -1.166$, $df = 57$, $p = 0.249$).“

4.2.4 Einfaktorielle Varianzanalyse in SPSS

Die einfaktorielle **ANOVA** stellt eine Alternative zum *t*-Test dar, wenn mehrere Gruppen miteinander verglichen werden sollen. Im Gegensatz zum *t*-Test können in der ANOVA zusätzlich Effektstärken wie η^2 berechnet werden, wodurch eine differenziertere Interpretation der Ergebnisse möglich wird (Sedlmeier & Renkewitz, 2018). Vor der Durchführung einer ANOVA ist sicherzustellen, dass die gewählten Einstellungen der spezifischen Fragestellung entsprechen.

Eine einfaktorielle ANOVA wird genutzt, um die Mittelwerte mehrerer unabhängiger Stichproben zu vergleichen, die durch eine kategoriale unabhängige Variable definiert sind (Döring & Bortz, 2016). In SPSS ist zu beachten, dass die Analyse ausschließlich numerische Werte verarbeiten kann. Da viele personenbezogene Daten im Levumi-Datensatz als Zeichenfolgen vorliegen, müssen diese vor der Analyse entsprechend transformiert werden. Dies erfolgt über die Funktion

Transformieren > Umcodieren in dieselbe Variable.

Nachdem die Werte definiert und die Variable als numerisch festgelegt wurde, kann die ANOVA durchgeführt werden.

Ein praktisches Beispiel für eine einfaktorielle ANOVA in SPSS ist die Untersuchung der Mittelwertsunterschiede zwischen vier unabhängigen Gruppen, beispielsweise verschiedener Jahrgangsstufen. Der Analyseweg in SPSS erfolgt über

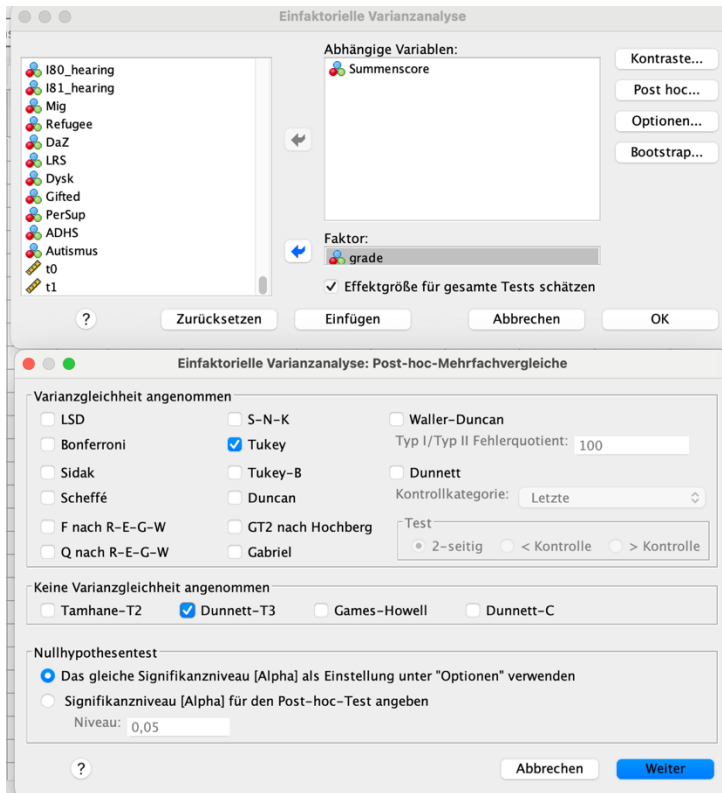
Analisieren > Mittelwerte vergleichen > Einfaktorielle Varianzanalyse.

Dabei werden die abhängigen Variablen (z. B. Summenscore) und der Faktor (z. B. grade für Klassenstufe) definiert und die Post-hoc Analyseverfahren festgelegt (Abbildung 47). Post-hoc Analysen dienen dazu konkrete Aussagen darüber zu treffen, welche Gruppen sich voneinander unterscheiden. Bei einer signifikanten ANOVA ohne Post-hoc Analysen kann sonst nur eine Aussage darüber getroffen werden, ob sich zwei Gruppen voneinander unterscheiden oder nicht, nicht aber welche.

Es gibt eine Vielzahl von Post-hoc Analyseverfahren. Wenn vor der Analyse schon eine Hypothese darüber besteht, welche Gruppen sich voneinander unterscheiden, werden Post-hoc Kontraste (also geplante Kontraste zwischen bestimmten Gruppen) berechnet. Wenn keine vorherige Hypothese besteht, werden Post-hoc Tests verwendet, die alle paarweisen Vergleiche vornehmen.

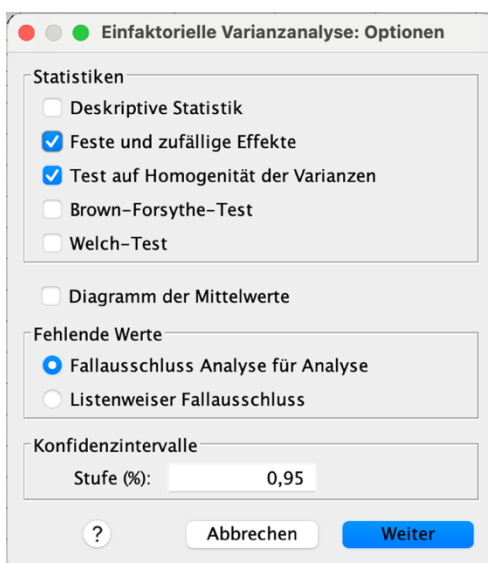
In SPSS sind die möglichen Analyseverfahren zudem unterteilt in Verfahren, für die Varianzgleichheit bestehen muss und in solche, bei denen das nicht der Fall sein muss. Die einzelnen Analyseverfahren unterscheiden sich hinsichtlich der Ausprägung (konservativ vs. Liberal). Welches Verfahren für den eigenen Test gewählt wird, sollte unter zu Hilfenahme eines Statistikbuches erfolgen. In diesem Beispiel wird der Tukey-Test als sehr ausgeglichener (zwischen konservativ und liberal) und weit verbreiteter Test gewählt. Der Dunnett-T3 ist hingegen zum Beispiel für kleinere Stichproben empfohlen. Es ist auch möglich, das Signifikanzniveau der Post-hoc Analysen anzupassen. Standardmäßig entspricht es $p = .05$.

Abbildung 47
SPSS Eingabefenster zur einfaktoriellen ANOVA



Zudem kann unter Optionen *Feste und zufällige Effekte* aktiviert werden, um Effektstärken zu berechnen und zu bestimmen, ob die ermittelten Mittelwertsunterschiede über zufällige Effekte hinausgehen. Auch die Varianzhomogenität kann unter der Option *Test auf Homogenität der Varianzen* getestet werden (Abbildung 48).

Abbildung 48
SPSS Eingabefenster zur einfaktoriellen ANOVA > Optionen



Anhand des Levene-Tests wird die Varianzhomogenität überprüft (Abbildung 49).

Abbildung 49a

SPSS-Output zur einfaktoriellen ANOVA > Deskriptive Statistiken

Deskriptive Statistik						
Summenscore		Std.- Abweichung	Std.-Fehler	95% Konfidenzintervall des Mittelwerts		Varianz zwischen den Komponenten
Modell				Untergrenze	Obergrenze	
	Feste Effekte	8,499	1,126	15,94	20,45	
	Zufällige Effekte		2,895	8,98	27,41	26,903

Abbildung 49b

SPSS-Output zur einfaktoriellen ANOVA > Tests der Varianzhomogenität

Tests der Varianzhomogenität					
		Levene- Statistik	df1	df2	Sig.
Summenscore	Basiert auf dem Mittelwert	1,546	3	53	,213
	Basiert auf dem Median	1,020	3	53	,391
	Basierend auf dem Median und mit angepaßten df	1,020	3	47,773	,392
	Basiert auf dem getrimmten Mittel	1,540	3	53	,215

Der Levene-Test prüft, ob die Annahme der Varianzhomogenität zutrifft. Bei einem signifikanten Levene Test ($p < .05$) ist von einer Verletzung der Varianzhomogenität auszugehen. Eine mögliche Formulierung der Auswertung für das vorliegende Beispiel lautet: Die Annahme der Varianzhomogenität wurde nicht verletzt, $F(3,53) = 1.55, p = .213$.

Die Interpretation der Ergebnisse der ANOVA erfolgt anschließend anhand der ANOVA-Tabelle (Abbildung 50).

Abbildung 50

SPSS-Output zur einfaktoriellen ANOVA > ANOVA-Tabelle und Effektgrößen

ANOVA

Summenscore

	Quadratsumme	df	Mittel der Quadrate	F	Sig.
Zwischen den Gruppen	1344,729	3	448,243	6,206	,001
Innerhalb der Gruppen	3828,148	53	72,229		
Gesamt	5172,877	56			

ANOVA-Effektgrößen^a

Summenscore		Punktschätzung	95% Konfidenzintervall	
			Unterer	Oberer
	Eta-Quadrat	,260	,054	,405
	Epsilon-Quadrat	,218	,000	,372
	Omega-Quadrat, fester Effekt	,215	,000	,368
	Omega-Quadrat, Zufallseffekt	,084	,000	,162

a. Eta-Quadrat und Epsilon-Quadrat werden basierend auf dem Modell mit festen Effekten geschätzt.

Besonders relevant ist die Zeile „Zwischen den Gruppen“, da hier der p-Wert angezeigt wird. Ein Beispiel für die Berichterstattung eines signifikanten Ergebnisses im Fließtext lautet: „Zwischen den Mittelwerten der Gruppen ergab sich ein signifikanter Mittelwertunterschied (F = 6.206, df = 3, p < .05).“ Zusätzlich werden die Ergebnisse der Post-Hoc Tests angegeben und in der Tabelle „Homogene Gruppen“ die verglichenen Gruppen aufgeschlüsselt (Abbildung 51).

Abbildung 51a

SPSS-Output zur einfaktoriellen ANOVA > Post-Hoc Analysen

Mehrere Vergleiche

Abhängige Variable: Summenscore

		(i) grade	(j) grade	Mittelwertdifferenz (i-j)	Std.-Fehler	Sig.	95% Konfidenzintervall	
							Untergrenze	Obergrenze
Tukey-HSD	1	2		-11,023 [*]	3,329	,009	-19,85	-2,19
		3		-10,641 [*]	3,220	,009	-19,18	-2,10
		4		-14,636 [*]	3,624	<,001	-24,25	-5,02
	2	1		11,023 [*]	3,329	,009	2,19	19,85
		3		,382	2,884	,999	-7,27	8,03
		4		-3,614	3,329	,700	-12,44	5,22
	3	1		10,641 [*]	3,220	,009	2,10	19,18
		2		-,382	2,884	,999	-8,03	7,27
		4		-3,995	3,220	,604	-12,54	4,55
	4	1		14,636 [*]	3,624	<,001	5,02	24,25
		2		3,614	3,329	,700	-5,22	12,44
		3		3,995	3,220	,604	-4,55	12,54

Gruppenvergleiche

Liegt ein signifikantes Ergebnis für die ANOVA vor, steht die Interpretation der SPSS-Ausgabe zu den Post-hoc Analyseverfahren an. Hier wird dies beispielhaft am Tukey erläutert. Im vorliegenden Test werden vier Gruppen miteinander verglichen (Anhand der Variable grade, die die Ausprägungen 1 – 4 beinhaltet). Insgesamt gibt es daher 12 Gruppenvergleiche. Zu beachten ist dabei, dass die Gruppen in beide Richtungen verglichen werden (1 und 2, aber auch 2 und 1). Die Spalte *Mittelwertsdifferenz (I-J)* stellt die Differenz der beiden miteinander verglichenen Gruppen dar. Außerdem werden die zugehörige *Standardabweichung* und die *Signifikanz* inklusive der *Konfidenzintervalle* angegeben.

In diesem Beispiel werden die Summenscores anhand der jeweiligen Jahrgangsstufe verglichen. Auf Grund des Alters und der Entwicklung des Sprachverständnisses wäre es erwartbar, dass die Schüler:innen der Klasse 4 einen höheren Summenscore erreichen als die der Klasse 1. Bei Inspektion dieses Gruppenvergleiches (der Gruppe 1 mit 4), bedeutet der Wert -14,023, dass die erste Klasse im Schnitt 14,023 Punkte weniger im Summenscore erreicht hat, als die vierte Klasse. Das Alphaniveau wurde auf 5 % (.05) festgelegt. Daher sind Ergebnisse signifikant, bei denen $p < .05$ ist. Diese Werte werden von SPSS zusätzlich mit einem Sternchen (*) in der Spalte *Mittelwertsdifferenz (I-J)* markiert. In diesem Fall ist der Unterschied der Klasse 1 zur Klasse 4 signifikant.

Abbildung 51b

SPSS-Output zur einfaktoriellen ANOVA > Homogene Gruppen

Summenscore				
Untergruppe für Alpha = 0.05.				
	grade	N	1	2
Tukey-HSD ^{a,b}	1	11	8,73	
	3	19		19,37
	2	16		19,75
	4	11		23,36
	Sig.			1,000

Mittelwerte für Gruppen in homogenen Untergruppen werden angezeigt.

- a. Verwendet Stichprobengrößen des harmonischen Mittels = 13,470
- b. Die Größen der Gruppen ist ungleich. Es wird das harmonische Mittel der Größe der Gruppen verwendet. Fehlerniveaus für Typ I werden nicht garantiert.

SPSS gibt außerdem eine Tabelle mit **homogenen Untergruppen** aus. Gruppen, die in derselben Spalte stehen, unterscheiden sich **nicht signifikant** voneinander. Gruppen, die in unterschiedlichen Spalten stehen, sind **signifikant verschieden**. Eine Gruppe

(grade 1) unterscheidet sich in diesem Beispiel signifikant von den anderen Gruppen (grade 2 – 4). Das ist daran zu erkennen, dass sie allein in der Untergruppe 1 aufgeführt ist.

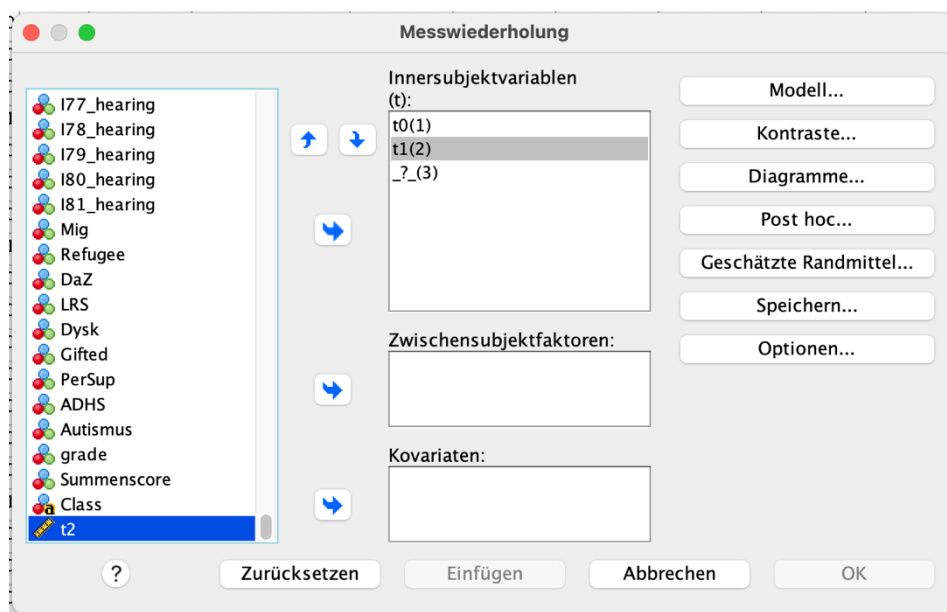
Die **ANOVA mit Messwiederholung** wird hingegen verwendet, wenn Mittelwerte innerhalb einer abhängigen Stichprobe über mehrere Zeitpunkte hinweg verglichen werden sollen. Eine Mindestanzahl von drei Messzeitpunkten ist erforderlich, um die ANOVA mit Messwiederholung sinnvoll anzuwenden. Die Analyse wird in SPSS über

Analysieren > Allgemeines lineares Modell > Messwiederholung

durchgeführt. Hier werden zunächst der Name der Messzeitpunkte und die Anzahl der Stufen definiert. Anschließend werden die Innersubjektvariablen spezifiziert und die Analyse ausgeführt (Abbildung 52).

Abbildung 52

SPSS Eingabefenster zur einfaktoriellen ANOVA mit Messwiederholung



Ein besonders wichtiger Bestandteil der ANOVA mit Messwiederholung ist der Test der Innersubjektkontraste, welcher anzeigt, ob zwischen den Messzeitpunkten signifikante Unterschiede bestehen (Abbildung 53).

Abbildung 53

SPSS-Output zur einfaktoriellen ANOVA mit Messwiederholung > Innersubjektkontraste

Tests der Innersubjektkontraste

Maß: MASS_1

Quelle	t	Typ III Quadratsumme	df	Mittel der Quadrate	F	Sig.
t	Linear	,094	1	,094	2,455	,123
	Quadratisch	,012	1	,012	,322	,573
Fehler(t)	Linear	2,215	58	,038		
	Quadratisch	2,141	58	,037		

Ein Beispiel für ein nicht-signifikantes Ergebnis wäre: „Der Test der Innersubjektkontraste ergab keinen signifikanten Unterschied zwischen den Messzeitpunkten ($p = .123$), was nicht auf eine signifikante Veränderung der Mittelwerte über die Zeit hinweg hinweist.“

Bei nur zwei Messzeitpunkten kann statt einer ANOVA mit Messwiederholung auch ein t-Test für abhängige Stichproben verwendet werden, da die ANOVA erst ab drei Messzeitpunkten eine methodische Notwendigkeit darstellt. Die korrekte Wahl des statistischen Verfahrens hängt somit von der Anzahl der Gruppen beziehungsweise Messzeitpunkten sowie von der spezifischen Fragestellung ab.

4.2.5 Korrelationen in SPSS

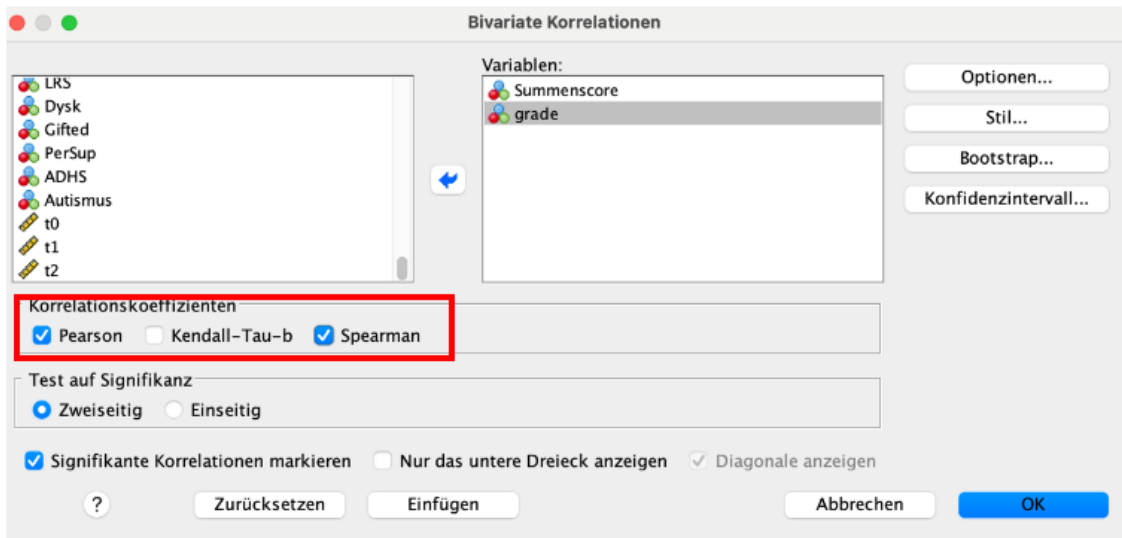
Die Korrelation ist ein statistisches Verfahren, das den Zusammenhang zwischen zwei Variablen misst. In SPSS können die Pearson- und Spearman-Korrelationen verwendet werden, um den Grad der Assoziation zwischen zwei Variablen zu untersuchen. Der Korrelationskoeffizient gibt dabei sowohl die Richtung als auch die Stärke des Zusammenhangs an und reicht von -1 (perfekte negative Korrelation) bis +1 (perfekte positive Korrelation). Ein Wert nahe 0 deutet darauf hin, dass kein oder nur ein sehr schwacher Zusammenhang besteht. Um die Signifikanz des Zusammenhangs zu überprüfen, wird der p-Wert herangezogen. Ist dieser kleiner als das gewählte Signifikanzniveau (z. B. $p < .05$), kann der Zusammenhang als statistisch signifikant betrachtet werden.

Für die Berechnung der **Korrelation nach Pearson** wird vorausgesetzt, dass zwischen den Variablen ein linearer Zusammenhang besteht und die Daten normalverteilt sind (Pearson, 1985). Wenn diese Voraussetzungen erfüllt sind, kann die Pearson-Korrelation, sonst die Spearman-Korrelation (Spearman, 1904) in SPSS über die Funktion

Analysieren > Korrelation > Bivariat

durchgeführt werden, wobei die zu korrelierenden Variablen ausgewählt und der entsprechende Koeffizient gewählt wird (Abbildung 54).

Abbildung 54
SPSS Eingabefenster zu Korrelationen



Die Ergebnisse der Berechnung erscheinen in einer Tabelle, in der der Korrelationswert sowie der p-Wert angegeben sind (Abbildung 55).

Abbildung 55a
SPSS-Output zu Korrelationen nach Pearson

Korrelationen

		Summenscore	grade
Summenscore	Pearson-Korrelation	1	,438**
	Sig. (2-seitig)		<,001
	N	59	57
grade	Pearson-Korrelation	,438**	1
	Sig. (2-seitig)	<,001	
	N	57	57

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

Ein Beispiel für eine solche Angabe im Fließtext könnte lauten: „Zwischen dem erreichten Summenscore und der Klassenstufe besteht ein positiver Zusammenhang ($r = 0.438$, $p = .01$).“

Abbildung 55b
SPSS-Output zu Korrelationen nach Spearman

Korrelationen

			Summenscore	grade
Spearman-Rho	Summenscore	Korrelationskoeffizient	1,000	,451**
		Sig. (2-seitig)	.	<,001
		N	59	57
grade		Korrelationskoeffizient	,451**	1,000
		Sig. (2-seitig)	<,001	.
		N	57	57

** . Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

Beide Korrelationen können visuell durch Streudiagramme dargestellt werden (siehe 4.2.2 Voraussetzungen für inferenzstatistische Tests prüfen). Diese helfen, den Zusammenhang zwischen den Variablen zu veranschaulichen. Bei der Pearson-Korrelation zeigt sich ein linearer Zusammenhang in einem Streudiagramm, während bei der Spearman-Korrelation auch nicht-lineare Zusammenhänge abgebildet werden können.

5 Abschluss der Datenauswertung

Nach Abschluss der Datenauswertung in SPSS ist es sinnvoll, alle relevanten Dateien systematisch in einem gemeinsamen Ordner zu speichern. Dazu gehören die Datendatei (.sav), die Syntax-Datei (.sps) sowie die Output-Datei (.spv). Diese Struktur erleichtert die Nachvollziehbarkeit der durchgeführten Analysen und ermöglicht es, die eigenen Auswertungen auch für andere verständlich zu machen. Besonders die Syntax-Datei ist dabei von großer Bedeutung, da sie sämtliche Analyseschritte dokumentiert und so die Reproduzierbarkeit der Ergebnisse sicherstellt. Ergänzend kann eine Readme-Datei erstellt werden, die eine kurze Beschreibung der Datenstruktur, der angewendeten Methoden und der zentralen Ergebnisse enthält. Durch diese Vorgehensweise wird die Transparenz der Analyse erhöht und eine spätere Weiterverarbeitung der Daten erleichtert.

Quellen

- Althelmig, M., & Jungjohann, J. (2025). *Testbeschreibung der Lernverlaufsdagnostik zum mündlichen Sprachverständnis (SpraveST): Beschreibung der Testkonstruktion sowie des Itempools* (Version 1.1). Technische Universität Dortmund. <https://doi.org/10.17877/DE290R-25303>
- Althelmig, M. & Jungjohann, J. (2024). Multiprofessionelle Teamarbeit in der schulischen Förderdiagnostik stärken – Lernverlaufstests der Onlineplattform „Levumi.de“ gemeinsam nutzen. In S. Martschinke, B. Oetjen & R. Baumann (Hrsg.), *Chance Heterogenität – Bildungsbrüche verhindern und Bildungsbrücken bauen*. (Bd. 159, S. 102–110). Grundschulverband e.V.
- de Vries, A., & Meys, J. (2012). *R für Dummies*. Wiley-VCH.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Springer.
- Field, A. (2022). *Discovering Statistics Using IBM SPSS Statistics* (5. Aufl.). Sage Publications.
- Pearson, K. (1895). *Note on regression and inheritance in the case of two parents*. *Proceedings of the Royal Society of London*, 58, 240–242.
- Jungjohann, J., DeVries, J. M., Gebhardt, M. & Mühling, A. (2018). Levumi: A web-based curriculum-based measurement to monitor learning progress in inclusive classrooms. In K. Miesenberger & G. Kouroupetroglou (Hrsg.), *Computers helping people with special needs. ICCHP 2018. Lecture notes in computer science* (S. 369–378). Springer International Publishing. https://doi.org/10.1007/978-3-319-94277-3_58
- Jungjohann, J., Diehl, K., Mühling, A. & Gebhardt, M. (2018). Graphen der Lernverlaufsdagnostik interpretieren und anwenden – Leseförderung mit der Onlineverlaufsmessung Levumi. *Forschung Sprache*, 6(2), 84–91. <https://doi.org/10.17877/DE290R-19806>
- Sedlmeier, P., & Burkhardt, J. (2021). *Datenanalyse mit R: Eine anwendungsorientierte Einführung* (2. Aufl.). Pearson.
- Sedlmeier, P., & Renkewitz, F. (2018). *Forschungsmethoden und Statistik: Ein Lehrbuch für Psychologen und Sozialwissenschaftler* (2. Aufl.). Pearson.
- Spearman, C. (1904). *The proof and measurement of association between two things*. *American Journal of Psychology*, 15(1), 72–101.
- Wickham, H. (2014). "Tidy Data". *Journal of Statistical Software*, 59(10), 1–23.

Software

IBM Corp. (2025). IBM SPSS Statistics for Mac (Version 29) [Computer software]. IBM Corp.

<https://www.ibm.com/products/spss-statistics>

Python Software Foundation. (2024). Python (Version 3.12) [Computer software].

<https://www.python.org>

R Core Team. (2025). R: A language and environment for statistical computing (Version 4.4.1)

[Computer software]. R Foundation for Statistical Computing. <https://www.r-project.org>

SAS Institute Inc. (2025). SAS (Version 9.4 M9) [Computer software]. SAS Institute Inc.

<https://www.sas.com>

The jamovi project. (2025). jamovi (Version 2.6.44) [Computer software].

<https://www.jamovi.org>

Weiterführende Literatur

Rasch, B. Friese, M., Hofmann, W. J. & Naumann, E. (2010). Quantitative Methoden 1. Einführung in die Statistik für Psychologen und Sozialwissenschaftler. Berlin: Springer.

Rasch, B. Friese, M., Hofmann, W. J. & Naumann, E. (2014). Quantitative Methoden 2. Einführung in die Statistik für Psychologen und Sozialwissenschaftler. Berlin: Springer.

Website Statistikguru: Unter dem Link finden Sie hilfreiche und kleinschrittige Anleitungen zur Arbeit in R, SPSS und Python <https://statistikguru.de>

Youtube Kanal: Educresem: Unter dem folgenden Link finden Sie Videos, die die Nutzung von SPSS in Englischer Sprache erklären. Bitte ziehen Sie bei Fragen zunächst diese Hilfe heran und wenden sich erst danach an die Personen des Lehrstuhls. <https://educresem.wordpress.com/tutorials/>