

www.statistikprofi.de

Leitfaden

Statistische Ergebnisdarstellung und
wissenschaftliches Schreiben nach APA 7

Ein Arbeitsdokument für Abschlussarbeiten und empirische Berichte

Römer & Römer Institut für Statistik · Statistikprofi.de

Inhaltsverzeichnis

1 Hinweise zur Nutzung dieses Dokuments.....	4
2 APA-Grundregeln der Zahlendarstellung.....	4
2.1 Grundregeln.....	4
2.2 Dezimaltrennzeichen: Punkt oder Komma	4
2.3 Ausschreiben von Zahlen.....	5
3 Symbol- und Abkürzungsverzeichnis.....	5
4 Verfahrenswahl nach Skalenniveau und Fragestellung	7
5 Hypothesen, Signifikanz und Teststärke.....	8
5.1 Hypothesen formulieren.....	8
5.2 Signifikanzniveau und multiples Testen	8
5.3 Teststärke und Stichprobenplanung.....	8
6 Voraussetzungen und Datenvorbereitung.....	9
6.1 Fehlende Werte	9
6.2 Ausreißer	9
6.3 Verteilungsannahmen und Robustheit	9
6.4 Diagnostische Plots im Überblick	10
7 Effektstärkemaße und Konventionen.....	10
7.1 Reliabilität und Eignung für Faktorenanalysen	10
8 Verfahren im Detail.....	12
8.1 Deskriptive Statistik.....	12
8.2 Korrelationsanalyse	12
8.3 t-Test für unabhängige Stichproben	13
8.4 t-Test für verbundene Stichproben	14
8.5 Einfache lineare Regression	14
8.6 Multiple lineare Regression.....	15
8.7 Mediationsanalyse	16
8.8 ANOVA mit Messwiederholung (RM-ANOVA).....	16
8.9 Gemischte ANOVA (Mixed ANOVA)	17
8.10 MANOVA	18
8.11 Binäre logistische Regression	18
8.12 Exploratorische Faktorenanalyse	19
9 Den Ergebnisteil schreiben.....	20
9.1 Aufbau und Reihenfolge.....	20
9.2 Zeitform	20
9.3 Was hineingehört – und was nicht	20
9.4 Umgang mit nicht signifikanten Ergebnissen	20
9.5 Abgrenzung von Ergebnis und Diskussion	20
9.6 Quantitative und qualitative Ergebnisse	21

9.7 Kommentierter Musterabschnitt	21
10 Hinweise zu weiteren Kapiteln	22
10.1 Methodenteil	22
10.2 Diskussion und Fazit	22
11 Tabellen und Abbildungen in Word	22
11.1 Beschriftungen und automatische Verzeichnisse	22
11.2 APA-konforme Tabellengestaltung	23
11.3 Abbildungen	23
12 Einsatz von KI-Werkzeugen	23
12.1 Transparenz und Kennzeichnung	23
12.2 Zitation von KI nach APA	24
12.3 Datenschutz	24
13 Häufige Fehler – Checkliste	24
14 Literaturverzeichnis	25

1 Hinweise zur Nutzung dieses Dokuments

Dieses Dokument bündelt die wichtigsten Regeln und Vorlagen für die Darstellung statistischer Ergebnisse nach den Richtlinien der American Psychological Association (APA, 7. Auflage). Es richtet sich an Studierende und Promovierende, die einen empirischen Ergebnisteil verfassen, und dient zugleich als Nachschlagewerk für Notation, Tabellengestaltung, Berichtsschemata und Formulierungshilfen.

Das Dokument ist in drei Bereiche gegliedert: ein Regel- und Notationsteil (Abschnitte 2 bis 7), ein Verfahrensteil mit Beispieltabellen und Musterformulierungen je Analyse (Abschnitt 8) sowie ein Schreib- und Formatteil zum Aufbau des Ergebnisteils, zur Arbeit mit Word und zur Kennzeichnung von KI (Abschnitte 9 bis 13).

Wichtig

Hochschulen und Betreuende setzen eigene Vorgaben. Diese gehen den hier genannten Konventionen im Zweifel vor. Vor der Abgabe sollten die Vorgaben des jeweiligen Instituts geprüft und mit diesem Leitfaden abgeglichen werden.

Alle Beispieltabellen enthalten frei erfundene Werte und dienen ausschließlich der Veranschaulichung des Formats.

2 APA-Grundregeln der Zahlendarstellung

Die folgenden Regeln gelten für die Darstellung statistischer Kennwerte gemäß dem Publication Manual (American Psychological Association, 2020).

2.1 Grundregeln

- Statistische Kennwerte werden mit zwei Dezimalstellen angegeben; p -Werte mit drei Dezimalstellen.
- Kennwerte, deren Betrag 1 nicht überschreiten kann (z. B. p , r , β), werden ohne führende Null angegeben. Beispiel: $r = .38$, $p = .041$.
- Kennwerte, die größer als 1 werden können (z. B. M , SD , t , F), behalten die führende Null. Beispiel: $M = 0.94$.
- Viele Kennwerte werden kursiv gesetzt (siehe Abschnitt 3). Griechische Buchstaben (α , β , η , χ) werden nicht kursiv gesetzt.
- Das Gleichheitszeichen erhält je ein Leerzeichen davor und danach: $p = .041$.
- Zwischen Zahl und Einheit steht ein Leerzeichen: 8 %, 14 kg, 5 h.
- Sehr kleine p -Werte werden als $p < .001$ berichtet, nicht als $p = .000$.

2.2 Dezimaltrennzeichen: Punkt oder Komma

Hier gibt es zwei legitime Konventionen, die nicht vermischt werden dürfen:

- *Punkt* (.) ist das APA-Standardtrennzeichen und in der internationalen psychologischen Literatur üblich. Beispiel: $M = 3.95$, $p = .041$.
- *Komma* (,) entspricht der deutschen Rechtschreibung und wird von vielen deutschsprachigen Hochschulen verlangt. Beispiel: $M = 3,95$, $p = ,041$.

Empfehlung

Maßgeblich ist die Vorgabe der Hochschule bzw. der Betreuungsperson. Liegt keine Vorgabe vor, ist für streng APA-orientierte Arbeiten der Punkt zu wählen, für deutschsprachige Arbeiten häufig das Komma.

Entscheidend ist die Konsistenz: Im gesamten Dokument – in Fließtext, Tabellen und Abbildungen – muss durchgängig dasselbe Trennzeichen verwendet werden.

In den Beispielen dieses Leitfadens wird einheitlich der Punkt verwendet. Bei Komma-Vorgabe sind alle Werte entsprechend umzustellen.

2.3 Ausschreiben von Zahlen

Im Fließtext werden Zahlen von null bis neun grundsätzlich ausgeschreiben, ab zehn als Ziffer dargestellt. Nicht ausgeschrieben werden hingegen:

- alle Zahlen ab 10;
- exakte Werte, Messwerte und Statistikkenwerte (z. B. 3 %, 5 kg, 2.34, 4-Punkte-Skala);
- Zahlen unmittelbar vor einer Maßeinheit (3 cm, 5 h);
- Zahlen vor oder nach statistischen Symbolen ($n = 9$);
- Zahlen in Tabellen, Abbildungen und deren Beschriftungen;
- Alters-, Prozent-, Zeit- und Datumsangaben (3 Jahre, 5 %, Tag 3, 2. Mai);
- Punkte einer Likert-Skala (Antwortoption 1 bis 5);
- Serien gleichartiger Zahlen (Messpunkte 1, 3, 7 und 9).

3 Symbol- und Abkürzungsverzeichnis

Die folgenden Tabellen führen die gebräuchlichen statistischen Symbole und Abkürzungen mit ihrer Bedeutung und Kursivsetzung auf. Symbole, die nicht Teil der APA-Standardnotation sind, sowie Abkürzungen müssen bei der ersten Verwendung im Text bzw. in der Tabellenanmerkung definiert werden.

Tabelle 1
Deskriptive Kennwerte

Symbol	Bedeutung	Kursiv
<i>M</i>	Stichprobenmittelwert	Ja
<i>SD</i>	Standardabweichung	Ja
<i>n</i>	Anzahl der Fälle (in der Regel Teilstichprobe)	Ja
<i>N</i>	Gesamtzahl der Fälle	Ja
β_2	Kurtosis (3 = Normalverteilung)	Nein
γ_2	Exzess der Kurtosis ($= \beta_2 - 3$; 0 = Normal)	Nein
<i>Sk</i>	Schiefte	Ja
<i>f</i>	Häufigkeit	Ja
<i>Mdn</i>	Median	Ja
<i>s</i>	Stichprobenstandardabweichung	Ja
s^2	Stichprobenvarianz (unverzerrte Schätzung)	Ja
S^2	Stichprobenvarianz (verzerrte Schätzung)	Ja
Σ	Summe	Nein

μ	Populationsmittelwert; Erwartungswert	Nein
σ	Populationsstandardabweichung	Nein
σ^2	Populationsvarianz	Nein
ΔM	Mittelwertsdifferenz	Ja
Min., Max.	Minimum, Maximum	Nein
Spannweite	Wertebereich	Nein
IQR	Interquartilsabstand	Nein

Anmerkung. Symbole, die nicht zur APA-Standardnotation gehören (z. B. ΔM , IQR), sind bei erster Nennung zu definieren.

Tabelle 2
Inferenzstatistische Kennwerte

Symbol	Bedeutung	Kursiv
F	F-Verteilung; Fishers F-Verhältnis	Ja
t	Studentsche t-Verteilung	Ja
df	Freiheitsgrade	Ja
p	Signifikanzwert (Überschreitungswahrscheinlichkeit)	Ja
α	Wahrscheinlichkeit eines Typ-I-Fehlers; Cronbachs Alpha	Nein
β	Wahrscheinlichkeit eines Typ-II-Fehlers; standardisierter Regressionskoeffizient	Nein
b	unstandardisierter Regressionskoeffizient	Ja
χ^2	Chi-Quadrat-Verteilung	Nein
Z	standardisierter Wert	Ja
SE	Standardfehler	Ja
SE_M	Standardfehler des Mittelwerts (SEM)	Ja
U	Mann-Whitney-Teststatistik	Ja
Wilcoxon-W	Wilcoxon-Teststatistik	Nein
Kruskal-Wallis-H	Kruskal-Wallis-Teststatistik	Nein
r	Pearson-Produkt-Moment-Korrelationskoeffizient	Ja
r^2	Bestimmtheitsmaß	Ja
R^2	Anteil aufgeklärter Varianz (multipel)	Ja
τ	Kendalls Rangkorrelationskoeffizient	Nein
ρ	Spearman-Rangkorrelation	Ja
d	Cohens d	Ja
g	Hedges' g (Effektgröße)	Ja
η^2	Eta-Quadrat	Nein
ηp^2	partielltes Eta-Quadrat	Nein

k	Anzahl der Studien in einer Meta-Analyse	Ja
r_{ti}	Trennschärfe (Item-Skala-Korrelation)	Ja
Λ	Faktorladung	Nein

Anmerkung. Griechische Buchstaben werden nicht kursiv gesetzt, lateinische Symbole überwiegend kursiv.

Tabelle 3
Weitere Abkürzungen

Abkürzung	Bedeutung	Kursiv
OR	Odds Ratio	Ja
CI	Konfidenzintervall	Nein
LL	Untergrenze (Lower Limit)	Ja
UL	Obergrenze (Upper Limit)	Ja
ANOVA	Varianzanalyse	Nein
ANCOVA	Kovarianzanalyse	Nein
MANOVA	multivariate Varianzanalyse	Nein
MANCOVA	multivariate Kovarianzanalyse	Nein
GLM	verallgemeinertes lineares Modell	Nein
RR	relatives Risiko	Ja
HR	Hazard Ratio	Ja
Q	Test auf Homogenität der Effektgrößen	Ja
H_0	Nullhypothese	Nein
H_1 (H_a)	Alternativhypothese	Nein
ns	nicht statistisch signifikant	Ja
MS	Mittel der Quadrate	Ja
SS (Typ III)	Quadratsumme (Typ III); nur SS kursiv	teils

Anmerkung. Abkürzungen für Verfahren (ANOVA, GLM) sind bei erster Nennung auszuschreiben.

4 Verfahrenswahl nach Skalenniveau und Fragestellung

Die Wahl des statistischen Verfahrens richtet sich nach der Fragestellung (Unterschied, Zusammenhang, Vorhersage) sowie nach dem Skalenniveau der abhängigen und unabhängigen Variablen. Nachfolgend findet sich eine Orientierungshilfe für die häufigsten Konstellationen.

Tabelle 4
Orientierung zur Verfahrenswahl

Fragestellung	UV	AV	Verfahren
Unterschied, 2 unabh. Gruppen	nominal (2)	metrisch	t-Test für unabhängige Stichproben
Unterschied, 2 verbundene Messungen	nominal (2)	metrisch	t-Test für verbundene Stichproben
Unterschied, ≥ 3 Gruppen	nominal (≥ 3)	metrisch	einfaktorielle ANOVA

Unterschied über Messzeitpunkte	Messwiederholung	metrisch	RM-ANOVA
Gruppen × Messwiederholung	gemischt	metrisch	gemischte ANOVA
mehrere AV gleichzeitig	nominal	mehrere metrisch	MANOVA
Zusammenhang zweier Variablen	metrisch	metrisch	Pearson-Korrelation
Zusammenhang, ordinal/nicht normal	ordinal	ordinal	Spearman / Kendall
Vorhersage, eine UV	metrisch	metrisch	einfache lineare Regression
Vorhersage, mehrere UV	metrisch/nominal	metrisch	multiple lineare Regression
Vorhersage binärer AV	beliebig	binär	binäre logistische Regression
Wirkpfad über Mediator	metrisch	metrisch	Mediationsanalyse
Dimensionsreduktion / Skalen	metrisch	—	Faktorenanalyse

Anmerkung. UV = unabhängige Variable; AV = abhängige Variable. Bei verletzten Voraussetzungen metrischer Verfahren sind verteilungsfreie Alternativen (z. B. Mann-Whitney-U, Wilcoxon, Kruskal-Wallis) zu prüfen.

5 Hypothesen, Signifikanz und Teststärke

5.1 Hypothesen formulieren

Jeder inferenzstatistische Test prüft eine Nullhypothese (H_0) gegen eine Alternativhypothese (H_1). Die H_0 postuliert in der Regel keinen Effekt bzw. keinen Unterschied, die H_1 den interessierenden Effekt. Hypothesen können gerichtet oder ungerichtet formuliert werden:

- *ungerichtet (zweiseitig)*: Es wird lediglich ein Unterschied bzw. Zusammenhang erwartet, ohne Richtung. Beispiel: Die Gruppen unterscheiden sich im Mittelwert.
- *gerichtet (einseitig)*: Es wird zusätzlich eine Richtung erwartet. Beispiel: Gruppe A erzielt einen höheren Mittelwert als Gruppe B.

Eine gerichtete Hypothese rechtfertigt einen einseitigen Test, der bei zutreffender Richtung mehr Teststärke besitzt. Einseitiges Testen ist jedoch vorab und theoretisch zu begründen und im Methodenteil festzulegen; eine nachträgliche Wahl ist nicht zulässig.

5.2 Signifikanzniveau und multiples Testen

- Das Signifikanzniveau (α) wird konventionell auf .05 festgelegt und vor der Analyse bestimmt. Ein Ergebnis gilt als statistisch signifikant, wenn $p < \alpha$.
- Statistische Signifikanz ist nicht gleich praktische Bedeutsamkeit. Daher sind stets Effektstärken (Abschnitt 7) und – wo möglich – Konfidenzintervalle zu berichten.
- Werden viele Tests an denselben Daten gerechnet, steigt die Wahrscheinlichkeit falsch-positiver Befunde (Alphafehler-Kumulierung). Eine Korrektur ist angezeigt, z. B. nach Bonferroni (Aufteilung von α auf die Anzahl der Tests) oder über Verfahren zur Kontrolle der False Discovery Rate.

5.3 Teststärke und Stichprobenplanung

Die Teststärke (Power) ist die Wahrscheinlichkeit, einen tatsächlich vorhandenen Effekt aufzudecken ($1 - \beta$); konventionell wird mindestens .80 angestrebt. Vor der Erhebung sollte über eine A-priori-Poweranalyse der erforderliche Stichprobenumfang bestimmt werden – abhängig von Signifikanzniveau, gewünschter Power und erwarteter Effektstärke. Hierfür eignet sich das Programm GPower (Faul et al., 2007).

Beispieltext

Auf Basis einer A-priori-Poweranalyse mit GPower (Faul et al., 2007) wurde für einen zweiseitigen *t*-Test bei einem mittleren Effekt ($d = 0.50$), einem Signifikanzniveau von $\alpha = .05$ und einer angestrebten Teststärke von $1 - \beta = .80$ ein erforderlicher Gesamtstichprobenumfang von $N = 128$ ermittelt.

Hinweis zur Verwendung

Die A-priori-Poweranalyse gehört in den Methodenteil. Im Ergebnisteil wird sie nicht wiederholt; dort wird lediglich der realisierte Stichprobenumfang berichtet.

6 Voraussetzungen und Datenvorbereitung

Vor der Hypothesenprüfung werden die Daten aufbereitet und die Verfahrensvoraussetzungen geprüft. Vorgehen und Ergebnis dieser Prüfung gehören in den Ergebnisteil (oder, je nach Vorgabe, in den Methodenteil) und werden vor den eigentlichen Hypothesentests berichtet.

6.1 Fehlende Werte

- Umfang und Muster fehlender Werte sind zu prüfen und zu berichten.
- Beim *listenweisen Ausschluss* werden Fälle mit fehlenden Werten vollständig entfernt; beim *paarweisen Ausschluss* nur für die jeweils betroffene Analyse. Das gewählte Vorgehen ist anzugeben.
- Bei umfangreichen Ausfällen können Imputationsverfahren (z. B. multiple Imputation) sinnvoll sein; ihre Anwendung ist zu begründen.

6.2 Ausreißer

- Ausreißer werden über Boxplots, standardisierte Residuen, Cooks Distanz und Hebelwerte identifiziert (siehe Plotübersicht in Abschnitt 6.4).
- Der Ausschluss von Ausreißern ist die Ausnahme und stets zu begründen; sinnvoll ist die Angabe der Ergebnisse mit und ohne Ausreißer (Sensitivitätsanalyse).
- Grenzwerte (Auswahl): Cooks Distanz > 1 (Larose, 2005); strengere Schwellen $4/(n - k)$ bzw. $4/n$ werden ebenfalls verwendet. Hebelwerte: Faustregel $(2 \cdot p)/n$ bei $n - p > 50$ (Salkind et al., 2010).

6.3 Verteilungsannahmen und Robustheit

- Die Normalverteilungsannahme bezieht sich bei *t*-Test und Regression auf die Residuen bzw. Differenzwerte, nicht auf die Rohdaten.
- *t*-Tests und Regressionsmodelle sind bei großen Stichproben robust gegenüber Verletzungen der Normalverteilungsannahme. Lineare Modelle gelten ab etwa $N = 65$ als robust, ab $N = 500$ als sehr robust gegenüber auch deutlichen Verletzungen (Lumley et al., 2002).
- Bei ungleichen Gruppengrößen oder heterogenen Varianzen ist dem Welch-*t*-Test der Vorzug zu geben (Delacre et al., 2017).

- Tests auf Verteilungsannahmen: Shapiro-Wilk (Shapiro & Wilk, 1965) und Kolmogorov-Smirnov (Stephens, 1974) auf Normalverteilung, Levene (Levene, 1960) auf Varianzhomogenität, Breusch-Pagan (Breusch & Pagan, 1979) auf Heteroskedastizität, Durbin-Watson (Durbin & Watson, 1950) auf Unabhängigkeit der Residuen.

6.4 Diagnostische Plots im Überblick

Tabelle 5

Diagnostische Plots und ihre Interpretation

Plot	Geprüfte Annahme	Erwartetes Muster
Histogramm der Residuen	Normalverteilung der Residuen	symmetrischer Glockenverlauf
Normal-Q-Q- bzw. P-P-Plot	Normalverteilung der Residuen	Punkte nahe der Diagonalen
Residuen vs. vorhergesagte Werte	Homoskedastizität / Linearität	zufällige Wolke, keine Trichterform
Residuen vs. einzelne Prädiktoren	Linearität	kein systematisches Muster
Cooks-Distanz-Diagramm	Einflussreiche Fälle	keine Werte deutlich über Schwellenwert
Hebelwert-Plot (Leverage)	Einfluss einzelner Fälle	wenige Punkte mit hoher Hebelwirkung

Anmerkung. ZRESID = standardisierte Residuen; ZPRED = standardisierte vorhergesagte Werte.

7 Effektstärkemaße und Konventionen

Lesart der Konventionen

Die angegebenen Werte sind keine Schwellenwerte, sondern Richtbereiche. Liegt z. B. f im Bereich um .25, spricht man von einem mittleren Effekt.

Tabelle 6

Konventionen gängiger Effektstärkemaße

Maß	klein	mittel	groß	Quelle
Cohens f	.10	.25	.40	Cohen, 2009, S. 355
Cohens d / d_z	.20	.50	.80	Cohen, 2009, S. 40
Pearsons r	.10	.30	.50	Cohen, 2009, S. 83
Hedges' g	.05	.15	.25	Cohen, 2009, S. 151
partiell η^2	$\approx .01$	$\approx .06$	$\approx .14$	abgel. aus Cohen, 2009
R^2	$\approx .02$	$\approx .13$	$\approx .26$	abgel. aus Cohen, 2009

Anmerkung. η^2 und R^2 ergeben sich über die Beziehung zu f ($\eta^2 = f^2 / (1 + f^2)$).

7.1 Reliabilität und Eignung für Faktorenanalysen

- **Cronbachs Alpha** (Cronbach, 1951): exzellent (.90–1.00), gut (.80–.89), akzeptabel (.70–.79), fragwürdig (.60–.69), schlecht (.50–.59), inakzeptabel (< .50) (Taber, 2018). Sehr hohe Werte (> .95) können auf Redundanz hindeuten.

- *KMO-Maß* (Kaiser & Rice, 1974): $\geq .90$ hervorragend, $\geq .80$ sehr gut, $\geq .70$ gut, $\geq .60$ mittelmäßig, $\geq .50$ schlecht, $< .50$ inakzeptabel.

8 Verfahren im Detail

Für jedes Verfahren werden die Menüpfade in SPSS und die entsprechenden Funktionen in R, die zu prüfenden Voraussetzungen, eine APA-konforme Beispieltabelle, das Berichtsschema sowie ein kommentierter Beispieltext angegeben. Die Beispielwerte sind frei gewählt.

8.1 Deskriptive Statistik

Menüpfade

SPSS: Analysieren → Deskriptive Statistiken → Häufigkeiten / Deskriptive Statistik / Explore

R: summary(); psych::describe(); für Skalen psych::alpha()

Tabelle 7

Deskriptive Kennwerte der eingesetzten Skalen

Skala	<i>M</i>	<i>SD</i>	<i>Sk</i>	<i>Y₂</i>	<i>α</i>
Mitarbeiterunterstützung	3.20	0.91	−0.30	−0.34	.91
Pulsation	3.15	0.93	−0.41	−0.10	.83
Sequenzielle Implementierung	3.33	0.88	−0.32	−0.02	.80
Digitaler Erfolg (gesamt)	3.58	0.88	−0.55	0.20	.95

Anmerkung. *M* = Mittelwert; *SD* = Standardabweichung; *Sk* = Schiefe; *Y₂* = Exzess; *α* = Cronbachs Alpha.

Beispieltext

Beispieltext

Die deskriptiven Kennwerte der eingesetzten Skalen sind in Tabelle 7 dargestellt. Die Skalenmittelwerte lagen zwischen *M* = 3.15 (*SD* = 0.93) für Pulsation und *M* = 3.58 (*SD* = 0.88) für digitalen Erfolg. Die internen Konsistenzen waren mit Werten zwischen *α* = .80 und *α* = .95 zufriedenstellend bis exzellent. Schiefe und Exzess lagen in einem für die Normalverteilungsannahme unkritischen Bereich.

Kommentar zum Beispieltext

Im Text werden die Kennwerte nicht vollständig wiederholt, sondern Spannweite und Auffälligkeiten zusammengefasst, mit Verweis auf die Tabelle.

Die Bewertung der Reliabilität (zufriedenstellend, exzellent) ist eine etablierte Konvention und keine inhaltliche Interpretation – diese bleibt der Diskussion vorbehalten.

8.2 Korrelationsanalyse

Menüpfade

SPSS: Analysieren → Korrelation → Bivariat (Pearson bzw. Spearman)

R: cor.test(x, y); für Matrizen psych::corr.test()

Voraussetzungen

- Pearson: metrische Variablen, linearer Zusammenhang, annähernde Normalverteilung.
- Bei ordinalen Daten oder Verletzung der Annahmen: Spearmans *ρ* oder Kendalls *τ*.

Tabelle 8

Interkorrelationen der Skalen

Variable	1	2	3	4
1 Nachhaltigkeit	—			

2 FOMO	-.28	—		
3 Preisorientierung	-.09	.13	—	
4 Kaufintention	.42	.05	-.01	—

Anmerkung. $p < .05$; $p < .01$.

Berichtsschema

$r(df) = .XX$, $p = .XXX$. Die Freiheitsgrade betragen $N - 2$.

Beispieltext

Beispieltext

Zwischen Nachhaltigkeit und Kaufintention bestand ein signifikanter mittlerer positiver Zusammenhang, $r(93) = .42$, $p < .001$. Nachhaltigkeit korrelierte zudem negativ mit FOMO, $r(93) = -.28$, $p = .006$. Zwischen Preisorientierung und Kaufintention zeigte sich kein bedeutsamer Zusammenhang, $r(93) = -.01$, $p = .925$.

Kommentar zum Beispieltext

Richtung (positiv/negativ), Stärke (mittel, nach Konvention) und Signifikanz werden benannt, ohne die Ursachen zu deuten.

Auch der nicht signifikante Befund wird berichtet – Vollständigkeit ist Pflicht.

8.3 t-Test für unabhängige Stichproben

Menüpfade

SPSS: Analysieren → Mittelwerte vergleichen → t-Test bei unabhängigen Stichproben

R: `t.test(y ~ gruppe)` (Welch-Test als Voreinstellung); `car::leveneTest()`; `effectsize::cohens_d()`

Voraussetzungen

- Normalverteilung in den Gruppen (bei großen Stichproben robust); Prüfung über Shapiro-Wilk.
- Varianzhomogenität über den Levene-Test: ist er signifikant, wird die Welch-Korrektur (untere Zeile in SPSS) berichtet, andernfalls der klassische t-Test.
- Ausreißer über Boxplots prüfen. Bei ungleichen Gruppengrößen oder Varianzen ist der Welch-Test vorzuziehen (Delacre et al., 2017).

Tabelle 9

Gruppenstatistik

Variable	Gruppe	<i>n</i>	<i>M</i>	<i>SD</i>	<i>SE_M</i>
Classroom-Management	weiblich	34	3.95	0.40	0.07
	männlich	16	3.94	0.53	0.13

Anmerkung. *n* = Gruppengröße; *M* = Mittelwert; *SD* = Standardabweichung; *SE_M* = Standardfehler des Mittelwerts.

Tabelle 10

Ergebnis des t-Tests

Variable	<i>t</i>	<i>df</i>	<i>p</i>	<i>MD</i>	95 % CI [<i>LL</i> , <i>UL</i>]	<i>d</i>
Classroom-Management	0.13	48	.959	0.01	[-0.26, 0.28]	0.04

Anmerkung. *MD* = Mittelwertsdifferenz; *CI* = Konfidenzintervall; *d* = Cohens *d*.

Berichtsschema

$t(df) = X.XX$, $p = .XXX$, $d = X.XX$, 95 % CI [LL , UL].

Beispieltext (nicht signifikant)

Beispieltext

Zur Prüfung des Gruppenunterschieds im Classroom-Management wurde ein t-Test für unabhängige Stichproben gerechnet. Der Levene-Test war nicht signifikant ($p = .312$), sodass von Varianzhomogenität ausgegangen wurde. Zwischen weiblichen ($M = 3.95$, $SD = 0.40$) und männlichen Lehrkräften ($M = 3.94$, $SD = 0.53$) zeigte sich kein signifikanter Unterschied, $t(48) = 0.13$, $p = .959$, $d = 0.04$, 95 % CI $[-0.26, 0.28]$.

Kommentar zum Beispieltext

Reihenfolge: zuerst die Voraussetzungsprüfung (Levene), dann das Testergebnis.
Gruppenmittelwerte werden in Klammern mitgeführt, sodass der Befund auch ohne Tabelle nachvollziehbar ist.
Effektstärke (d) und Konfidenzintervall werden auch bei nicht signifikantem Ergebnis berichtet.

8.4 t-Test für verbundene Stichproben

Menüpfade

SPSS: Analysieren → Mittelwerte vergleichen → t-Test bei verbundenen Stichproben

R: `t.test(t1, t2, paired = TRUE); effectsize::cohens_d(paired = TRUE)`

Voraussetzungen

- Normalverteilung der Differenzwerte (zuvor Differenz bilden, dann Explore/Shapiro-Wilk).
- Ausreißer im Boxplot der Differenzwerte prüfen (Talikan et al., 2024).

Tabelle 11

Ergebnis des t-Tests für verbundene Stichproben

Vergleich	<i>MD</i>	<i>SD</i>	95 % CI [<i>LL</i> , <i>UL</i>]	<i>t</i>	<i>df</i>	<i>p</i>	<i>d_z</i>
Prä vs. Post	2.45	3.10	[1.57, 3.33]	5.57	49	< .001	0.79

Anmerkung. *MD* = mittlere Differenz zwischen den Messzeitpunkten; *d_z* = Effektgröße für verbundene Stichproben.

Berichtsschema

$t(df) = X.XX$, $p = .XXX$, $d_z = X.XX$, 95 % CI [LL , UL].

Beispieltext (signifikant)

Beispieltext

Die Differenzwerte waren annähernd normalverteilt (Shapiro-Wilk, $p = .214$). Vom Prä- zum Posttest stieg der Wert signifikant an, $t(49) = 5.57$, $p < .001$, $d_z = 0.79$, 95 % CI [1.57, 3.33]. Die mittlere Differenz betrug $MD = 2.45$ ($SD = 3.10$).

Kommentar zum Beispieltext

Da das Konfidenzintervall die Null nicht einschließt, ist der Unterschied auf dem 5%-Niveau signifikant – dies muss nicht zusätzlich behauptet werden.
Die Richtung (Anstieg) wird beschrieben, nicht jedoch erklärt.

8.5 Einfache lineare Regression

Menüpfade

SPSS: Analysieren → Regression → Linear

R: `lm(y ~ x); summary(); plot(model) für Diagnostik; lmtest::bptest()`

Voraussetzungen

- Linearität (Streudiagramm mit Regressionsgerade).
- Normalverteilung der Residuen (Histogramm, Normal-P-P-Plot); Modelle sind ab $N \approx 65$ robust (Lumley et al., 2002; Schmidt & Finan, 2018).
- Homoskedastizität (ZRESID gegen ZPRED; Breusch-Pagan- bzw. White-Test); bei Verletzung HC3-Schätzer für $N = 25-500$ (Long & Ervin, 2000).
- Unabhängigkeit der Residuen (Durbin-Watson, Richtwert 1.5–2.5).
- Ausreißer/Einfluss über standardisierte Residuen, Cooks Distanz und Hebelwerte (Wooldridge, 2013).

Tabelle 12

Regression von digitalem Erfolg auf Mitarbeiterunterstützung

Prädiktor	<i>b</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Konstante	0.88	0.25		3.54	.001
Mitarbeiterunterstützung	0.79	0.08	.67	9.48	< .001

Anmerkung. *b* = unstandardisierter Koeffizient; β = standardisierter Koeffizient. $R^2 = .45$, $F(1, 93) = 76.10$, $p < .001$.

Berichtsschema

Modellgüte: $R^2 = .XX$, $F(df_1, df_2) = X.XX$, $p = .XXX$. Koeffizient: $b = X.XX$, $SE = X.XX$, $\beta = .XX$, $t = X.XX$, $p = .XXX$.

Beispieltext (signifikant)

Beispieltext

Die Voraussetzungen der linearen Regression waren erfüllt (lineare Beziehung, homoskedastische und annähernd normalverteilte Residuen, Durbin-Watson = 1.94). Mitarbeiterunterstützung sagte den digitalen Erfolg signifikant vorher undklärte 45 % der Varianz auf, $R^2 = .45$, $F(1, 93) = 76.10$, $p < .001$. Mit jeder Einheit Mitarbeiterunterstützung stieg der digitale Erfolg um $b = 0.79$ ($SE = 0.08$), $\beta = .67$, $t(93) = 9.48$, $p < .001$.

Kommentar zum Beispieltext

Erst wird die Voraussetzungsprüfung knapp zusammengefasst, dann das Gesamtmodell, dann der einzelne Koeffizient.

Die Varianzaufklärung wird als Prozentwert benannt und durch R^2 belegt.

Die Aussage „mit jeder Einheit ... steigt“ beschreibt den Koeffizienten; eine ursächliche Deutung wäre erst in der Diskussion zulässig.

8.6 Multiple lineare Regression

Menüpfade

SPSS: Analysieren → Regression → Linear (mehrere Prädiktoren); in Statistiken VIF und Toleranz aktivieren

R: `lm(y ~ x1 + x2 + x3); car::vif(); summary()`

Voraussetzungen

- Wie einfache Regression, zusätzlich Prüfung auf Multikollinearität: $VIF < 10$, besser < 5 (Wooldridge, 2013).

Tabelle 13

Multiple Regression auf die Kaufintention

Prädiktor	<i>b</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	95 % CI [LL, UL]
Konstante	1.12	0.31		3.61	< .001	[0.51, 1.73]

Nachhaltigkeit	0.38	0.09	.36	4.22	< .001	[0.20, 0.56]
FOMO	0.11	0.08	.11	1.38	.171	[-0.05, 0.27]
Preisorientierung	-0.04	0.07	-.04	-0.57	.571	[-0.18, 0.10]

Anmerkung. $R^2 = .21$, korrigiertes $R^2 = .19$, $F(3, 91) = 8.06$, $p < .001$. Alle VIF < 1.5.

Beispieltext (gemischt)

Beispieltext

Das Gesamtmodell mit drei Prädiktoren war signifikant undklärte 21 % der Varianz der Kaufintention auf, $R^2 = .21$, $F(3, 91) = 8.06$, $p < .001$. Multikollinearität lag nicht vor (alle VIF < 1.5). Von den Prädiktoren leistete allein Nachhaltigkeit einen signifikanten Beitrag, $b = 0.38$, $\beta = .36$, $t(91) = 4.22$, $p < .001$, 95 % CI [0.20, 0.56]. FOMO ($p = .171$) und Preisorientierung ($p = .571$) trugen nicht bedeutsam bei.

Kommentar zum Beispieltext

Reihenfolge: Gesamtmodell, dann Multikollinearität, dann die einzelnen Prädiktoren.

Nicht signifikante Prädiktoren werden mit p -Wert genannt, aber nicht über ihre fehlende Wirkung spekuliert.

8.7 Mediationsanalyse

Menüpfade

SPSS: über Regression oder PROCESS-Makro (Analysieren → Regression → PROCESS); Bootstrapping mit Bias-corrected-CI und z. B. 5.000 Stichproben aktivieren

R: mediation::mediate() oder lavaan::sem() bzw. psych::mediate()

Voraussetzungen

- Regressionsvoraussetzungen wie bei multipler Regression.
- Mit Bootstrapping sind Verteilungsannahmen für den indirekten Effekt weniger kritisch.

Berichtsschema

Indirekter Effekt $ab = X.XX$, 95 % Bootstrap-CI [LL , UL] (signifikant, wenn das CI die Null nicht einschließt). Zusätzlich direkter Effekt c' und totaler Effekt c .

Beispieltext (signifikant)

Beispieltext

Geprüft wurde, ob der Zusammenhang zwischen Mitarbeiterunterstützung und digitalem Erfolg über die sequenzielle Implementierung vermittelt wird. Der indirekte Effekt war signifikant, $ab = 0.18$, 95 % Bootstrap-CI [0.07, 0.31] (5.000 Stichproben). Der direkte Effekt blieb bedeutsam, $c' = 0.42$, $p < .001$, was auf eine partielle Mediation hindeutet.

Kommentar zum Beispieltext

Beim Bootstrapping ist das Konfidenzintervall das entscheidende Kriterium; ein p -Wert für den indirekten Effekt ist nachrangig.

Die Einordnung als „partielle Mediation“ ist eine etablierte Beschreibung; die inhaltliche Deutung des Mechanismus gehört in die Diskussion.

8.8 ANOVA mit Messwiederholung (RM-ANOVA)

Menüpfade

SPSS: Analysieren → Allgemeines lineares Modell → Messwiederholte Messungen

R: afex::aov_ez() oder ez::ezANOVA()

Voraussetzungen

- Sphärizität (Mauchly-Test wird automatisch ausgegeben); bei Verletzung Korrektur nach Greenhouse-Geisser oder Huynh-Feldt.
- Normalverteilung je Bedingung (Residuenplots, Shapiro-Wilk); Ausreißer je Messzeitpunkt (Boxplots).

Tabelle 14*Innersubjekteffekt des Messzeitpunkts*

Quelle	SS	df	MS	F	p	η^2
Messzeitpunkt	18.42	1.83	10.07	12.64	< .001	.21
Fehler	69.31	87.0	0.80			

Anmerkung. Freiheitsgrade nach Greenhouse-Geisser korrigiert ($\epsilon = .91$). η^2 = partielles Eta-Quadrat.

Berichtsschema

$F(df_1, df_2) = X.XX$, $p = .XXX$, $\eta^2 = .XX$. Bei verletzter Sphärizität korrigierte Freiheitsgrade und ϵ angeben.

Beispieltext (signifikant)**Beispieltext**

Der Mauchly-Test deutete auf eine Verletzung der Sphärizität hin ($p = .028$), daher wurden die Freiheitsgrade nach Greenhouse-Geisser korrigiert ($\epsilon = .91$). Der Wert veränderte sich signifikant über die Messzeitpunkte, $F(1.83, 87) = 12.64$, $p < .001$, $\eta^2 = .21$, was einem großen Effekt entspricht.

Kommentar zum Beispieltext

Die Sphärizitätsprüfung steht vor dem Testergebnis; die Korrektur wird transparent gemacht (ϵ). Die Effektgröße wird gemäß Konvention eingeordnet, ohne den Befund inhaltlich zu deuten.

8.9 Gemischte ANOVA (Mixed ANOVA)**Menüpfade**

SPSS: Analysieren → Allgemeines lineares Modell → Messwiederholte Messungen (mit Zwischensubjektfaktor)

R: afex::aov_ez() mit between- und within-Faktor

Voraussetzungen

- Varianzhomogenität zwischen den Gruppen (Levene); Sphärizität für den Innersubjektfaktor (Mauchly).
- Residualdiagnostik wie bei der RM-ANOVA.

Berichtsschema

Für jeden Haupteffekt und die Interaktion: $F(df_1, df_2) = X.XX$, $p = .XXX$, $\eta^2 = .XX$.

Beispieltext (mit Interaktion)**Beispieltext**

Die Interaktion aus Messzeitpunkt und Gruppe war signifikant, $F(2, 176) = 6.41$, $p = .002$, $\eta^2 = .07$. Während die Interventionsgruppe vom Prä- zum Posttest deutlich zunahm, blieb die Kontrollgruppe weitgehend stabil. Der Haupteffekt der Gruppe war nicht signifikant, $F(1, 88) = 1.92$, $p = .169$, $\eta^2 = .02$.

Kommentar zum Beispieltext

Bei signifikanter Interaktion steht diese im Vordergrund; das Muster wird beschrieben (welche Gruppe verändert sich wie). Haupteffekte werden ergänzend berichtet, auch wenn sie nicht signifikant sind.

8.10 MANOVA

Menüpfade

SPSS: Analysieren → Allgemeines lineares Modell → Multivariat

R: `manova(cbind(y1, y2) ~ gruppe); summary(m, test = "Pillai")`

Voraussetzungen

- Multivariate Normalverteilung (näherungsweise über univariate Residuenplots beurteilen).
- Homogenität der Kovarianzmatrizen (Box-M-Test); multivariate Ausreißer über Mahalanobis-Distanzen; paarweise Linearität der AVs.

Berichtsschema

Multivariater Test: Pillai-Spur = .XX, $F(df_1, df_2) = X.XX$, $p = .XXX$. Bei Signifikanz folgen die univariaten ANOVAs.

Beispieltext (signifikant)

Beispieltext

Der multivariate Test fiel signifikant aus, Pillai-Spur = .14, $F(2, 87) = 7.02$, $p = .001$. Die nachgeordneten univariaten Analysen zeigten einen signifikanten Gruppenunterschied in der Kundenzufriedenheit, $F(1, 88) = 12.10$, $p < .001$, $\eta^2 = .12$, nicht jedoch in der Kostenminderung, $F(1, 88) = 1.44$, $p = .233$, $\eta^2 = .02$.

Kommentar zum Beispieltext

Zuerst der multivariate Gesamttest, dann die univariaten Folgetests – nur wenn der multivariate Test signifikant ist.

Pillais Spur ist die robusteste der gängigen Teststatistiken und wird daher bevorzugt berichtet.

8.11 Binäre logistische Regression

Menüpfade

SPSS: Analysieren → Regression → Binär logistisch

R: `glm(y ~ x1 + x2, family = binomial); performance::check_collinearity();`

`ResourceSelection::hoslem.test()`

Voraussetzungen

- Linearität im Logit für metrische Prädiktoren (Box-Tidwell); keine starke Multikollinearität (vorab VIF prüfen).
- Einflussreiche Fälle über Cooks Distanz und standardisierte Residuen; Modellanpassung über den Hosmer-Lemeshow-Test.

Tabelle 15

Binäre logistische Regression der Kaufentscheidung

Prädiktor	<i>b</i>	<i>SE</i>	Wald	<i>p</i>	<i>OR</i>	95 % CI [<i>LL</i> , <i>UL</i>]
Konstante	-1.84	0.52	12.52	< .001	0.16	—
Nachhaltigkeit	0.62	0.18	11.86	.001	1.86	[1.31, 2.64]
Preisorientierung	-0.21	0.16	1.72	.190	0.81	[0.59, 1.11]

Anmerkung. *OR* = Odds Ratio. Modell $\chi^2(2) = 18.94$, $p < .001$; Nagelkerkes $R^2 = .24$; Hosmer-Lemeshow $p = .61$.

Berichtsschema

Modell: $\chi^2(df) = X.XX$, $p = .XXX$. Prädiktor: *OR* = $X.XX$, 95 % CI [*LL*, *UL*], $p = .XXX$.

Beispieltext (signifikant)

Beispieltext

Das Gesamtmodell war signifikant, $\chi^2(2) = 18.94$, $p < .001$, und klärte nach Nagelkerke 24 % der Varianz auf. Der Hosmer-Lemeshow-Test deutete auf eine gute Modellanpassung hin ($p = .61$). Mit steigender Nachhaltigkeitsorientierung erhöhte sich die Chance einer nachhaltigen Kaufentscheidung, $OR = 1.86$, 95 % CI [1.31, 2.64], $p = .001$. Die Preisorientierung trug nicht bedeutsam bei ($p = .190$).

Kommentar zum Beispieltext

Odds Ratios werden über das Konfidenzintervall abgesichert; schließt es die 1 ein, ist der Prädiktor nicht signifikant.

Modellgüte (χ^2 , Pseudo- R^2 , Hosmer-Lemeshow) steht vor den einzelnen Prädiktoren.

8.12 Exploratorische Faktorenanalyse

Menüpfade

SPSS: Analysieren → Dimensionsreduktion → Faktorenanalyse

R: psych::fa(); Eignung über psych::KMO() und psych::cortest.bartlett()

Voraussetzungen / Richtwerte

- Stichprobengröße: mindestens vier Fälle pro Variable (Little, 2013); mindestens drei, besser drei bis vier Variablen pro Faktor (Velicer & Fava, 1998).
- Eignung der Daten: KMO $\geq .60$ und signifikanter Bartlett-Test (Bartlett, 1950; Kaiser, 1970).
- Faktorextraktion nach Eigenwert > 1 (Kaiser, 1960); Ausreißer ausschließen (Osborne, 2014). Bei nicht normalverteilten Daten liefert die Hauptfaktorenanalyse stabilere Ergebnisse als die Maximum-Likelihood-Methode (Fabrigar et al., 1999).

Tabelle 16

Faktorladungen nach Varimax-Rotation

Item	Faktor 1	Faktor 2
Item 1	.82	.11
Item 2	.78	.09
Item 3	.74	-.04
Item 4	.07	.81
Item 5	.12	.76

Anmerkung. KMO = .84; Bartlett $\chi^2(10) = 312.40$, $p < .001$. Ladungen $\geq .40$ sind hervorgehoben zu interpretieren.

Beispieltext (signifikant)**Beispieltext**

Die Daten waren für eine Faktorenanalyse gut geeignet (KMO = .84; Bartlett-Test $\chi^2(10) = 312.40$, $p < .001$). Nach dem Eigenwertkriterium (> 1) wurden zwei Faktoren extrahiert, die gemeinsam 64 % der Varianz aufklärten. Nach Varimax-Rotation luden die Items 1 bis 3 eindeutig auf den ersten, die Items 4 und 5 auf den zweiten Faktor (Ladungen $\geq .74$); Querladungen blieben unter .15.

Kommentar zum Beispieltext

Erst die Eignungsprüfung (KMO, Bartlett), dann die Faktorenlösung und die Varianzaufklärung, dann das Ladungsmuster.

Die inhaltliche Benennung der Faktoren erfolgt theoriegeleitet und kann je nach Vorgabe im Ergebnis- oder Diskussionsteil stehen.

9 Den Ergebnisteil schreiben

Im Ergebnisteil werden die Befunde der eigenen Untersuchung sachlich dargestellt und den Forschungsfragen bzw. Hypothesen zugeordnet. Die Ergebnisse werden ausschließlich beschrieben, nicht interpretiert; die Einordnung erfolgt in der Diskussion.

9.1 Aufbau und Reihenfolge

Bewährt hat sich folgende Gliederung:

1. Stichproben- und Datenbeschreibung: realisierter Stichprobenumfang, Umgang mit fehlenden Werten und Ausreißern.
2. Deskriptive Statistik der zentralen Variablen, häufig in einer Tabelle zusammengefasst.
3. Voraussetzungsprüfung für die geplanten Verfahren (sofern nicht im Methodenteil).
4. Hypothesenprüfung in der Reihenfolge der Hypothesen; pro Hypothese: Verfahren, Kennwerte, Effektstärke, Bezug zur Hypothese.

Die Ergebnisse werden nach Hypothesen oder nach Wichtigkeit geordnet. Jede Teilfrage und jede Hypothese erhält einen klar erkennbaren Bezug, sodass die Leserschaft auf einen Blick erkennt, welcher Befund zu welcher Hypothese gehört.

9.2 Zeitform

Der Ergebnisteil wird im Präteritum verfasst, da abgeschlossene Analysen beschrieben werden („der Unterschied war signifikant“). Das Präteritum wirkt wissenschaftlicher als das Perfekt. Auf Tabellen und Abbildungen wird im Präsens verwiesen („Tabelle 7 zeigt ...“).

9.3 Was hineingehört – und was nicht

- Alle relevanten Befunde werden berichtet – auch nicht signifikante und solche, die den Erwartungen widersprechen.
- Rohdaten gehören nicht in den Ergebnisteil, sondern in den Anhang.
- Keine Interpretation, keine Ursachenzuschreibung, kein Bezug auf fremde Studien. Formulierungen wie „daraus lässt sich schließen“ oder „das bedeutet, dass“ sind hier zu vermeiden.
- Tabellen und Text dürfen sich nicht doppeln: Im Text werden die zentralen Werte hervorgehoben oder zusammengefasst und auf die Tabelle verwiesen, nicht alle Zellen wiederholt.

9.4 Umgang mit nicht signifikanten Ergebnissen

Ein nicht signifikantes Ergebnis ist ein Ergebnis und wird vollständig berichtet (Teststatistik, p -Wert, Effektstärke, Konfidenzintervall). Ein nicht signifikanter Befund belegt nicht das Fehlen eines Effekts, sondern nur, dass kein hinreichender Beleg für einen Effekt gefunden wurde. Formulierungen wie „es zeigte sich kein signifikanter Unterschied“ sind angemessen, „es gibt keinen Unterschied“ hingegen nicht.

9.5 Abgrenzung von Ergebnis und Diskussion

Tabelle 17

Ergebnisteil im Vergleich zur Diskussion

Ergebnisteil	Diskussion
--------------	------------

beschreibt Befunde	interpretiert Befunde
ordnet Befunde den Hypothesen zu	wertet Hypothesen inhaltlich aus
nur Ergebnisse mit Bezug zur Fragestellung	Raum für weitere interessante Befunde
keine fremden Studien	Einordnung in den Forschungsstand
beantwortet die Forschungsfrage noch nicht	Fazit beantwortet die Forschungsfrage

Anmerkung. In qualitativen Arbeiten werden Ergebnis und Diskussion mitunter zusammengeführt; hier sind die Vorgaben der Hochschule maßgeblich.

9.6 Quantitative und qualitative Ergebnisse

In quantitativen Arbeiten werden je Hypothese das verwendete Verfahren, die relevanten Kennwerte und der Bezug zur Hypothese angegeben; komplexe Befunde werden über Tabellen oder Abbildungen veranschaulicht. In qualitativen Arbeiten werden die Ergebnisse nach Themen oder Relevanz geordnet, von allgemeinen zu detaillierten Befunden geführt, wiederkehrende Muster herausgearbeitet und durch prägnante, korrekt zitierte Aussagen der Teilnehmenden belegt; vollständige Transkripte gehören in den Anhang.

9.7 Kommentierter Musterabschnitt

Der folgende Abschnitt zeigt einen vollständigen, knappen Ergebnisabschnitt für eine Hypothese:

Beispieltext

Zur Prüfung von Hypothese 1, wonach sich nachhaltigkeitsorientierte und konventionell orientierte Konsumierende in der Kaufintention unterscheiden, wurde ein t-Test für unabhängige Stichproben durchgeführt. Die Voraussetzungen waren erfüllt: Die Differenzen waren annähernd normalverteilt (Shapiro-Wilk, $p = .18$) und die Varianzen homogen (Levene, $p = .44$). Nachhaltigkeitsorientierte Konsumierende ($M = 3.82$, $SD = 0.61$) wiesen eine signifikant höhere Kaufintention auf als konventionell orientierte ($M = 3.21$, $SD = 0.74$), $t(93) = 4.36$, $p < .001$, $d = 0.90$, 95 % CI [0.33, 0.89]. Hypothese 1 wurde damit gestützt.

Kommentar zum Musterabschnitt

Satz 1 nennt Hypothese und Verfahren; die Leserschaft weiß sofort, worum es geht.
 Satz 2 fasst die Voraussetzungsprüfung knapp zusammen, bevor das eigentliche Ergebnis folgt.
 Satz 3 berichtet Gruppenmittelwerte, Teststatistik, Signifikanz, Effektstärke und Konfidenzintervall in einem Zug.
 Der Schlusssatz ordnet den Befund der Hypothese zu („gestützt“), ohne ihn inhaltlich zu deuten – die Erklärung folgt erst in der Diskussion.
 Durchgehend Präteritum; auf eine Tabelle könnte mit „siehe Tabelle X“ verwiesen werden.

10 Hinweise zu weiteren Kapiteln

10.1 Methodenteil

Der Methodenteil beschreibt nachvollziehbar, wie die Forschungsfrage beantwortet wurde, sodass die Untersuchung reproduzierbar ist. Er folgt in der Gliederung auf Einleitung und theoretischen Rahmen und umfasst als Richtwert etwa 10 % des Textes. Hilfreich ist die Beantwortung von sechs Leitfragen:

5. Welche Art von Forschung (quantitativ, qualitativ, Mixed Methods; induktiv oder deduktiv)?
6. Wie wurden die Daten erhoben (Literaturrecherche, Feldstudie, Befragung, Experiment)?
7. Welche Eigenschaften haben die Daten, und welche Ein- und Ausschlusskriterien galten?
8. Wie ist die Erhebung praktisch abgelaufen (Zeitraum, Rahmenbedingungen)?
9. Wie wurden die Daten analysiert (Software wie SPSS oder R, verwendete Verfahren)?
10. Sind die Gütekriterien erfüllt (quantitativ: Objektivität, Reliabilität, Validität; qualitativ: Transparenz, Reichweite, Intersubjektivität)?

Der Methodenteil wird üblicherweise im Präteritum verfasst („die Zielgruppe war ...“), der Bezug auf laufende Analyseschritte im Präsens; Ausführungen zu Gütekriterien stehen im Präsens. Ein- und Ausschlusskriterien sind stets anzugeben.

Beispieltext (Methodenteil)

Zur Beantwortung der Forschungsfrage wurde eine quantitative Querschnittsbefragung durchgeführt. Die Datenerhebung erfolgte vom 1. bis 14. März 2025 über einen Online-Fragebogen. In die Auswertung einbezogen wurden ausschließlich vollständig ausgefüllte Fragebögen volljähriger Teilnehmender; unvollständige Datensätze wurden ausgeschlossen. Die Daten wurden mit SPSS (Version 29) ausgewertet; zur Hypothesenprüfung kamen t-Tests und multiple Regressionen zum Einsatz.

Kommentar

Art der Forschung, Erhebung, Ein-/Ausschlusskriterien, Software und Verfahren sind in wenigen Sätzen abgedeckt.

Wie ein Verfahren funktioniert, wird hier nicht erklärt – das gehört in den theoretischen Rahmen.

10.2 Diskussion und Fazit

In der Diskussion werden die Ergebnisse interpretiert, in den Forschungsstand eingeordnet und Limitationen reflektiert. Das Fazit beantwortet schließlich die Forschungsfrage. Beobachtungen, die über die Fragestellung hinausgehen, finden hier ihren Platz – nicht im Ergebnisteil.

11 Tabellen und Abbildungen in Word

11.1 Beschriftungen und automatische Verzeichnisse

Tabellen und Abbildungen sollten über die Word-Funktion Beschriftung nummeriert werden. Dadurch lässt sich ein automatisches, stets aktuelles Verzeichnis erzeugen.

- Beschriftung einfügen: Tabelle markieren → Register *Referenzen* (bzw. *Verweise*) → *Beschriftung einfügen* → Bezeichnung „Tabelle“ wählen. Word vergibt und aktualisiert die Nummerierung automatisch.
- Verzeichnis erstellen: an der gewünschten Stelle → *Referenzen* → *Abbildungsverzeichnis einfügen* → unter „Beschriftung“ „Tabelle“ wählen. Analog für Abbildungen.
- Querverweise: im Text mit *Referenzen* → *Querverweis* auf die Beschriftung verweisen. Werden Tabellen verschoben oder ergänzt, bleibt die Nummerierung über Felder aktualisieren (Strg + A, dann F9) korrekt.

Wichtiger APA-Stolperstein

Word setzt standardmäßig „Tabelle 1: Titel“ in eine Zeile. APA verlangt jedoch die Tabellenummer (fett) und darunter den Titel (kursiv) in getrennten Zeilen über der Tabelle.

Lösung: Im Dialog *Beschriftung einfügen* das Trennzeichen anpassen bzw. die Beschriftungsformatvorlage so ändern, dass Nummer fett und Titel kursiv in zwei Zeilen erscheinen. Das Verzeichnis greift weiterhin auf die Beschriftung zu.

11.2 APA-konforme Tabellengestaltung

- Nur horizontale Linien verwenden (über und unter dem Tabellenkopf sowie unter der letzten Zeile); keine vertikalen Linien und keine inneren Gitterlinien.
- Über der Tabelle: Nummer (fett), darunter Titel (kursiv). Unter der Tabelle die Anmerkung, eingeleitet mit kursivem „Anmerkung.“
- Anmerkungen in der Reihenfolge: allgemeine Anmerkung (z. B. Erläuterung der Abkürzungen) → spezifische Anmerkungen → Signifikanzanmerkungen ($p < .05$ usw.).
- Tabellen als echte Word-Tabellen einfügen, nie als Bild (Screenshots sind nicht durchsuchbar und schlecht lesbar).
- Bei seitenübergreifenden Tabellen: Kopfzeile wiederholen (Tabelleneigenschaften → Zeile → „Gleiche Kopfzeile auf jeder Seite wiederholen“) und „Zeilenwechsel auf Seiten zulassen“ deaktivieren.

11.3 Abbildungen

- Abbildungen erhalten oberhalb die Nummer (fett) und den kursiven Titel; eine Anmerkung steht darunter.
- Achsen sind vollständig zu beschriften, Skalen eindeutig anzugeben. Auf eine klare, nicht überladene Darstellung achten.
- Im Text wird auf jede Abbildung und Tabelle verwiesen, ohne deren Inhalt vollständig zu wiederholen.

12 Einsatz von KI-Werkzeugen

Der Einsatz generativer KI (z. B. Sprachmodelle) ist in Abschlussarbeiten je nach Hochschule unterschiedlich geregelt. Maßgeblich sind stets die Vorgaben der eigenen Institution; die folgenden Punkte gelten als gute Praxis.

12.1 Transparenz und Kennzeichnung

- Viele Hochschulen verlangen zusätzlich zur Eigenständigkeitserklärung eine Erklärung über den Einsatz von KI, in der Werkzeug, Zweck und Umfang offengelegt werden.

- Zulässig ist meist Unterstützung bei Recherche, Strukturierung, Sprache und Formulierung. Die inhaltliche und analytische Eigenleistung muss von der verfassenden Person stammen; die Verantwortung für alle Inhalte verbleibt bei ihr.
- KI kann fehlerhafte Inhalte und nicht existierende Quellen erzeugen. Fakten und Zitationen sind grundsätzlich an den Originalquellen zu prüfen.

12.2 Zitation von KI nach APA

Wird der Output eines KI-Werkzeugs wiedergegeben, behandelt APA dieses als Software. Im Text wird mit Hersteller und Jahr zitiert; im Literaturverzeichnis wird das Werkzeug mit Version aufgeführt, z. B.:

Beispiel

Im Text: (OpenAI, 2025)

Im Verzeichnis: OpenAI. (2025). ChatGPT (Version vom 1. Juni) [Großes Sprachmodell].
<https://chat.openai.com>

12.3 Datenschutz

- In KI-Werkzeuge dürfen keine personenbezogenen Rohdaten, vertraulichen oder urheberrechtlich geschützten Inhalte eingegeben werden, sofern keine entsprechende Rechtsgrundlage und sichere Verarbeitung vorliegt.
- Bei der Verarbeitung sensibler Daten ist auf datenschutzkonforme Werkzeuge und auf Anonymisierung zu achten.

13 Häufige Fehler – Checkliste

- Führende Null bei p , r , β gesetzt (sollte fehlen) oder bei M , t , F weggelassen (sollte stehen).
- $p = .000$ statt korrekt $p < .001$ berichtet.
- Dezimaltrennzeichen uneinheitlich (Punkt und Komma gemischt).
- Effektstärke und/oder Konfidenzintervall fehlen.
- Normalverteilung der Rohdaten statt der Residuen bzw. Differenzwerte geprüft.
- Levene-Ergebnis ignoriert und falsche Zeile (Student statt Welch) berichtet.
- Nicht signifikanter Befund als „kein Effekt vorhanden“ fehlinterpretiert.
- Interpretation im Ergebnisteil statt in der Diskussion.
- Tabelleninhalte im Text vollständig wiederholt.
- Tabellen als Bild eingefügt; vertikale Linien verwendet; Nummer und Titel nicht APA-konform.
- Verfahren bei verletzten Voraussetzungen ohne robuste oder verteilungsfreie Alternative berichtet.
- KI-Nutzung nicht gekennzeichnet; von KI erfundene Quellen ungeprüft übernommen.

14 Literaturverzeichnis

- Allevato, E. (2019). A project-based approach to examine university teachers attitudes towards visually impaired students. *Scholarly Journal of Psychology and Behavioral Sciences*, 2(3).
<https://doi.org/10.32474/sjpbs.2019.02.000137>
- American Psychological Association (Hrsg.). (2020). *Publication manual of the American Psychological Association* (7. Aufl.). American Psychological Association. <https://doi.org/10.1037/0000165-000>
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Statistical Psychology*, 3(2), 77–85. <https://doi.org/10.1111/j.2044-8317.1950.tb00285.x>
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5), 1287–1294. <https://doi.org/10.2307/1911963>
- Cohen, J. (2009). *Statistical power analysis for the behavioral sciences* (2. Aufl., Nachdruck). Psychology Press.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/bf02310555>
- Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of Student's t-test. *International Review of Social Psychology*, 30(1), 92–101. <https://doi.org/10.5334/irsp.82>
- Durbin, J., & Watson, G. S. (1950). Testing for serial correlation in least squares regression: I. *Biometrika*, 37(3/4), 409–428. <https://doi.org/10.2307/2332391>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Gauss, C. F. (1823). *Theoria combinationis observationum erroribus minimis obnoxiae*. H. Dieterich.
- Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, 19(2), 149–161. <https://doi.org/10.1007/bf02289162>
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141–151. <https://doi.org/10.1177/001316446002000116>
- Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, 35(4), 401–415. <https://doi.org/10.1007/bf02291817>
- Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement*, 34(1), 111–117. <https://doi.org/10.1177/001316447403400115>
- Larose, D. T. (2005). *Data mining methods and models* (1. Aufl.). Wiley. <https://doi.org/10.1002/0471756482>
- Levene, H. (1960). Robust tests for equality of variances. In *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (S. 278–292). Stanford University Press.
- Little, T. D. (Hrsg.). (2013). *The Oxford handbook of quantitative methods in psychology: Vol. 2* (1. Aufl.). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199934898.001.0001>
- Long, J. S., & Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3), 217–224. <https://doi.org/10.2307/2685594>
- Lumley, T., Diehr, P., Emerson, S., & Chen, L. (2002). The importance of the normality assumption in large public health data sets. *Annual Review of Public Health*, 23, 151–169. <https://doi.org/10.1146/annurev.publhealth.23.100901.140546>
- OpenAI. (2025). *ChatGPT* (Version vom 1. Juni) [Großes Sprachmodell]. <https://chat.openai.com>
- Osborne, J. W. (2014). *Best practices in exploratory factor analysis*. CreateSpace Independent Publishing Platform.

- Salkind, N. J. (Hrsg.). (2010). *Encyclopedia of research design*. SAGE.
<https://doi.org/10.4135/9781412961288>
- Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of Clinical Epidemiology*, 98, 146–151. <https://doi.org/10.1016/j.jclinepi.2017.12.006>
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4), 591–611. <https://doi.org/10.2307/2333709>
- Stephens, M. A. (1974). EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347), 730–737. <https://doi.org/10.2307/2286009>
- Student. (1908). The probable error of a mean. *Biometrika*, 6(1), 1–25. <https://doi.org/10.2307/2331554>
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296.
<https://doi.org/10.1007/s11165-016-9602-2>
- Talikan, A. I., Salapuddin, R., Aksan, J. A., Rahimulla, R. J., Ismael, A., Jimlah, R., Idris, N., Dammang, R. B., Jamar, D. A., Sarahadil, E., & Ajan, R. A. (2024). *On paired samples t-test: Applications, examples and limitations*. Zenodo. <https://doi.org/10.5281/ZENODO.10987546>
- Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, 3(2), 231–251. <https://doi.org/10.1037/1082-989x.3.2.231>
- Velleman, P. F., & Welsch, R. E. (1981). Efficient computing of regression diagnostics. *The American Statistician*, 35(4), 234–242. <https://doi.org/10.2307/2683296>
- Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, 34(1/2), 28–35. <https://doi.org/10.2307/2332510>
- Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5. Aufl.). South-Western Cengage Learning.