

EINFÜHRUNG IN SPSS

Version vom 07.02.2024

I. VERFÜGBARKEIT

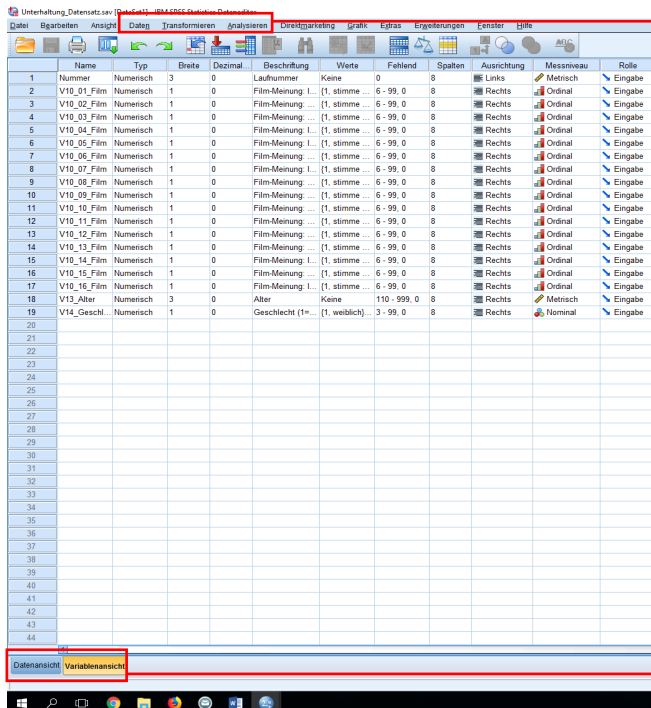
- In den PC-Pools und über Citrix können Sie die SPSS-Lizenz der Uni nutzen.
- IBM bietet eine [kostenlose 30-tägige Testversion](#) an.
- [PSPP](#) ist eine freie Alternative

II. FENSTER UND DATEIFORMATE

Die Benutzeroberfläche von SPSS umfasst drei Fenster:

1. Den **Dateneditor**:

- Ist in eine Variablenansicht und eine Datenansicht unterteilt. Die **Variablenansicht** zeigt in den Zeilen die einzelnen Variablen und in den Spalten deren Eigenschaften. Die **Datenansicht** zeigt in den Zeilen die untersuchten Fälle und in den Spalten die einzelnen Variablen. Ein Wert in der Datenansicht entspricht somit der codierten Ausprägung eines bestimmten Falles in einer bestimmten Variable.
- Über die Menüs *Daten*, *Transformieren* und *Analysieren* in der Menüleiste können die vorhandenen Daten bearbeitet und ausgewertet werden.
- Dateiformat ist `.sav`.



Menüleiste mit den wichtigsten Analysemenüs

Umschalten zwischen Datenansicht und Variablenansicht

2. Den **Syntaxeditor**:

- Bietet die Möglichkeit, Daten über Syntaxbefehle zu bearbeiten und auszuwerten.
- Kann damit (bei ausreichenden Kenntnissen) die Menüs im Dateneditor ersetzen oder ergänzen.
- Befehle, die über das Menü ausgelöst wurden, können im Syntaxeditor abgespeichert und die Datenauswertung so dokumentiert werden (Transparenz).
- Potentielle Vorteile gegenüber den Befehlen im Dateneditor: Schnellere Befehlseingabe, leichtere Wiederholbarkeit von Berechnungen, Dokumentation der durchgeführten Aktionen
- Dateiformat ist *.sps*.

3. Den **Viewer**:

- Dient zur Anzeige von Ergebnissen und zur Dokumentation von Befehlen (dazu: *Bearbeiten > Optionen > Viewer > Haken bei Befehle im Protokoll anzeigen* setzen).
- Dateiformat ist *.spv*.

III. VARIABLEN

- Zum **Erstellen einer Variable** muss zunächst im Dateneditor die Variablenansicht geöffnet werden.
- Klickt man dort in ein Feld in Spalte 1 *Name*, so kann man über die Tastatur einen Namen für die erste Variable eingeben. Mit *Enter* wird die Eingabe bestätigt und die Variable erstellt. Achtung: Der Variablenname darf unter anderem keine Leerzeichen enthalten. Er sollte so gewählt werden, dass er zur schnellen Identifikation der Variable dient, also kurz und prägnant sein. Zum Beispiel: „V01_sex“ für eine erste Variable, die nach dem Geschlecht einer Person fragt.
- Die weiteren Spalten, d. h. Eigenschaften der Variable werden von SPSS automatisch erstellt und müssen dann nach den eigenen Bedürfnissen angepasst werden. Die Anpassung erfolgt durch einen Klick auf das jeweilige Feld bzw. einen dort hinterlegten Button. Angepasst werden können die folgenden Spalten:
 - *Typ*: Hier wird der Variablentyp definiert. Meist wird der Typ *Numerisch*, d. h. eine ganze Zahl oder Dezimalzahl benötigt. Bei einer offenen Kategorie wird der Typ *Zeichenfolge* benötigt – Mit diesem können jedoch keine Berechnungen durchgeführt werden.
 - *Breite*: Gibt die Spaltenbreite in Zeichen an.
 - *Dezimalstellen*: Gibt die Anzahl zulässiger Dezimalstellen an.
 - *Beschriftung*: Hier können Sie Ihre Variable mit einer Beschreibung versehen, aus der hervorgeht, was diese abfragt. Zum Beispiel wäre „Geschlecht der befragten Person“ eine Beschriftung für die Variable „V01_sex“.
 - *Werte*: Hier können Sie Werte für die einzelnen Ausprägungen Ihrer Variable hinterlegen und diese mit Beschriftungen versehen. Zum Beispiel legen Sie für die Variable „V01_sex“ die Werte 1=„weiblich“, 2=„männlich“ und 3=„divers“ an.

In der Inhaltsanalyse entsprechen diese Werte den Codes der Ausprägungen einer Kategorie.

- *Fehlend*: Hier können Sie Werte definieren, welche keine sinnvolle Entsprechung in Ihren Ausprägungen finden. Diese werden dann in der Analyse Ihrer Daten nicht berücksichtigt, obwohl sie in den Daten vorkommen.
 - *Spalten*: Spaltenbreite im Datensatz in Zeichen (nur optische Bedeutung)
 - *Ausrichtung*: Bündigkeit der Werte in der Darstellung im Datensatz (nur optische Bedeutung)
 - *Messniveau*: Dient der Angabe des Skalenniveaus. Zur Auswahl stehen *Nominal*, *Ordinal* und *Metrisch*. Achtung: Diese Option dient in erster Linie der eigenen Übersicht und schränkt NICHT prinzipiell die Analysemöglichkeiten des Programms ein. D. h.: Die Sinnhaftigkeit der durchgeführten Berechnungen muss selbst erkannt werden!
 - *Rolle* (für Sie nicht relevant): Dient zur Vorauswahl von bestimmten Variablen für die Analyse.
- Zum **Löschen** einer Variable wird diese zunächst in der Variablenansicht des Dateneditors durch einen Klick auf die Zeilennummerierung ausgewählt. Durch einen Rechtsklick öffnet sich ein Dialog mit der Option *Löschen*. Mehrere Variablen können über *Umschalt* + Klick oder *Strg* + Klick ausgewählt werden. Achtung: Mit dem Löschen einer Variable werden auch die dazugehörigen Daten gelöscht.

IV. DATENEINGABE

- Um **Daten von Hand einzugeben** öffnen Sie die *Datenansicht* im *Dateneditor*. Nach einem Klick auf ein Feld im Tabellenblatt können Sie dort einen Wert eingeben. Andere Felder können wie in Excel auch über die Pfeiltasten angesteuert werden.
- Um **Daten zu importieren** wählen Sie im *Dateneditor*: *Datei* > *Daten importieren* > *[Dateiformat]*. Wie gewohnt können Sie nun die gewünschte Datei auswählen. Wenn Sie zum Beispiel aus einer Excel-Datei neben den Werten auch die Variablennamen übernehmen wollen, so müssen diese in der ersten Zeile der Excel-Datei stehen.


V. DATENBEREINIGUNG

- Daten können aus unterschiedlichen Gründen fehlerhaft sein. Zum Beispiel können bei der Erhebung durch Messfehler oder bei der Dateneingabe durch Flüchtigkeitsfehler falsche Werte vergeben werden.
- Bevor man die Daten auswertet, sollte man zuerst diese Fehler so weit wie möglich und sichtbar korrigieren → Datenbereinigung.
- In SPSS können Sie eventuell vorhandene Fehler aber nur finden, wenn:
 - ein Wert außerhalb des zulässigen Bereichs liegt (z. B. wurde eine 8 codiert, obwohl in den Werten gar keine 8 vorgesehen ist).
 - logische Inkonsistenzen auftreten (z.B. ist ein Alter mit 5 Jahren angegeben, der Befragte ist aber verheiratet und hat zwei Kinder).

- außergewöhnliche Werte oder Kombination auftreten, die auf einen Fehler hinweisen (z.B. eine Familie mit 20 Mitgliedern, ein studentisches Einkommen von 1 Mio. Euro im Monat).
- bei einer Frage, die gar nicht gestellt werden durfte (Filterfrage), ein Wert angegeben wurde.
- Wie finden Sie diese Fehler?
 - Unzulässige Werte: Grundauszählung (Häufigkeiten) aller Variablen und Unstimmigkeiten suchen (Im Dateneditor: *Analysieren* > *Deskriptive Statistiken* > *Häufigkeiten* > *[Variablen auswählen]* > *OK*)
 - Unlogische Kombinationen finden Sie durch Kreuztabellen (Im Dateneditor: *Analysieren* > *Deskriptive Statistiken* > *Kreuztabellen* > *[Variablen auswählen]* > *OK*)
 - Praktischer Tipp: In der Datenansicht lassen sich die Werte einzelner Variablen einfach ab- oder aufsteigend sortieren: Mit einem rechts-Klick auf die entsprechende Spalte (=Variable) und auf- oder absteigend sortieren auswählen. So lassen sich einzelne Variablen überprüfen und man erkennt relativ schnell unsinnige oder fehlende Werte. Dieses Vorgehen ist aber nicht unbedingt zur systematischen Überprüfung geeignet.
- Wie behandeln Sie diese Fehler?
 - Variable, in der der Fehler gefunden wurde, durch Klick markieren, auf das Fernglas in der Symbolleiste klicken und den entsprechenden Wert, der aufgefallen ist, suchen
 - entsprechenden Fall im Untersuchungsmaterial (Fragebögen, Codebögen) noch einmal nachsehen und korrigieren
 - im Zweifel: Fall löschen oder den falschen Wert als fehlenden Wert angeben

VI. DATEN BEARBEITEN

Eine Datei aufteilen:

- Öffnen Sie den Dateneditor.
- Wählen Sie *Daten* > *Datei aufteilen* oder klicken Sie auf das Symbol 
- Die Standardeinstellung im folgenden Dialogfenster ist *Alle Fälle analysieren, keine Gruppen bilden*. Dies bestätigen Sie mit einem Klick auf *OK*, falls alle Fälle in Ihre Berechnung einbezogen werden sollen.
- Falls Sie die Fälle anhand einer Variable getrennt betrachten wollen, wählen Sie zunächst *Gruppen vergleichen* oder *Ausgabe nach Gruppen aufteilen* und anschließend aus der Liste die entsprechende Variable.
- Anschließend treffen Sie eine Auswahl zwischen *Datei nach Gruppierungsvariablen sortieren* und *Datei ist sortiert*, je nachdem, ob Ihre Werte bereits anhand der ausgewählten Variable sortiert sind oder nicht. Sie bestätigen mit *OK*.
- Hätten Sie die Beispielvariable „V01_sex“ mit den Werten 1=“weiblich“, 2=“männlich“ und 3=“divers“ ausgewählt, so würde Ihnen die Auswahl *Gruppen vergleichen* bei jeder

weiteren Berechnung eine Tabelle ausgeben, in welcher die Fälle getrennt nach Geschlecht ausgewiesen werden. Die Auswahl *Ausgabe nach Gruppen aufteilen* würde Ihnen je eine Tabelle pro Geschlecht ausgeben.

- Haben Sie Ihre Datei aufgeteilt, betrifft diese Aufteilung alle weiteren Berechnungen. Möchten Sie Ihre Fälle wieder zusammenführen, wählen Sie die Option *Alle Fälle analysieren, keine Gruppen bilden*.

Fälle auswählen:

- Um bestimmte Fälle auszuwählen, d. h. Ihr Datenset zu filtern, öffnen Sie zunächst den Dateneditor.
- Wählen Sie *Daten > Fälle auswählen* oder klicken Sie auf das Symbol 
- Das folgende Fenster unterteilt sich in *Auswählen* und *Ausgabe*.
- Die Standardeinstellung bei *Auswählen* ist *Alle Fälle*. Dies bestätigen Sie mit einem Klick auf *OK*, falls alle Fälle in Ihre Berechnung einbezogen werden sollen und Sie keinen Filter setzen wollen.
- Mit der Auswahl von *Falls Bedingung zutrifft* und einem Klick auf *Falls* öffnet sich ein Dialogfenster. In diesem können Sie mittels mathematischer Operatoren Falls-Bedingungen definieren, um Filter zu erstellen. Angenommen Sie wollen mit der Beispielformelvariable „V01_sex“ einen Filter erstellen, der nur die weiblichen Fälle auswählt, so definieren Sie die Falls-Bedingung $V01_sex=1$.
- Mit der Auswahl von *Zufallsstichprobe* können Sie Bedingungen für eine zufällige Stichprobe Ihrer Fälle definieren.
- Mit der Auswahl von *Nach Zeit- oder Fallbereich* können Sie eine Auswahl hintereinanderliegender Fälle treffen (wird nur selten benötigt).
- Mit der Auswahl von *Filtervariable verwenden* können Sie eine zuvor selbst definierte Filtervariable auswählen. Diese ist dichotom: Fälle mit dem Wert „1“ werden ausgewählt, Fälle mit dem Wert „0“ gefiltert.
- Im Bereich *Ausgabe* haben Sie die Wahl zwischen *Nicht ausgewählte Fälle filtern*, *ausgewählte Fälle in neues Dataset kopieren* und *Nicht ausgewählte Fälle löschen*. In der Regel wählen Sie *Nicht ausgewählte Fälle filtern* und bestätigen mit *OK*.
- Ist ein Filter aktiviert, so sehen Sie dies unten rechts im Dateneditor an der Anzeige *Filter aktiv*. Außerdem erstellt SPSS automatisch eine neue Variable Namens „filter_\$“, welche allen ausgewählten Fällen den Wert „1“ und allen gefilterten Fällen den Wert „0“ zuweist. Gefilterte Fälle erkennen Sie zudem daran, dass die Fallnummer in der ersten Spalte durchgestrichen ist.
- Möchten Sie einen Filter deaktivieren, so wählen Sie unter *Daten > Fälle auswählen* die Option *Alle Fälle*.

Fälle gewichten:

- Um Fälle zu gewichten erstellen Sie zunächst eine sog. Häufigkeitsvariable. D. h. eine neue Variable, welche jedem Fall sein Gewicht numerisch zuweist.
- Zum Beispiel weisen Sie einem Fall für die Häufigkeitsvariable den Wert „3“ zu, wenn Sie möchten, dass dieser in der Auswertung wie drei Fälle gezählt wird.
- Fälle, bei denen die Häufigkeitsvariable den Wert „0“, einen negativen Wert oder einen fehlenden Wert aufweist, werden von der Analyse ausgeschlossen.
- Nach dem Erstellen der Häufigkeitsvariable wählen Sie *Daten > Fälle gewichten*
- Wählen Sie in diesem Fenster die Option *Fälle gewichten mit* und die Häufigkeitsvariable. Im Dateneditor sehen Sie nun unten rechts die Information *Gewichtung aktiv*
- Um die Gewichtung zu deaktivieren wählen Sie im *Fälle gewichten*-Fenster die Option *Fälle nicht gewichten*.

VII. VARIABLEN TRANSFORMIEREN

Eine Transformation von Variablen kann hilfreich sein, wenn Sie Komplexität reduzieren, Ihre Daten verdichten oder latent gemessene Konstrukte in einer manifesten Variable abbilden wollen. Sie können z. B. Ausprägungen zusammenfassen, Ausprägungen umdrehen oder additive Indices berechnen.

Neue Variablen berechnen:

- Klicken Sie im Dateneditor auf *Transformieren > Variable berechnen*.
- Im neuen Fenster benennen Sie zunächst eine *Zielvariable*. Achtung: Entspricht die Benennung dem Namen einer bestehenden Variable, so wird diese überschrieben. Tipp: Überschreiben Sie niemals (!) eine bestehende Variable, sondern legen Sie immer eine neue Variable an. Ist eine Variable einmal überschrieben, lässt sich die ursprüngliche Version nicht mehr wiederherstellen.
- Im Feld *Numerischer Ausdruck* können Sie mittels mathematischer Operatoren und der bestehenden Variablen den Inhalt der Zielvariable definieren.
- Zum Beispiel: Sie haben die tägliche Nutzungsdauer verschiedener Medien in Minuten erhoben und wollen nun eine Variable für den täglichen Medienkonsum in Stunden berechnen:

Zielvariable:

V0X_Medienkonsum

=

Numerischer Ausdruck:

(V01_TV + V02_Radio + V03_Zeitung + ...) / 60

- Durch den Klick auf *OK* wird die neu berechnete Variable erstellt. Deren Eigenschaften können Sie in der *Variablenansicht*, wie in Kapitel III beschrieben, definieren.

Variablen umcodieren:

- Klicken Sie im Dateneditor auf *Transformieren > Umcodieren in andere Variablen*.
- (Eine Umcodierung in dieselbe Variable ist auch möglich, wird hier jedoch nicht empfohlen, da das Umcodieren immer mit einem Informationsverlust einhergeht und dieser beim Umcodieren in dieselbe Variable unwiederbringlich ist.)
- Das Fenster zur Umcodierung hat drei relevante Bestandteile: *Eingabevariable -> Ausgabevariable*, *Ausgabevariable* und *Alte und neue Werte*.
- Wählen Sie zunächst aus der Variablenliste links Ihre „alte“ Variable, d. h. diejenige welche umcodiert werden soll aus und fügen Sie diese im Feld *Eingabevariable -> Ausgabevariable* ein.
- Benennen Sie nun im Feld *Ausgabevariable* Ihre neue Variable.
- Mit einem Klick auf *Alte und neue Werte* können Sie die Werte der „alten“ Variable in neue Werte überführen, welche dann die Ausprägungen Ihrer neuen Variable darstellen. In der einfachsten Variante geben Sie unter *Alter Wert* einen bestehenden Wert und unter *Neuer Wert* einen neuen Wert ein und klicken auf *Hinzufügen*. Analog können Sie auch fehlende Werte und/oder Wertbereiche definieren.
- Zum Beispiel: Sie haben das Alter der befragten Personen in Jahren erhoben. Nun wollen Sie aber Kohorten bilden, um in einem zweiten Schritt zu erfahren, ob Teenager, Twens und Über-30-Jährige Medien anders nutzen. Für die Teenager definieren Sie als alten Wert die *Spannweite* „13“ bis „19“ und als neuen Wert „1“. Für die Twens definieren Sie als alten Wert die *Spannweite* „20“ bis „29“ und als neuen Wert „2“. Für die Über-30-Jährigen definieren Sie als alten Wert die *Spannweite*, *Wert bis GRÖSSTER* „30“ und als neuen Wert „3“. Ihr Fenster *Alt -> Neu* gestaltet sich dann wie folgt:

13 thru 19 -> 1
20 thru 29 -> 2
30 thru Highest -> 3

Nun legt SPSS eine neue Variable an, welche die drei Alterskohorten beinhaltet. Deren Eigenschaften können Sie in der *Variablenansicht*, wie in Kapitel III beschrieben, definieren.

Variablen mit Mehrfachantworten:

- Zunächst wird entsprechend der Anzahl der möglichen Antworten oder Codierungen bei diesem Item oder dieser Kategorie für jede mögliche Antwort/Codierung eine eigene Variable in SPSS angelegt.
- Zum Beispiel: Sie interessieren sich für die Themen von Zeitungsartikeln. Für jeden Artikel sollen bis zu drei Themen codiert werden können. Obwohl Sie in Ihrem Codebuch lediglich eine Kategorie „Thema“ mit zahlreichen Ausprägungen haben, legen Sie in SPSS drei Variablen an („Thema_1“, „Thema_2“, „Thema_3“). Behandelt ein Artikel

nur ein Thema wird dies unter „Thema_1“ codiert, bei zwei Themen wird jeweils ein Code bei „Thema_1“ und ein Code bei „Thema_2“ vergeben usw.

- Klicken Sie im Dateneditor auf *Analysieren > Mehrfachantworten > Variablensets definieren*.
- Fügen Sie die Variablen mit Mehrfachantworten oder Mehrfachcodierungen dem Feld *Variablen im Set* hinzu.
- Im Feld *Variablen codiert als* legen Sie unter *Kategorien* den Wertebereich Ihrer Codes fest (= niedrigster bis höchster Wert, der im Fragebogen bzw. Codebuch vergeben werden konnte).
- Fügen Sie analog zur Erstellung von Variablen in den Feldern *Name* und *Beschriftung* Eine Benennung und Beschriftung für das Mehrfachantwortenset hinzu. Bestätigen Sie mit *Hinzufügen* und *Schließen*.
- Die Datenanalyse erfolgt dann über *Analysieren > Mehrfachantworten > Häufigkeiten* oder *Kreuztabellen*

VIII. DATENANALYSE: DESKRIPTIVE STATISTIKEN

Die deskriptiven Statistiken können u. a. genutzt werden, um die Datenbereinigung vorzubereiten sowie um die Verteilung einer Variable oder das Verhältnis zweier Variablen zueinander darzustellen.

Häufigkeiten

- Klicken Sie im Dateneditor auf *Analysieren > Deskriptive Statistiken > Häufigkeiten*.
- Im folgenden Fenster können Sie zunächst eine oder mehrere Variablen auswählen, für die Sie die Häufigkeiten auszählen lassen wollen.
- Unter *Statistiken* können Sie auswählen, welche Informationen zu der/n Variable/n angezeigt werden sollen. Zur Auswahl stehen verschiedene *Perzentilwerte*, *Lagemaße*, *Streuungsmaße* und *Verteilungsmaße*.
- Unter *Diagramme* können Sie sich ein Diagramm anzeigen lassen. Bei nominal- oder ordinalskalierten Variablen kann ein Balkendiagramm, bei metrischen Variablen ein Histogramm sinnvoll sein, um einen schnellen Überblick über die Verteilung der Werte zu erhalten

Deskriptive Statistik

- Klicken Sie im Dateneditor auf *Analysieren > Deskriptive Statistiken > Deskriptive Statistik*.
- Im folgenden Fenster können Sie zunächst eine oder mehrere Variablen auswählen, für die Sie die deskriptiven Statistiken anzeigen lassen wollen.

- Unter *Optionen* können Sie auswählen, welche Informationen zu der/n Variable/n angezeigt werden sollen. Zur Auswahl stehen verschiedene, *Lagemaße*, *Streuungsmaße*, *Verteilungsmaße* und die *Anzeigenreihenfolge*.
- Im Vergleich zu den *Häufigkeiten*, zeichnet sich die deskriptive Statistik durch eine höhere Übersichtlichkeit aus. So können Sie beispielsweise anhand der Minima und Maxima schnell erkennen, ob in Ihren Daten falsche oder unrealistische Werte vorliegen.

Kreuztabellen

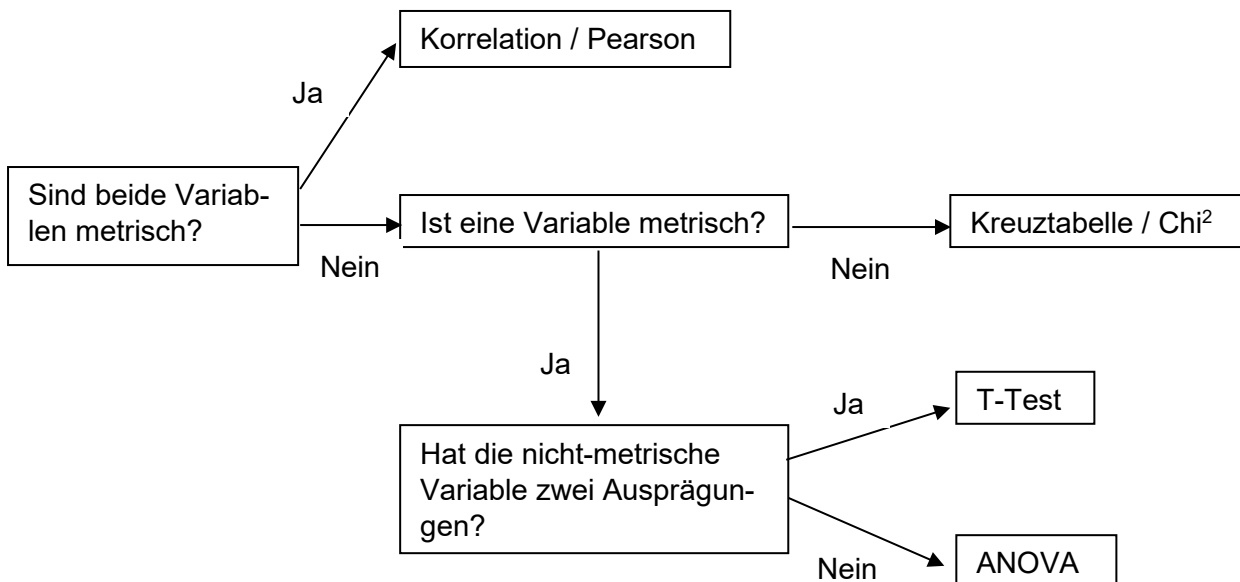
- Mit einer Kreuztabelle lassen sich die Zusammenhänge zwischen zwei nicht-metrischen Variablen einfach aufzeigen.
- Klicken Sie im Dateneditor auf *Analysieren > Deskriptive Statistiken > Kreuztabellen*.
- Im folgenden Fenster können Sie zunächst eine oder mehrere Variablen für Zeilen und Spalten der Tabelle auswählen. Üblicherweise steht dabei die unabhängige Variable in der Spalte und die abhängige Variable in der Zeile
- Unter *Zellen* können Sie zusätzliche Informationen auswählen, die angezeigt werden sollen. Zur Interpretation eines Zusammenhangs zwischen zwei Variablen empfiehlt sich hier die Auswahl von *spaltenweisen Prozentwerten*.

IX. INFERENZSTATISTIK

SPSS ermöglicht zahlreiche weitere statistische Auswertungen zu Gruppenunterschieden oder Zusammenhängen zwischen Variablen. Ein guter Ausgangspunkt hierfür ist das Angebot der [Methodenberatung der Universität Zürich](#).

Die folgenden Abschnitte beschreiben einige Formen der Dependenzanalyse bzw. von Signifikanztests, die häufig im Kontext von Abschlussarbeiten benötigt werden. Mittels Dependenzanalysen sollen mögliche Abhängigkeiten zwischen verschiedenen Variablen untersucht werden. Aus der Forschungsfrage bzw. den Hypothesen ergibt sich, ob Unterschiede oder Zusammenhänge zwischen den Variablen relevant sind. Mit sogenannten Signifikanztests lässt sich überprüfen, ob sich die Abhängigkeiten zwischen zwei Variablen mit einer gewissen Wahrscheinlichkeit (meist 95%) von der Stichprobe auf die Grundgesamtheit übertragen lassen.

Welcher Signifikanztest geeignet ist, hängt vom Skalenniveau der zu untersuchenden Variablen ab. Folgender Entscheidungsbaum hilft bei der Auswahl:



Korrelation

Mit einer Korrelation kann keine Aussage über Kausalzusammenhänge zwischen zwei Variablen getroffen werden. Untersucht wird lediglich, ob zwischen den Variablen ein linearer Zusammenhang besteht.

Eine Voraussetzung zur Durchführung der Analyse ist, dass die Variablen normalverteilt sind. Dies kann folgendermaßen mit einem Shapiro Wilk Test überprüft werden:

- Klicken Sie im Dateneditor auf *Analysieren > Deskriptive Statistiken > Explorative Datenanalyse*
- Wählen Sie die gewünschten Variablen für *Abhängige Variablen* aus.

- Klicken Sie auf *Diagramme* und wählen anschließend unter *Boxplots* die Option *Faktorstufen zusammen* aus, unter *Deskriptive Statistik* setzen Sie **keinen** Haken und bei *Normalverteilungsdiagramm mit Tests* setzen Sie einen Haken.
- Bestätigen Sie zuerst mit *Weiter* und dann mit *OK*.
- Die Tabelle *Tests auf Normalverteilung* im Ergebnisviewer zeigt das Ergebnis des Shapiro Wilk Tests an. Ist das Ergebnis **nicht signifikant**, ist eine Variable normalverteilt.
- Falls die Variablen nicht normalverteilt sind, ist statt der Korrelation ein nichtparametrischer Test zu wählen. Zum Beispiel der [Spearman-Rangkorrelationskoeffizient](#).

Eine weitere Voraussetzung ist ein linearer Zusammenhang zwischen den untersuchten Variablen. Hierzu kann ein Streudiagramm erstellt werden.

- Klicken Sie im Dateneditor auf *Grafik > Diagrammerstellung* und ziehen Sie anschließend das Element *Streudiagramm* aus der *Galerie* in die *Diagrammvorschau* (das weiße Feld über der Galerie)
- Ziehen Sie eine der benötigten Variablen aus *Variablen* auf die *Y-Achse* und die andere Variable auf die *X-Achse* der *Diagrammvorschau*. Bestätigen Sie mit *Okay*.
- Im Ergebnisviewer sehen Sie das Streudiagramm. Verläuft die Punktwolke tendenziell gerichtet in einer geraden Linie, deutet das auf eine mögliche Korrelation hin (s. beispielsweise Abbildungen 1 und 2). Bei einer anderweitig verteilten Punktwolke besteht kein linearer Zusammenhang (s. beispielsweise Abbildung 3).

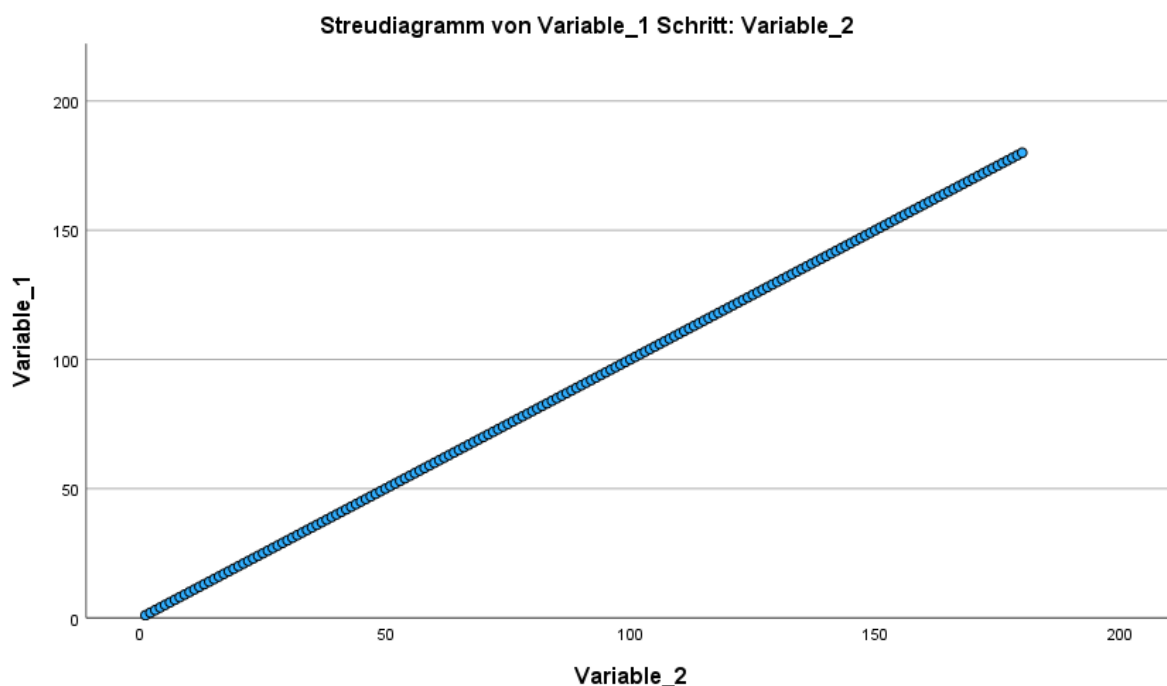


Abbildung 1. perfekter positiver linearer Zusammenhang zwischen zwei Beispielvariablen

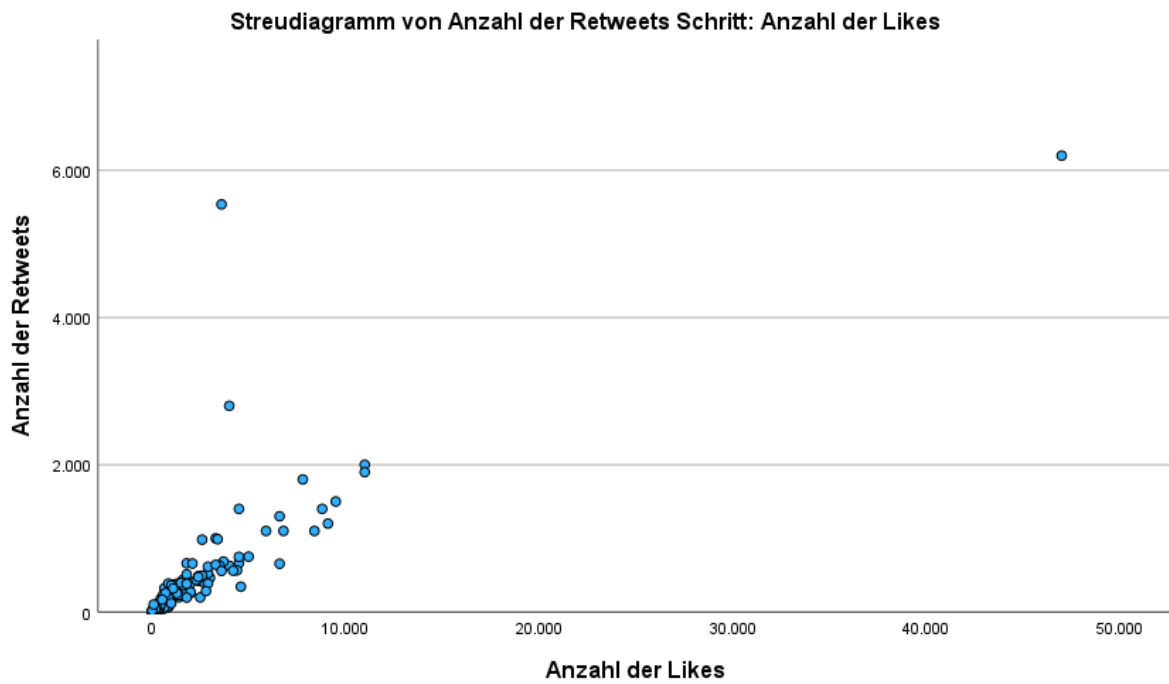


Abbildung 2. positiver linearer Zusammenhang zwischen der Anzahl der Retweets und der Anzahl der Likes von Tweets.

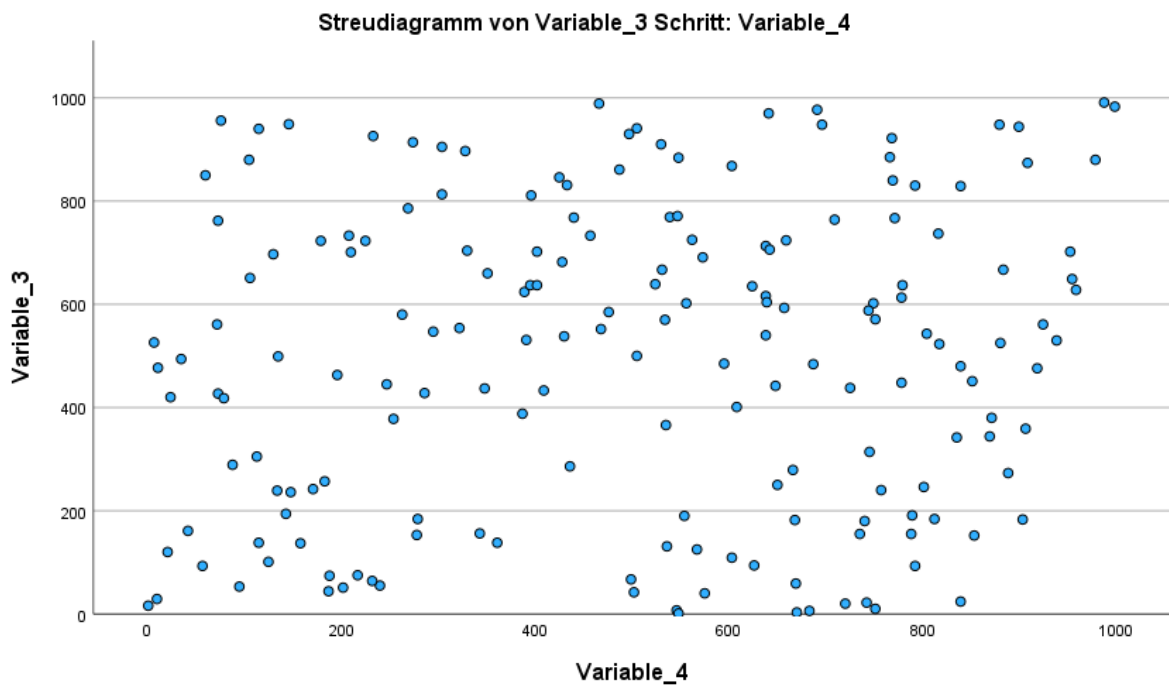


Abbildung 3. ungerichtete Punktwolke

- Falls kein linearer Zusammenhang besteht, ist statt der Korrelation [Spearman's-Rangkorrelationskoeffizient](#) als nichtparametrischer Test zu wählen

Sind alle Voraussetzungen erfüllt, kann die Korrelation berechnet werden:

- Klicken Sie im Dateneditor auf *Analysieren > Korrelation > Bivariat*.
- Wählen Sie zwei Variablen zur Berechnung der Korrelation aus
- Klicken Sie auf *Optionen* und Wählen Sie in diesem Fenster die Option *Mittelwerte und Standardabweichungen* im Feld *Statistik* sowie die Option *Paarweiser Fallausschluss* im Feld *Fehlende Werte*. Bestätigen Sie mit *Weiter*.
- Wählen Sie im Feld *Korrelationskoeffizienten* den Koeffizienten *Pearson* aus.
- Wählen Sie im Feld *Test auf Signifikanz* die Option *Zweiseitig* aus.
- Setzen Sie einen Haken bei *Signifikante Korrelationen markieren* und bestätigen Sie mit *OK*
- Im Output von SPSS erhalten Sie eine Tabelle, aus welcher der Korrelationskoeffizient nach Pearson, die Signifikanz des Ergebnisses sowie die Größe der Stichprobe hervorgehen.
- Ist das Ergebnis signifikant, so wird der Korrelationskoeffizient von SPSS mit einem oder mehreren Sternchen (*) markiert. Unterhalb der Tabelle wird dann, ebenfalls mit Sternchen markiert, das Signifikanzniveau angezeigt.
- Die richtige Interpretation des Korrelationskoeffizienten geht aus folgender Tabelle hervor:

Wert r	Bedeutung
0	Kein Zusammenhang
$r \geq -1$ und < 0	Negativer Zusammenhang
$r > 0$ und ≤ 1	Positiver Zusammenhang
$r > 0,3$	Geringer Zusammenhang
$r > 0,3$ und $< 0,5$	Mittlerer Zusammenhang
$r > 0,5$	Starker Zusammenhang

Kreuztabelle / Chi²-Test

Der Chi²-Test für Unabhängigkeit überprüft, ob zwei oder mehr Variablen statistisch unabhängig sind. Dabei wird ebenfalls kein Kausalzusammenhang untersucht.

- Klicken Sie im Dateneditor auf *Analysieren > Deskriptive Statistiken > Kreuztabellen*.
- Wählen Sie zwei Variablen zur Analyse aus
- Klicken Sie auf *Statistiken* und wählen Sie dort die Optionen *Chi-Quadrat* und *Phi und Cramer-V* aus. Bestätigen Sie mit *Weiter*
- Klicken Sie auf *Zellen* und wählen Sie dort die Optionen *Beobachtet* und *Erwartet* im Feld *Häufigkeiten* sowie die Option *Anzahl in den Zellen runden* im Feld *Nicht ganzzahlige Gewichtungen* aus. Bestätigen Sie mit *Weiter*.

- Bestätigen Sie mit *OK*.
- Im Output von SPSS erhalten Sie unter anderem eine Tabelle, die mit *Chi-Quadrat-Tests* überschrieben ist.
 - Anhand der Angabe in der Spalte *df* können Sie überprüfen, ob eine der Voraussetzungen für den Chi²-Test erfüllt ist. Nämlich, dass die Freiheitsgrade >1 sind. Ist diese Voraussetzung nicht erfüllt, kann alternativ die [Korrektur nach Yates](#) verwendet werden (gleiches gilt für eine Stichprobengröße zwischen 20 und 50).
 - Anhand der Angaben unter dieser Tabelle (mit einem „a.“ gekennzeichnet) können Sie überprüfen, ob eine weitere der Voraussetzungen für den Chi²-Test erfüllt ist. Nämlich, dass die erwarteten Zelhäufigkeiten >5 sind. Ist diese Voraussetzung nicht erfüllt, kann alternativ der [exakte Test nach Fisher](#) durchgeführt werden (gleiches gilt für eine Stichprobengröße unter 20).
 - Die Angabe *Wert* dient der Berechnung der Signifikanz, das Signifikanzniveau selbst geht aus der Spalte *Asymptotische Signifikanz (zweiseitig)* hervor.
- Im Output finden Sie zudem eine Tabelle namens *symmetrische Maße*. Aus dieser geht die Stärke des Zusammenhangs der beiden Variablen hervor. Dabei kann der Phi-Koeffizient nur bei dichotomen Variablen verwendet werden, Cramer-V hingegen bei jeder beliebigen Anzahl an Ausprägungen. Die Interpretation der Koeffizienten entspricht der Interpretation der Korrelationskoeffizienten (s. Abschnitt Korrelation). Cramer-V kann dabei jedoch nur Werte im Bereich 0 bis 1 annehmen.

T-Test

Der T-Test überprüft, ob sich die Mittelwerte zweier unabhängiger Stichproben unterscheiden. Die abhängige Variable ist dabei metrisch, die unabhängige Variable ist dichotom. Hat die unabhängige Variable mehr als zwei Ausprägungen, so können die Daten entweder für den T-Test getrennt werden, so dass zwei zu vergleichende Gruppen im Datensatz entstehen oder es kann eine ANOVA gerechnet werden (siehe nächster Abschnitt).

Eine Voraussetzung zur Durchführung der Analyse ist, dass die abhängige Variable normalverteilt ist. Dies kann folgendermaßen mit einem Shapiro Wilk Test überprüft werden:

- Klicken Sie im Dateneditor auf *Analysieren > Deskriptive Statistiken > Explorative Datenanalyse*
- Wählen Sie die gewünschte Variable für *Abhängige Variablen* aus.
- Wählen Sie die unabhängige Variable für *Faktorenliste* aus.
- Klicken Sie auf *Diagramme* und wählen anschließend unter *Boxplots* die Option *Faktorstufen zusammen* aus, unter *Deskriptive Statistik* setzen Sie **keinen** Haken und bei *Normalverteilungsdiagramm mit Tests* setzen Sie einen Haken.
- Bestätigen Sie zuerst mit *Weiter* und dann mit *OK*.

- Die Tabelle *Tests auf Normalverteilung* im Ergebnisviewer zeigt das Ergebnis des Shapiro Wilk Tests an. Ist das Ergebnis **nicht signifikant**, ist eine Variable normalverteilt.
- Falls jede der zu vergleichenden Gruppen mindestens 30 Fälle enthält, kann der T-Test unabhängig von einer Normalverteilung der Variable durchgeführt werden. Bei weniger als 30 Fällen und einer fehlenden Normalverteilung, kann alternativ der [Mann-Whitney-U-Test](#) durchgeführt werden.

Falls der T-Test durchgeführt werden kann, geschieht das wie folgt:

- Klicken Sie im Dateneditor auf *Analysieren > Mittelwerte und Proportionen vergleichen > t-Test bei unabhängigen Stichproben*.
- Wählen Sie Ihre abhängige Variable(n) für das Feld *Testvariable(n)* aus.
- Wählen Sie Ihre unabhängige Variable für das Feld *Gruppierungsvariable* aus.
- Klicken Sie auf Optionen. Definieren Sie dort den *Prozentsatz Konfidenzintervall* mit 95% und wählen Sie im Feld *Fehlende Werte* die Option *Fallausschluss Analyse für Analyseaus*. Bestätigen Sie mit *Weiter*.
- Klicken Sie auf *Gruppen definieren*. Bei einer dichotomen unabhängigen Variable wählen Sie *angegebene Werte verwenden* und geben für jede Gruppe den dazugehörigen Code ein (Haben Sie zum Beispiel nach dem Geschlecht gefragt und für ‚weiblich‘ die 1 codiert sowie für ‚männlich‘ die 2, dann geben Sie in das Feld *Gruppe 1* die 1 ein und in das Feld *Gruppe 2* die 2). Bei einer polytomen unabhängigen Variable wählen Sie *Trennwert* und geben den Code derjenigen Ausprägung ein, anhand derer der Datensatz in zwei Gruppen geteilt werden soll. Die Gruppen umfassen dann alle Daten \geq Trennwert sowie alle Daten $<$ Trennwert. Bestätigen Sie mit *Weiter*.
- Bestätigen Sie mit *OK*.
- Im Output von SPSS betrachten Sie zunächst die zweite Spalte der Ergebnistabelle. Diese zeigt Ihnen die *Signifikanz des Levene-Tests der Varianzgleichheit* an. Ist das Ergebnis **nicht signifikant**, so liegt Varianzhomogenität vor. In diesem Fall geht das Ergebnis des T-Tests sowie dessen Signifikanz aus der Zeile *Varianzen sind gleich* hervor. Ist das Ergebnis des Levene-Tests signifikant, so ist die Zeile *Varianzen sind nicht gleich* relevant. Die wichtigsten Spalten, deren Werte dann auch üblicherweise in Ergebnisberichten angegeben werden, sind *T* und *Signifikanz Zweiseitiges p*.

ANOVA (Auch Einfaktorielle Varianzanalyse)

Die ANOVA überprüft, ob sich die Mittelwerte zweier unabhängiger Stichproben unterscheiden. Die abhängige Variable ist dabei metrisch, die unabhängige Variable ist polytom.

Eine Voraussetzung zur Durchführung der Analyse ist, dass die abhängige Variable normalverteilt ist. Dies kann folgendermaßen mit einem Shapiro Wilk Test überprüft werden:

- Klicken Sie im Dateneditor auf *Analysieren > Deskriptive Statistiken > Explorative Datenanalyse*
- Wählen Sie die gewünschte Variable für *Abhängige Variablen* aus.
- Wählen Sie die unabhängige Variable für *Faktorenliste* aus.
- Klicken Sie auf *Diagramme* und wählen anschließend unter *Boxplots* die Option *Faktorstufen zusammen* aus, unter *Deskriptive Statistik* setzen Sie **keinen** Haken und bei *Normalverteilungsdiagramm mit Tests* setzen Sie einen Haken.
- Bestätigen Sie zuerst mit *Weiter* und dann mit *OK*.
- Die Tabelle *Tests auf Normalverteilung* im Ergebnisviewer zeigt das Ergebnis des Shapiro Wilk Tests an. Ist das Ergebnis **nicht signifikant**, ist eine Variable normalverteilt.
- Falls jede der zu vergleichenden Gruppen mindestens 25 Fälle enthält, kann die ANOVA unabhängig von einer Normalverteilung der Variable durchgeführt werden. Bei weniger als 25 Fällen und einer fehlenden Normalverteilung, kann alternativ der [Kruskal-Wallis-Test](#) durchgeführt werden.

Falls die ANOVA durchgeführt werden kann, geschieht das wie folgt:

- Klicken Sie im Dateneditor auf *Analysieren > Allgemeines lineares Modell > Univariat*.
- Wählen Sie Ihre abhängige Variable für das Feld *Abhängige Variable* aus.
- Wählen Sie Ihre unabhängige Variable für das Feld *Feste Faktoren* aus.
- Klicken Sie auf *Post hoc* und wählen Sie Sie dort im Dialogfenster Ihre unabhängige Variable für das Feld *Post-hoc-Tests für* aus. Wählen Sie im Feld *Varianzgleichheit angenommen* die Option *Bonferroni* aus. Bestätigen Sie mit *Weiter*.
- Klicken Sie auf *Optionen* und wählen Sie Sie dort im Dialogfenster im Feld *Anzeige* die Optionen *Deskriptive Statistiken*, *Schätzungen der Effektgröße* und *Homogenitätstests* aus. Bestätigen Sie mit *Weiter*.
- Bestätigen Sie mit *OK*.
- Im SPSS-Output betrachten Sie zunächst die Tabelle *Levene-Test auf Gleichheit der Fehlervarianzen*. Ist dessen Ergebnis **nicht signifikant**, so ist mit der sog. Varianzhomogenität eine Voraussetzung für die ANOVA gegeben. Bei einem signifikanten Ergebnis kann alternativ der Brown-Forsythe-Test oder der Welch-Test durchgeführt werden (In SPSS unter *Analysieren > Mittelwerte und Proportionen vergleichen > Einfaktorielle Varianzanalyse* durch einem Klick auf *Optionen* zu finden).

- Nun betrachten Sie im Output die Tabelle *Tests der Zwischensubjekteffekte*. Hier entnehmen Sie aus der Zeile *Korrigiertes Modell* die Werte der Spalten *F*, *Sig.* und *Partielles Eta-Quadrat*. Ersterer ist das Testergebnis, der zweite gibt die Signifikanz des Ergebnisses an und letzterer die Effektgröße. Ein partieller Eta-Quadrat-Wert von zum Beispiel ,416 bedeutet, dass 41,6% der Variation in der abhängigen Variable durch die unabhängige Variable erklärt werden können. In der Fußnote *a* zeigt der Wert *korrigiertes R-Quadrat* schließlich die Güte Ihres Modells an. Ein Wert von ,343 bedeutet zum Beispiel, dass 34,3% der Streuung um den Gesamtmittelwert erklärt werden können.
- Die Tabelle *Mehrere Vergleiche* zeigt Ihnen das Ergebnis des Post-Hoc-Tests nach Bonferroni an. Hier werden alle Gruppen, die Sie mittels Ihrer unabhängigen Variable gebildet haben, paarweise darauf verglichen, ob zwischen ihnen signifikante Unterschiede bestehen. Diese gehen jeweils aus der Spalte *Sig.* hervor.