

# Der Einsatz generativer KI in der statistischen Datenanalyse und ihre Auswirkungen auf die Lehre der Statistik an Fachhochschulen<sup>1</sup>

JOACHIM SCHWARZ, EMDEN

<sup>1</sup> Übersetzung von Manfred Borovcnik basierend auf: Schwarz, J. (2025). The use of generative AI in statistical data analysis and its impact on teaching statistics. *Teaching Statistics*, 47(2), 118–128. Original Open Access. Original und Übersetzung CC BY 4.0: <https://doi.org/10.1111/test.12398>

**Zusammenfassung:** Diese Studie untersucht den Einsatz generativer KI, insbesondere ChatGPT, in der statistischen Datenanalyse und deren Auswirkungen auf die Statistikausbildung an Fachhochschulen. Der Beitrag beginnt mit einer Diskussion über die zukünftige Arbeitsteilung zwischen Mensch und Maschine im Kontext der statistischen Datenanalyse nach der Einführung von ChatGPT. Durch die Analyse künstlich generierter Datensätze werden die Fähigkeiten und Grenzen von ChatGPT bei der Durchführung statistischer Analysen aufgezeigt. Die Ergebnisse zeigen, dass generative KI die Datenanalyse für Personen mit minimalen Statistikkenntnissen erleichtern kann, hauptsächlich durch die Generierung geeigneter Codes, aber nur teilweise durch die Befolgung von Standardverfahren. Daher bleiben menschliche Aufsicht und gute Statistikkenntnisse weiterhin von entscheidender Bedeutung. Die Studie legt nahe, dass sich der Schwerpunkt der Statistikausbildung von der Vermittlung software-spezifischer Fähigkeiten hin zur Förderung eines tieferen Verständnisses statistischer Konzepte verlagern sollte. Diese Verlagerung wird die Studierenden darauf vorbereiten, KI-Tools effektiv zu nutzen und gleichzeitig strenge Standards für die Datenanalyse einzuhalten.

## 1 Einleitung

Seit OpenAI im November 2022 sein *Large Language Model* (LLM) ChatGPT 3.5 der Öffentlichkeit zugänglich gemacht hat, haben Anwendungen generativer KI an Hochschulen erheblich an Bedeutung gewonnen. Eine Studie der University of Baltimore ergab beispielsweise, dass 67 % der Lehrkräfte und 53 % der Studierenden ChatGPT bereits genutzt hatten (Stansbury et al., 2023). Eine ähnliche Studie der University of Michigan ergab, dass 58 % der Lehrkräfte und bis zu 66 % der Studierenden mindestens einmal ein generatives KI-Tool genutzt hatten (Duraisamy et al., 2023).

Generative KI bietet sowohl Chancen als auch Herausforderungen für die Hochschulbildung. Zu den Chancen zählen die Verbesserung der Lernerfahrungen der Studierenden, die Rationalisierung von Verwaltungsaufga-

ben und die Unterstützung von Forschungsaktivitäten (siehe z. B. die narrative Literaturübersicht von Chukwuere, 2024). Zu den Herausforderungen zählen u. a. Plagiate, akademische Integrität, Fehlinformationen sowie Datenschutz- und Urheberrechtsbedenken (Bobula, 2024; Neumann et al., 2023).

Die Auswirkungen oder Folgen des Einsatzes generativer KI im Bereich der Statistikausbildung und Datenanalyse wurden in der bisherigen Diskussion jedoch nur wenig beachtet. Eine Literaturrecherche in den Fachzeitschriften *American Statistician*, *Journal of Applied Learning and Teaching*, *Journal of Educational and Behavioral Statistics*, *Journal of Statistics and Data Science Education*, *Statistics Education Research Journal*, *Teaching Statistics* und *Technology Innovations in Statistics Education* ergab nur wenige Quellen zu diesem Thema. Die Suche umfasste alle Beiträge, die im Zeitraum von Januar 2023 bis August 2024 veröffentlicht wurden. Die Suchbegriffe für die Recherche lauteten „generative KI“, „Large Language Models (LLM)“ und „ChatGPT“.

Die ersten, die die Datenanalysefähigkeiten von ChatGPT erwähnten, waren Rasul et al. (2023), obwohl ihre Forschung keine Tests dieser Fähigkeiten umfasste. Dennoch schlugen die Autoren vor, dass „Hochschuleinrichtungen Studierende und Wissenschaftler in Umgang und Missbrauch von ChatGPT für Forschung und Datenanalyse schulen müssen“ (S. 46). Ein weiterer früher Beitrag zu dieser Debatte stammt von Ellis und Slade (2023), die die Verwendung von ChatGPT zur Erstellung von Kursinhalten, Prüfungsfragen und Code-Generierung in der Statistik diskutieren.

In einer weiteren Forschungsarbeit diskutiert Xing (2024) die Verwendung von ChatGPT zur Beantwortung konzeptioneller Fragen in der Statistik und kommt zum Schluss, dass ChatGPT einige Fragen beantworten und diese Antworten mit entsprechenden Erklärungen untermauern kann. Xing beschreibt auch, wie ChatGPT Datenanalysen mit Richtlinien, Modellauswahl, Programmcode unter Verwendung von R und der Berichterstattung über Ergebnisse unterstützt. Da Xing jedoch ChatGPT 3.5 verwendet, konnte er dies nicht für die Analyse von Datensätzen testen.

Dies sind bis zum oben genannten Zeitpunkt die einzigen Beiträge, die sich mit der Datenanalysefähigkeit von ChatGPT im Kontext von Unterricht befassen.

Über den Fokus auf den Unterricht hinaus gibt es jedoch noch weitere Arbeiten. So untersuchten Irvine et al. (2023) die Fähigkeiten von ChatGPT für die Analyse hydrologischer Datensätze. Die Autoren gehen über die reine Codegenerierung durch ChatGPT hinaus und untersuchen auch die Fähigkeiten zur Darstellung der Daten. Die Autoren weisen darauf hin, dass ChatGPT es Personen, die keine Kenntnisse in R, Python oder einer anderen Programmiersprache haben, ermöglichen kann, Datenanalysen durchzuführen (Irvine et al., S. 2).

Cribben und Zeinali (2023) berichten über einen ersten Versuch, ChatGPT zur Erstellung eines Entscheidungsbaums für einen bekannten Datensatz, den Boston House Price Dataset (S. 23), zu verwenden. Die Autoren berichten, dass die von ChatGPT präsentierten Ergebnisse von den Ergebnissen ihrer eigenen Analyse abweichen. Diese Abweichung in den Ergebnissen veranlasste die Autoren, zur Vorsicht bei der Verwendung von ChatGPT aufzurufen. (S. 24)

Schließlich testeten Huang et al. (2024) explizit die analytischen Fähigkeiten von ChatGPT anhand eines medizinischen Datensatzes und verglichen die Ergebnisse mit klassischen statistischen Softwaretools, die in der Medizin verwendet werden (SAS, Python und R). Sie erstellten sowohl deskriptive Statistiken als auch Vergleiche von Gruppen unter Verwendung von Hypothesentests. Während die deskriptiven Statistiken von ChatGPT konsistent waren, stellten die Autoren geringfügige Abweichungen bei den Hypothesentests von ChatGPT fest. In ihrer Diskussion weisen die Autoren darauf hin, dass „ChatGPT-4 bei der Durchführung gängiger Datenanalyseaufgaben mit Einfachheit und Effizienz mit traditioneller Statistiksoftware mithalten kann, [aber] traditionelle biostatistische Software für komplexe und fortgeschrittene Aufgaben nach wie vor besser geeignet ist“ (Huang et al., 2024, S. 6).

Diese ersten Studien zeigen, dass generative KI für die Datenanalyse eingesetzt werden kann. Dies ist besonders relevant für die Statistikausbildung an Fachhochschulen, deren Schwerpunkt eher auf der Vorbereitung ihrer Studierenden auf eine Karriere in der Privatwirtschaft als auf einer akademischen Laufbahn liegt. Die aktuellen Entwicklungen in der Wirtschaft lassen die Annahme, dass Unternehmen in Zukunft LLMs für ihre tägliche Arbeit einsetzen werden, sehr plausibel erscheinen. Eine Umfrage unter 786 Beschäftigten in Deutschland, die von Statista Research Department (2024) durchgeführt wurde, zeigt, dass dies bereits heute der Fall ist. In dieser Umfrage gaben 21,5 % der Befragten an, dass sie ChatGPT für die Datenanalyse nutzen. Auch wenn in dieser Umfrage nicht näher spezifiziert wurde, was genau unter Datenanalyse mit ChatGPT zu verste-

hen ist und ob es sich überhaupt um eine repräsentative Stichprobe handelt, unterstreicht dies die Relevanz des Themas für die Ausbildung an Fachhochschulen.

Studierende lernen während ihres Studiums je nach gewähltem Studiengang in unterschiedlichem Umfang Statistik. Dieser Beitrag konzentriert sich auf Studierende, die keine Karriere als Datenanalyst oder Datenwissenschaftler anstreben, beispielsweise Studierende der Sozialwissenschaften oder der Betriebswirtschaftslehre. Zusätzlich zu den grundlegenden theoretischen Konzepten lernen diese Studierenden in der Regel auch, wie sie diese Konzepte auf reale Datensätze anwenden können, wofür sie Analysesoftware wie Excel oder R verwenden. Dies steht in völligem Einklang mit den „Guidelines for Assessment and Instruction in Statistics Education“ (GAISE), in denen darauf hingewiesen wird, dass die Verwendung realer Daten in der Statistikausbildung für die Entwicklung eines Verständnisses von Daten und den relevanten statistischen Konzepten unerlässlich ist (GAISE, 2016, S. 17). Wenn jedoch große Teile, wenn nicht sogar die gesamte Datenanalyse nun von einer generativen KI durchgeführt werden, stellt sich die Frage, wie der Statistikerunterricht aussehen kann und welche Schwerpunkte gesetzt werden sollen. Kann man so weit gehen zu sagen, dass „generative KI das Ende des Lehrerberufs bedeuten wird?“ (Rudolph et al., 2023, S. 8).

Die hier diskutierte Frage lautet daher, worauf sich der Statistikerunterricht in Zukunft konzentrieren kann. Bei deren Beantwortung stütze ich mich zunächst auf die Arbeit von Frey und Osborne (2023), welche die Auswirkungen von LLMs auf die zukünftige Arbeitsteilung zwischen Mensch und Maschine untersucht hat. Auf der Grundlage dieser Überlegungen wird in diesem Beitrag die zukünftige Arbeitsteilung zwischen Mensch und Maschine im Bereich der Datenanalyse diskutiert. Anhand einfacher Datensätze werden dann die Möglichkeiten und Grenzen von ChatGPT in der Datenanalyse aufgezeigt und Empfehlungen für Lehrkräfte abgeleitet.

## 2 Die zukünftige Arbeitsteilung zwischen Mensch und Maschine

Ein guter Ausgangspunkt für die Erstellung von Lehrplänen für berufsorientierte Bildungsgänge ist es, die Anforderungen der Unternehmen an zukünftige Absolventen zu betrachten. Dies gilt umso mehr für Fachhochschulen. Die Herausforderung besteht darin, heute für den Arbeitsmarkt von morgen auszubilden, der sich aufgrund der disruptiven Natur der generativen KI massiv verändern wird. Es liegt auf der Hand, dass ein LLM, das Bilder generieren und Texte schreiben kann, eine akute Bedrohung für die Arbeitsplätze von Designern und Schriftstellern darstellt. Eine Analyse

von Stellenanzeigen auf Freelancer-Plattformen ergab beispielsweise einen Rückgang von 21 % bei der Anzahl der Stellenanzeigen für automatisierungsanfällige Tätigkeiten im Bereich Schreiben und Programmieren und einen Rückgang von 17 % bei der Anzahl der Stellenanzeigen im Bereich Grafikdesign und 3D-Modellierung nach der Einführung von ChatGPT (Demirci et al., 2023, S. 2).

Angesichts dieser Zahlen stellt sich die Frage, ob sich nur die Erstellung von Inhalten im weitesten Sinne ändern wird oder ob auch andere Berufe betroffen sein werden. Frey und Osborne (2023) nähern sich dieser Frage, indem sie die zukünftige Arbeitsteilung zwischen Mensch und Maschine diskutieren. Als ersten Punkt nennen sie den Faktor Mensch, beispielsweise in Bezug auf Motivation oder Vertrauen in Geschäfts- und Kundenbeziehungen. Hier gehen die Autoren davon aus, dass dieser in Zukunft nicht durch generative KI ersetzt werden kann und immer einen Menschen erfordern wird (S. 3).

Der zweite Punkt, den die Autoren ansprechen, ist Kreativität, also die Fähigkeit, etwas wirklich Neues zu schaffen. Generative KI ist sehr gut darin, vorhandenes Wissen neu zu kombinieren und auf andere Weise zusammenzusetzen, aber das führt nicht zu etwas wirklich Neuem. Außerdem hat eine generative KI kein konzeptionelles Verständnis dafür, was sie tut oder worüber sie spricht (Frey & Osborne, 2023, S. 5).

In ihrem dritten Punkt beziehen sich die Autoren auf die Halluzinationen, die generative KI hervorrufen kann, und verweisen dabei auf die Erzeugung falscher oder irreführender Ergebnisse durch KI. Die Autoren sprechen hier zwei Aspekte an. Ein Aspekt ist, in welchen Situationen Fehler überhaupt akzeptabel sind: „KI – in ihrer derzeitigen Form – wird eher in weniger risikoreichen Bereichen wie Kundenservice oder Lagerhaltung eingesetzt als in risikoreicheren Bereichen wie dem Autofahren.“ (Frey & Osborne, 2024, S. 7). Der zweite Aspekt bezieht sich auf die Struktur des zu lösenden Problems: Je einfacher eine Aufgabe oder ein Problem strukturiert ist, desto leichter kann generative KI es lösen.

Frey & Osborne (2023, S. 9) kommen daher zu folgendem Schluss:

„Die unmittelbare Auswirkung der jüngsten Welle generativer KI wird jedoch weder Automatisierung noch neue Industrien sein, sondern die Transformation bestehender Jobs im Bereich der Inhaltserstellung, wodurch diese leichter auszuführen sein werden. [...] Zum einen müssen Menschen eine Eingabe vornehmen und dann die gewünschte Ausgabe auswählen (und meist auch bearbeiten), und genau in dieser Eingabe und Auswahl liegt ein Großteil der tatsächlichen Kreativität.“

Die Autoren weisen auch darauf hin, dass generative KI tendenziell eine durchschnittliche menschliche Leistung erbringt, da die LLMs auf einer breiten Masse von Informationen und nicht auf den Informationen ausgewählter Experten trainiert werden. Eine Folge davon ist auch, dass insbesondere die Leistung von Menschen mit unterdurchschnittlicher Ausbildung gesteigert werden kann, während gut ausgebildete Experten in geringerem Maße davon profitieren (Frey & Osborne, 2023, S. 9).

### **3 Zukünftige Arbeitsteilung Mensch – Maschine in der Datenanalyse**

Die Anwendung der Überlegungen von Frey und Osborne (2023) auf die statistische Datenanalyse impliziert unmittelbar, dass der Einsatz generativer KI die Durchführung solcher Analysen erleichtern wird. Darüber hinaus wird dies auch für Personen möglich, die nur über geringe oder gar keine Statistikkennntnisse verfügen, und zwar aus zwei Hauptgründen:

Wie oben erläutert, entwickelt generative KI nichts wirklich Neues, sondern kombiniert lediglich vorhandenes Wissen neu. Wendet man dies auf die Analyse eines Datensatzes an, wird schnell klar, dass hier, wenn überhaupt, nur in den seltensten Fällen wirklich etwas Neues entsteht. Jedes Verfahren zur Analyse eines Datensatzes folgt weitgehend denselben Schritten. Sobald eine generative KI weiß, welches Skalenniveau für eine bestimmte Variable gilt, weiß sie auch, welche statistischen Methoden sie für jedes einzelne Datenfeld anwenden kann und welche nicht.

Das Gleiche gilt für die Datenaufbereitung und deskriptive Datenanalyse. Hier gibt es Standardschritte, die immer durchgeführt werden müssen, wie die Berechnung deskriptiver Kennzahlen, die Datenvisualisierung, der Umgang mit fehlenden Werten und so weiter. Die Auswahl eines geeigneten statistischen Testverfahrens basiert üblicherweise auf Regeln und der Überprüfung von Annahmen und kann daher vollständig von einer generativen KI übernommen werden. Je einfacher die Struktur des zu analysierenden Datensatzes ist, desto mehr trifft dies zu.

Der zweite Grund ist, dass es nicht mehr notwendig ist, Kenntnisse im Umgang mit Auswertungssoftware wie R oder Python zu erwerben. Aufgrund der Fähigkeit der generativen KI, Programmcode zu generieren, kann sie eine Analyse auch selbstständig durchführen. Es müssen lediglich noch die richtigen Anweisungen („Prompts“) formuliert werden. Wenn statistische Analysen jedoch mehr oder weniger aus immer gleichen Schritten bestehen, ist es plausibel anzunehmen, dass auch hier zumindest für Probleme mit einer einfacheren Struktur keine oder nur geringe statistische Kenntnisse erforderlich sind.

Es bleibt zu überlegen, was dies für die Statistikausbildung an Universitäten bedeutet. An Fachhochschulen und insbesondere für Studierende der Betriebswirtschaftslehre oder Sozialwissenschaften besteht der Statistikkunterricht aus einer Mischung aus der Vermittlung theoretischer Grundlagen und Anwendungsübungen mit der Interpretation von Ergebnissen. Wenn, wie oben erläutert, die Durchführung der Datenanalyse vollständig von einer generativen KI übernommen werden kann, reicht es vielleicht aus, den Unterricht allein auf die theoretischen Grundlagen zu konzentrieren. Oder vielleicht kann sogar ganz darauf verzichtet werden, da auch die Erläuterung der einzelnen Begriffe und Konzepte von einer generativen KI übernommen werden kann.

## 4 Anwendung von ChatGPT auf Datensätze

Wie oben erläutert, kann davon ausgegangen werden, dass generative KI auch Personen mit geringen oder gar keinen Statistikkennnissen die statistische Datenanalyse ermöglicht, zumindest bei Problemen mit einer einfacheren Struktur. Um dies zu überprüfen, werde ich mich im Folgenden auf Analysen konzentrieren, die mit den grundlegenden Methoden des  $t$ -Tests, der linearen Regression und der ANOVA durchgeführt werden können. Diese Methoden sind Teil des Standardlehrplans eines zweisemestrigen Statistikkurses in Betriebswirtschaftslehre und Sozialwissenschaften. Darüber hinaus werde ich kurz untersuchen, ob und inwieweit generative KI eine Orientierungshilfe für die Datenanalyse bieten kann.

Die statistisch zu untersuchenden Fragestellungen wurden mit künstlich generierten Datensätzen getestet. Insgesamt wurden vier Datensätze generiert, die alle recht einfach gehalten waren. Sie enthielten nur die Variablen, die für die jeweilige Analysefragestellung relevant waren. Die Datensätze wurden so generiert, dass zumindest eine geringfügige Verletzung der Anwendungsvoraussetzungen für statistische Testverfahren vorlag.

Für die für eine ANOVA geeignete Fragestellung wurde zusätzlich ein zweiter Datensatz mit schwerwiegenden Verletzungen der Voraussetzungen generiert. Gemäß der obigen Erläuterung ist die Überprüfung der Voraussetzungen einer der Standardschritte einer statistischen Datenanalyse, zumindest sollten die schwerwiegenden Verletzungen erkannt werden. Im Folgenden werden die Ergebnisse für die beiden Datensätze mit für die ANOVA geeigneten Fragestellungen detailliert dargestellt. Die Ergebnisse für die übrigen Datensätze sind im Abschnitt „Diskussion“ zusammengefasst. Alle vier Datensätze und der generierende  $R$ -Code sind in einem Repository gespeichert.

Als generative KI wurde der GPT *Data Analyst* von ChatGPT verwendet. OpenAI bezeichnet GPTs, die auf bestimmte Anforderungen zugeschnitten sind, als Versionen von ChatGPT (OpenAI, 2023b). Der *Data Analyst* ist für Hochladen und Auswerten von Datensätzen konzipiert und erfordert eine kostenpflichtige ChatGPT Plus-Lizenz (OpenAI, 2023a).

Bei der Formulierung der Eingabeaufforderungen für *Data Analyst* wird von einem Benutzer ausgegangen, der noch nie einen Statistikkurs besucht hat. Ein solcher theoretischer Benutzer ist mit grundlegenden grafischen Darstellungen wie Balkendiagrammen oder Kreisdiagrammen aus anderen Quellen vertraut und weiß auch, was ein arithmetischer Mittelwert ist. Er oder sie verfügt noch über Grundkenntnisse der Wahrscheinlichkeitstheorie, wie z. B. Urnenmodelle, aus der Schulzeit. Ein solcher Benutzer wäre vor der Einführung von LLMs nicht in der Lage gewesen, inferenzstatistische Datenanalysen durchzuführen. Darüber hinaus kennt er keine fachspezifischen Fachbegriffe oder Ausdrücke, die er zur Formulierung der Eingabeaufforderungen verwenden könnte.

### 4.1 Orientierungshilfe zur Vorgehensweise bei der Datenanalyse

Auch wenn dies nicht im Mittelpunkt dieses Artikels steht, werde ich zunächst kurz zeigen, ob und inwieweit *Data Analyst* in der Lage ist, Anleitungen für die Herangehensweise an die Datenanalyse zu geben. Dazu geht es zunächst um allgemeine beschreibende Kennzahlen und Diagrammtypen zur Zusammenfassung und Visualisierung von Daten. Anschließend folgen spezifischere Fragen zu bestimmten Datensätzen. Die Eingabeaufforderungen und die Dokumentation der Antworten von ChatGPT finden Sie im selben Repository wie die Datensätze; hier wird nur eine kurze Zusammenfassung gegeben.

Die Frage an *Data Analyst* „Welche Kennzahlen und Grafiken sind für eine deskriptive Datenanalyse erforderlich?“ führte zu einer Zusammenstellung der wichtigsten davon für die univariate sowie die bivariate und multivariate deskriptive Datenanalyse. *Data Analyst* lieferte auch Informationen zu ihrer Eignung für metrische und nicht-metrische Variablen. Für einen bestimmten Datensatz werden die Empfehlungen von *Data Analyst* spezifischer. Darüber hinaus werden Informationen dazu bereitgestellt, was bei der Beurteilung der jeweiligen Diagramme und Kennzahlen zu beachten ist.

Auf die Frage, wie man bei der Analyse einer bestimmten Forschungsfrage vorgehen sollte (z. B. „Was muss ich mit diesem Datensatz tun, wenn ich untersuchen möchte, ob sich Variable  $A$  für verschiedene Stufen von

Variable  $B$  unterscheidet?“), antwortete ChatGPT *Data Analyst* mit mehreren Schritten, die als Leitfaden zur Beantwortung der gestellten Forschungsfrage dienen können.

Der erste Schritt bestand in der Regel aus der Formulierung der Nullhypothese, gefolgt von deskriptiven Statistiken und Visualisierungen, und endete mit einem Hypothesentest und einem Verweis auf die erforderlichen Anwendungsannahmen und alternativen Testverfahren. Als Anwendungsannahme wurde jedoch nur die Normalverteilung zum Testen empfohlen. Es wurde kein Bezug auf die Prüfung der Gleichheit der Varianzen (für ANOVA und  $t$ -Test) oder Ausreißer genommen. In einem Fall wurde die lineare Regression nur als deskriptives Instrument und nicht als inferenzstatistisches Verfahren empfohlen, was bedeutet, dass kein Hypothesentest empfohlen wurde. Insgesamt muss daher festgestellt werden, dass die bereitgestellten Leitlinien nicht vollständig waren.

## 4.2 Analyse des ersten Datensatzes

Der erste Datensatz besteht aus zwei Variablen mit insgesamt 30 Beobachtungen. Die erste Variable enthält normalverteilte Werte aus drei verschiedenen Normalverteilungen mit jeweils 10 Beobachtungen. Bei den drei Normalverteilungen handelt es sich um eine Standardnormalverteilung, eine Normalverteilung mit Mittelwert 1 und Standardabweichung 3 sowie eine Normalverteilung mit Mittelwert 2 und Standardabweichung 5. Die zweite Variable repräsentiert einen Fak-

tor mit den Faktorstufen 1, 2 und 3. Die Zufallszahlen wurden mit der  $R$ -Funktion `rnorm()` generiert, und die Analyse begann mit dem Hochladen der Datendatei in ChatGPT.

Die zu analysierende Frage lautete: Unterscheidet sich die metrische Zielvariable für verschiedene Faktorstufen? Es wird davon ausgegangen, dass eine Person mit den oben beschriebenen rudimentären Statistikkenntnissen diese Frage auf einfache Weise stellt. Die Antwort von *Data Analyst* ist recht lang, weshalb sie hier in mehrere Teile aufgeteilt wird. Der Beginn der Konversation enthält eine Übersicht über den allgemeinen Ansatz, den *Data Analyst* zur Analyse der Daten verfolgt (siehe Abb. 1).

Obwohl dies die üblichen Schritte sind, die man zur Durchführung dieser Analyse befolgen würde, fehlt die Überprüfung der Anwendungsannahmen. *Data Analyst* erstellte daraufhin die Boxplots und berechnete die relevanten deskriptiven Kennzahlen (Abb. 2). Dabei ist zu beachten, dass sich die Standardabweichungen in den drei Faktorstufen erheblich unterscheiden (abweichend vom Original werden hier die deskriptiven Statistiken aller Faktorstufen gezeigt.) *Data Analyst* ging jedoch nicht auf diese Tatsache ein, sondern führte ohne weitere Diskussion eine ANOVA durch. Für das Testergebnis antwortete *Data Analyst* (Abb. 3).

Zunächst ist anzumerken, dass *Data Analyst* den Datensatz tatsächlich vollständig analysiert hat. Jeder Analyseschritt wurde zunächst beschrieben und dann umgesetzt. Das bedeutet, dass auch jemand ohne Sta-

**Der Datensatz enthält die folgenden Spalten:**

- `Unnamed` :  $\emptyset$ : Dies scheint eine Indexspalte zu sein.
- `MetricVar` : Eine metrische Variable.
- `Factor` : Eine kategoriale Variable.

**Wie möchtest du mit diesen Daten verfahren?** [`<_`]

Bitte untersuche statistisch, ob sich `MetricVar` für verschiedene `Factor` -Stufen unterscheidet.

**Um zu untersuchen, ob sich `MetricVar` für verschiedene `Factor` -Stufen unterscheidet, können wir eine statistische Analyse durchführen. Hier ist der Plan:**

- Deskriptive Statistik:** Führe zusammenfassende Statistiken für `MetricVar` für jede Stufe von `Factor` durch.
- Visualisierung:** Erstelle Boxplots, um die Verteilung von `MetricVar` über die verschiedenen `Factor` -Stufen zu visualisieren.
- Statistischer Test:** Führe einen ANOVA-Test durch, um zu prüfen, ob es signifikante Unterschiede der Mittelwerte von `MetricVar` zwischen den `Factor` -Gruppen gibt.

Abb. 1: Beginn der Analyse des ersten Datensatzes.

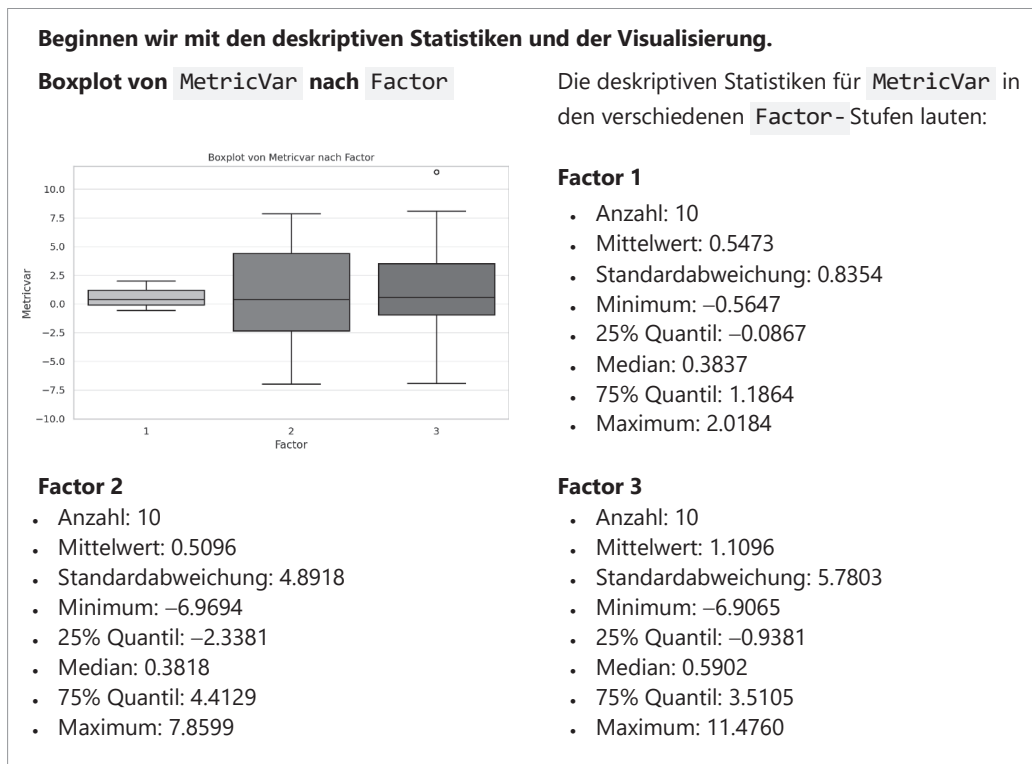


Abb. 2: Boxplots und deskriptive Zahlen

tistikkenntnisse eine inferenzstatistische Datenanalyse durchführen kann, was vor der Existenz generativer KI nicht möglich gewesen wäre.

Bei genauerer Betrachtung wird jedoch auch deutlich, dass diese Analyse in mehrfacher Hinsicht oberflächlich und unvollständig war. Wie bereits erwähnt, wurden die Unterschiede in den Standardabweichungen nicht weiter diskutiert, obwohl die Überprüfung der Anwendungsannahmen dies erfordert hätte. Eine Analyse fehlender Werte und Ausreißer wäre ebenfalls sinnvoll gewesen. Ein weiterer wichtiger Mangel der Analyse besteht darin, dass die geringe Stichprobengröße von 10 Beobachtungen pro Faktorstufe nicht berücksichtigt wurde. Eine Analyse mit der *R*-Funktion `kruskal`.

`test()` ergab einen anderen *p*-Wert ( $p = 0,9758$ ), aber keine andere Testentscheidung. Folglich kann man sagen, dass die von *Data Analyst* durchgeführte Analyse zumindest in diesem Fall insgesamt keine gravierenden Fehler enthielt.

Die Durchführung der Analyse ist natürlich nur ein Teil der Aufgabe. Die Interpretation und das Verständnis der Ergebnisse gehören ebenfalls dazu. Viele der von *Data Analyst* vorgestellten Analysen und Konzepte sind für statistisch unversierte Nutzer höchstwahrscheinlich nicht verständlich, aber auch hier können sie ChatGPT um Hilfe bitten (Abb. 4).

Obwohl die Erklärungen noch viele statistische Fachausdrücke enthalten, ist der letzte Satz auch für jeman-

**Als Nächstes führen wir einen ANOVA-Test durch, um zu bestimmen, ob sich die Mittelwerte von MetricVar signifikant zwischen den Factor-Stufen unterscheiden.**

[<\_]

**Die Ergebnisse des ANOVA-Tests lauten wie folgt:**

- **F-Statistik:** 0.0584
- **p-Wert:** 0.9434

Da der *p*-Wert deutlich größer als 0.05 ist, können wir die Nullhypothese nicht verwerfen. Dies legt nahe, dass **kein statistisch signifikanter Unterschied** zwischen den Mittelwerten von MetricVar über die verschiedenen Factor-Stufen hinweg besteht.

**Falls du weitere Fragen hast oder zusätzliche Analysen brauchst, sag mir einfach Bescheid!**

[<\_]

Abb. 3: ANOVA für den ersten Datensatz

**Ich habe keine statistischen Vorkenntnisse.  
Kannst du mir bitte das Ergebnis erklären?**

Abb. 4: Zweite Eingabeaufforderung für den ersten Datensatz

den, der mit Hypothesentesten nicht vertraut ist, leicht verständlich.

In dieser Hinsicht kann man feststellen, dass *Data Analyst* einen einfach strukturierten Datensatz weitgehend korrekt analysieren kann, sodass auch Personen ohne Statistikkenntnisse Datenanalyseaufgaben durchführen können.

Eine weitere Frage, die offen blieb, war, welches Tool *Data Analyst* für die Durchführung der Analysen verwendete. Auf Nachfrage zeigte sich, dass Python verwendet wurde, anschließend wurde der Code ausgegeben. Dieser Code wurde mit Python getestet, wodurch die Ergebnisse bestätigt wurden.

### 4.3 Analyse des zweiten Datensatzes

Der zweite Datensatz besteht ebenfalls aus zwei Variablen, einer numerischen und einem Faktor mit drei Stufen, aber er umfasst 180 Beobachtungen, mit je 60 pro Faktorstufe. Darüber hinaus wurde der Datensatz so erstellt, dass ANOVA und Kruskal-Wallis-Test zu unterschiedlichen Testergebnissen führen.

Zu diesem Zweck wurden 5 Ausreißer und 25 fehlende Werte hinzugefügt. Diese wurden zufällig aus den 180 Werten unter Verwendung einer Stichprobe ohne Zurücklegen ausgewählt. Bei den Ausreißern wurden die ursprünglichen Werte mit 6 multipliziert. Die Boxplots in Abb. 6 veranschaulichen den Datensatz.

#### Natürlich! Hier ist eine vereinfachte Erklärung des Ergebnisses:

##### 1. Was wir gemacht haben:

- Wir wollten prüfen, ob sich die durchschnittlichen Werte (Mittelwerte) von **MetricVar** zwischen den verschiedenen **Factor**-Stufen (1, 2 und 3) unterscheiden.
- Dafür haben wir einen statistischen Test namens **ANOVA** (Analyse der Varianz) verwendet. Dieser Test vergleicht die Mittelwerte von **MetricVar** für jede **Factor**-Stufe, um zu sehen, ob sie sich signifikant voneinander unterscheiden.

##### 2. Ergebnis:

- Der ANOVA-Test liefert zwei wichtige Kennzahlen: die **F-Statistik** und den **p-Wert**.
- Die **F-Statistik** (0.0584) misst das Verhältnis der Varianz zwischen den Gruppen zur Varianz innerhalb der Gruppen. Für unsere Interpretation ist aber der **p-Wert** entscheidend.

##### 3. p-Wert:

- Der **p-Wert** (0.9434) sagt aus, wie wahrscheinlich es ist, dass die beobachteten Unterschiede rein zufällig zustande gekommen sind.
- Allgemein gilt: Ist der p-Wert kleiner als 0,05, dann ist der Unterschied statistisch signifikant – es spricht also etwas dafür, dass sich die Gruppen wirklich unterscheiden.
- In unserem Fall ist der p-Wert **0.9434**, also deutlich größer als 0.05. Das heißt, es ist sehr wahrscheinlich, dass die beobachteten Unterschiede rein zufällig sind – also **keine echten Unterschiede zwischen den Gruppen** vorliegen.

##### 4. Fazit:

- Weil der p-Wert so hoch ist, schließen wir, dass **kein signifikanter Unterschied** zwischen den Mittelwerten von **MetricVar** über die verschiedenen **Factor**-Stufen besteht.
- Einfacher gesagt: **MetricVar** unterscheidet sich **nicht nennenswert** zwischen den **Factor**-Stufen – zumindest basierend auf unseren Daten.

Abb. 5: Erläuterung zu den Ergebnissen des ersten Datensatzes

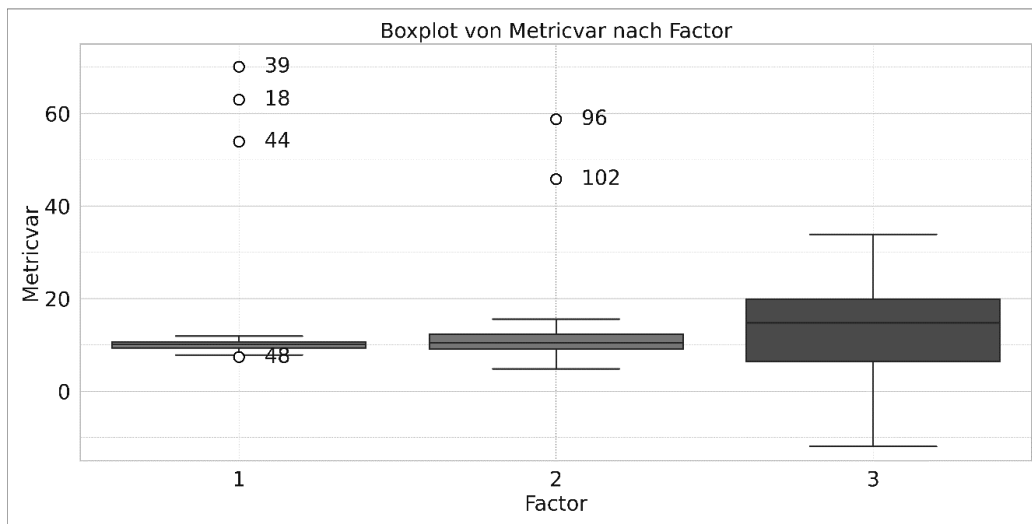


Abb. 6: Boxplots für den zweiten Datensatz

Die Analyse begann mit dem Hochladen der Daten in ChatGPT. *Data Analyst* lud den Datensatz ohne jegliche Vorverarbeitung hoch.

Der Benutzer wurde zwar über fehlende Werte informiert, ChatGPT erwähnte jedoch die Ausreißer nicht. Für diesen Datensatz führte *Data Analyst* die vollständige Analyse ohne Erläuterung des Verfahrens und ohne Überprüfung der Anwendungsannahmen durch. Die deskriptive Statistik beschränkte sich auf Mittelwert und Standardabweichung, für die ANOVA wurden  $F$ - und entsprechender  $p$ -Wert angegeben. Das Vorhandensein wesentlicher Ausreißer wurde weder erkannt noch untersucht.

Würde jedoch ein gegen Ausreißer robustes Verfahren wie der Kruskal-Wallis-Test verwendet, ergäbe sich ein signifikanter Unterschied zwischen den Faktorstufen ( $p = 0,032$  mit der Funktion `kruskal.test()` in R). *Data Analyst* liefert hier ein falsches Ergebnis, gerade weil die Annahmen für die ANOVA nicht überprüft wurden. Abb. 7 enthält die gesamte Ausgabe einschließlich des Prompts.

#### 4.4 Stabilitätsprüfungen und Analysen weiterer Datensätze

Die beiden hier vorgestellten Datensätze wurden zur Stabilitätsprüfung erneut analysiert. ChatGPT bietet die Möglichkeit, die Ergebnisse früherer Eingaben in aktuellen Eingaben zu verwenden; diese Option wurde für die Stabilitätsprüfungen deaktiviert.

Jeder der Datensätze wurde zwei weitere Male analysiert, sodass für jeden Datensatz drei Analysen durchgeführt wurden. Dabei zeigte sich, dass *Data Analyst* in einer der beiden Wiederholungsanalysen die Anwendungsanforderungen sowohl für den ersten als auch für den zweiten Datensatz überprüft hat, mit dem Ergeb-

nis, dass anstelle einer klassischen ANOVA ein anderer Test (Welch-ANOVA für den ersten Datensatz und Kruskal-Wallis-Test für den zweiten Datensatz) durchgeführt wurde. Der Kruskal-Wallis-Test für den zweiten Datensatz erkannte und meldete den signifikanten Unterschied zwischen den drei Faktorstufen.

Die beiden verbleibenden Datensätze, die für einen  $t$ -Test für zwei Stichproben und eine lineare Regression geeignet waren, wurden ebenfalls jeweils dreimal mit *Data Analyst* analysiert. Für den Datensatz für den  $t$ -Test für zwei Stichproben führte *Data Analyst* in allen drei Analysen einen  $t$ -Test durch, ohne die Anwendungsannahmen zu überprüfen. In einer der drei Analysen wurden auch die Mittelwerte beider Messreihen angezeigt, während in den beiden anderen nur  $t$ - und  $p$ -Werte dargestellt wurden.

Bei der Analyse des für die lineare Regression geeigneten Datensatzes, bestehend aus drei unabhängigen und einer abhängigen Variablen, führte *Data Analyst* in allen drei Fällen multiple lineare Regressionen durch. In einer der drei Analysen berechnete ChatGPT vor der Analyse paarweise Korrelationen zwischen den unabhängigen und der abhängigen Variablen. Auch nur in einer der drei Analysen wies *Data Analyst* darauf hin, dass die lineare Regression Anwendungsannahmen hat, testete diese jedoch nicht. In den beiden anderen Analysen wurden keine Anwendungsannahmen und deren Überprüfung erwähnt.

## 5 Diskussion

Die beiden oben dargestellten Beispiele sowie die Analyse der beiden anderen Datensätze zeigen, dass die Schlussfolgerungen von Frey und Osborne (2023) nur teilweise zutreffen, wenn sie auf die statistische Datenanalyse angewendet werden. Obwohl die Überprü-

**Die Datei enthält drei Spalten:**

1. **Unnamed** : 0 – scheint ein Index oder Identifikationslabel zu sein.
2. **MetricVar** – vermutlich eine Variable oder Messgröße von Interesse.
3. **Factor** – scheint dazu zu dienen, die Daten in Gruppen zu kategorisieren.  
In diesem Fall haben alle Werte den Wert 1.

**Es gibt einige fehlende Werte in der Spalte MetricVar.**

**Wie möchtest du mit diesen Daten weiter verfahren? [<\_]**

Bitte untersuche statistisch, ob sich die Werte der Variable **MetricVar** zwischen den verschiedenen Stufen von **Factor** unterscheiden.

**Die deskriptiven Statistiken für MetricVar , gruppiert nach Factor–Stufen, lauten wie folgt:**

- **Factor 1:** Mittelwert = 12.87, Standardabweichung = 12.37
- **Factor 2:** Mittelwert = 12.08, Standardabweichung = 8.64
- **Factor 3:** Mittelwert = 11.13, Standardabweichung = 10.49

Ein ANOVA-Test wurde durchgeführt, um zu prüfen, ob sich **MetricVar** signifikant zwischen den verschiedenen **Factor**–Stufen unterscheidet. Das Ergebnis lautet:

- **F-Statistik** = 0.135
- **p-Wert** = 0.873

Da der p-Wert deutlich größer als 0,05 ist, gibt es **keinen statistisch signifikanten Unterschied** in **MetricVar** zwischen den verschiedenen **Factor**–Stufen.

**Möchtest du weitere Analysen oder Visualisierungen? [<\_]**

Abb. 7: Vollständige Analyse des zweiten Datensatzes

fung der Anwendungsannahmen ein fester Bestandteil der Testauswahl ist, hat ChatGPT diesen Schritt nicht konsequent oder in einigen Fällen gar nicht durchgeführt. Dies ist aus zwei Gründen überraschend: Erstens weist *Data Analyst* auf die Notwendigkeit hin, die Anwendungsannahmen zu überprüfen, wenn es darum geht, die Schritte zur Durchführung eines *t*-Tests, einer ANOVA oder einer linearen Regression zu erklären. Zweitens zeigen die Stabilitätsprüfungen, dass *Data Analyst* durchaus in der Lage ist, die Anwendungsanforderungen zu überprüfen, was bedeutet, dass eine vollständige statistische Datenanalyse möglich wäre. Eine Person, die über wenig oder gar keine Kenntnisse in Statistik verfügt und Datenanalyseaufgaben mit ChatGPT durchführen möchte, kann den vorgestellten Ansatz nicht befolgen oder, schlimmer noch, wäre nicht in der Lage, die Mängel der Ergebnisse zu erkennen. Selbst wenn diese Person zunächst nach den erforderlichen Analyseschritten fragt und diese dann Schritt für Schritt mit geeigneten Prompts ausführt, können wichtige Analyseschritte, wie z. B. die Überprüfung fehlender Werte, dennoch fehlen (siehe Abschnitt 4.1).

Selbst wenn eine Person ohne statistische Kenntnisse nun technisch in der Lage ist, einen Datensatz mit KI zu analysieren, sind dennoch statistische Kenntnisse erforderlich, um die richtigen Eingabeaufforderungen zu formulieren und sicherzustellen, dass die KI keinen Schritt der Analyse auslässt.

Diese Erkenntnis hat erhebliche Auswirkungen darauf, wie Statistik unterrichtet werden kann oder sollte. Nach aktuellem Stand der Debatte wird empfohlen, ChatGPT einzusetzen, um die Lernenden mit potenziellen Fallstricken vertraut zu machen, die entstehen können, wenn sie die Erklärungen von ChatGPT ungeprüft übernehmen (Xing, 2024).

Die größte Wirkung von *Data Analyst* im Bereich der speziellen und generativen KI ist jedoch in den praktischen Übungen zu erwarten. Eine generative KI wie *Data Analyst* kann die Rolle der Auswertungssoftware übernehmen. Es wird erwartet, dass diese Art der Nutzung den Lernenden die Arbeit erleichtert, da weniger Wert auf die Programmierung selbst gelegt werden muss. Stattdessen können Lehrkräfte Übungen ein-

bauen, welche den Lernenden helfen, mit Hilfe von KI effizienter und genauer zu programmieren. Dadurch verlagert sich der Fokus vom Finden des richtigen Befehls oder der richtigen Funktion hin zum genaueren Untersuchen und Verstehen der Daten. Infolgedessen sind die Lernenden besser in der Lage, die Ergebnisse statistischer Auswertungssoftware richtig zu interpretieren, wodurch Ziel 8 des GAISE-Berichts (GAISE, 2016, S. 8) erfüllt wird.

Dies gilt auch für das Verfassen von Hausarbeiten in Statistik und anderen Arten von Statistikübungsaufgaben. Es ist nicht mehr notwendig, die Studierenden darauf zu prüfen, wie gut sie ein Problem in einer Programmiersprache wie R oder Python lösen können. Der Schwerpunkt sollte nun vielmehr darauf liegen, ob die Studierenden die richtigen Analyseschritte durchführen. Darüber hinaus kann eine generative KI wie ChatGPT Code in R oder Python erstellen, wodurch Übungsaufgaben zum Testen der Programmierkenntnisse überflüssig werden.

Es bleibt jedoch unerlässlich, die theoretischen Konzepte der Statistik zu vermitteln und diese Lerninhalte durch Anwendungsübungen zu festigen. Denn eine solide Grundlage in statistischer Methodik ist nach wie vor erforderlich, um die richtigen Eingabeaufforderungen zu formulieren und vor allem die Vollständigkeit der Ergebnisse zu verstehen. Diese Schlussfolgerung findet sich elegant in den Worten von Bobula (2024, S. 12) wieder:

Sobald die Studierenden über die erforderlichen Kenntnisse und Fähigkeiten verfügen, können sie die von der KI generierten Ergebnisse bewerten, Ungenauigkeiten aussortieren und nützliche und hilfreiche Komponenten der von der Technologie angebotenen Antworten nutzen. Dieser Ansatz fördert die Anwendung von kritischem Denken in ihrem Studium.

Hier sind noch weitere Aspekte zu berücksichtigen. Wenn eine generative KI die Datenanalyse durchführt, bedeutet dies immer, dass die Daten auf einen fremden Server hochgeladen werden. Dies ist natürlich nur zulässig, wenn keine Datenschutzbestimmungen oder Geheimhaltungsvereinbarungen verletzt werden. Wenn eine generative KI die Datenanalyse durchführen soll, müssen solche Aspekte sorgfältig geprüft werden. Es gibt auch ethische Bedenken, beispielsweise bei der Bewertung der eigenständigen Leistung in Aufsätzen oder anderen Arten von Aufgaben. Wenn *Data Analyst* lediglich die Rolle einer Auswertungs-Software wie R oder Python übernimmt und die Eingabeaufforderungen alltägliche Sprachäquivalente von Programmierbefehlen darstellen, erscheint dies vergleichsweise unproblematisch. So wie es üblich ist, die verwendete Ana-

lysesoftware mit Versionsnummer zu nennen, könnte man hier ChatGPT *Data Analyst* 4.0 angeben.

Wenn jedoch die Verwendung von *Data Analyst* über den Rahmen einer Auswertungs-Software hinausgeht, beispielsweise wenn zumindest Teile der Datenanalyse vollständig von *Data Analyst* übernommen werden, muss hinterfragt werden, inwieweit dies noch eine eigenständige Leistung der Studierenden darstellt und darstellen kann. Diese Frage ist nicht nur für Prüfungsleistungen relevant, sondern auch für wissenschaftliche Publikationen. Peng und Zhao (2024, S. 2) werfen beispielsweise die Frage auf, ob einer generativen KI wie ChatGPT die Urheberschaft zugeschrieben werden sollte. Für Datenanalysen im Rahmen von Aufgaben, die in die Gesamtleistung der Studierenden einfließen, ist es ratsam, dass klare Richtlinien festgelegt werden, ob und inwieweit ChatGPT als Auswertungsinstrument zulässig ist.

Abschließend müssen die Grenzen der vorliegenden Studie angesprochen werden. Die Eingabeaufforderungen wurden unter der Annahme formuliert, dass der Nutzer keine Kenntnisse in Statistik hat. Es wurde jedoch noch nicht getestet, ob jemand ohne statistische Kenntnisse tatsächlich so vorgegangen wäre. Darüber hinaus waren die hier analysierten Beispiele eher einfach, mit vergleichsweise einfach strukturierten Datensätzen. Hier müsste überprüft werden, wie eine generative KI wie ChatGPT *Data Analyst* mit komplexeren Datensätzen und fortgeschritteneren statistischen Verfahren umgeht.

Schließlich sollte in einem Experiment untersucht werden, ob ein Statistikkurs, der generative KI einsetzt, tatsächlich zu einer Verbesserung des Unterrichts führt. So könnte beispielsweise getestet werden, ob Kursteilnehmer aus einem Kurs, der generative KI einsetzt, zu besseren Ergebnissen kommen als Kursteilnehmer aus einem Kurs, der nur Auswertungs-Software wie R oder Python verwendet.

## Danksagung

Der Autor bedankt sich bei zwei anonymen Gutachtern (des Originalartikels in *Teaching Statistics*) für ihre wertvollen Hinweise und Verbesserungsvorschläge sowie bei Herrn Borovcnik für die Übersetzung des englischen Originaltextes ins Deutsche. Darüber hinaus hat der Autor bei der Erstellung dieses Manuskripts die folgenden generativen KI-Tools verwendet: ChatGPT 4.0 zusammen mit ChatGPT *Data Analyst* und der Suchmaschine perplexity.ai. ChatGPT *Data Analyst* wurde zur Durchführung der in der Arbeit beschriebenen Datenanalysen verwendet, ChatGPT diente zur Un-

terstützung beim Verfassen der Zusammenfassung und perplexity.ai wurde zur Suche nach relevanten Publikationen genutzt. Open-Access-Finanzierung ermöglicht und organisiert durch Projekt DEAL.

Zur authentischen Wiedergabe der Chat-Protokolle mit ChatGPT *Data Analyst* vom Englischen ins Deutsche wurde *Scholar GPT* von ChatGPT verwendet. Die Daten, der R code zu ihrer Generierung und der Chatverlauf mit ChatGPT *Data Analyst* sind zugänglich: <https://doi.org/10.25625/OHXGA4>

## References

- Bobula, M. (2024). Generative artificial intelligence (AI) in higher education: A comprehensive review of challenges, opportunities, and implications. *Journal of Learning Development in Higher Education*, 30, 1–27. <https://doi.org/10.47408/jldhe.vi30.1137>
- Chukwuere, J. E. (2024). The future of generative AI chatbots in higher education. *arXiv*. <https://doi.org/10.48550/arXiv.2403.13487>
- Cribben, I. & Zeinali, Y. (2023). The benefits and limitations of ChatGPT in business education and research: A focus on management science, operations management and data analytics. *SSRN eJournals*. <https://doi.org/10.2139/ssrn.4404276>
- Demirci, O., Hannane, J. & Zhu, X. (2023). Who is AI replacing? The impact of generative AI on online freelancing platforms. *SSRN eJournals*. <https://dx.doi.org/10.2139/ssrn.4602944>
- Duraisamy, K., Agrawal, V., Athey, B. D., Daniels, T., Das, K., Gaspar, N., Gere, A. R., Hayward, C., Hemphill, L., Jones, R., Melville, N. P., Méndez, D., Mihalcea, R., Rodriguez, J., Spector-Bagdady, K., Tinkle, T., van Lent, C., Watta, P., Kwiatkowski, M., & Burns, T. (2023). *Generative artificial intelligence advisory committee report*. University of Michigan. <https://drive.google.com/file/d/101zhMpzr67SRRePbbxfHc87j-5mSlkuOL/view>
- Ellis, A. R. & Slade, E. (2023). A new era of learning: Considerations for ChatGPT as a tool to enhance statistics and data science education. *Journal of Statistics and Data Science Education*, 31(2), 128–133. <https://doi.org/10.1080/26939169.2023.2223609>
- Frey, C. B. & Osborne, M. (2023). *Generative AI and the future of work: A reappraisal. The Oxford Martin working paper series on the future of work, No. 2023*. University of Oxford. <https://www.oxfordmartin.ox.ac.uk/downloads/academic/2023-FoW-Working-Paper-Generative-AI-and-the-Future-of-Work-A-Reappraisal-combined.pdf>
- GAISE (2016). *Guidelines for Assessment and Instruction in Statistics Education College Report*. College Report ASA Revision Committee. [https://www.amstat.org/asa/files/pdfs/gaise/gaisecollege\\_full.pdf](https://www.amstat.org/asa/files/pdfs/gaise/gaisecollege_full.pdf)
- Huang, Y., Wu, R., He, J., & Xiang, Y. (2024). Evaluating ChatGPT-4.0's data analytic proficiency in epidemiological studies: A comparative analysis with SAS, SPSS, and R. *Journal of Global Health*, 14 (2024), 04070. <https://doi.org/10.7189/jogh.14.04070>
- Irvine, D. J., Halloran, L. J. S., & Brunner, P. (2023). Opportunities and limitations of the ChatGPT advanced data analysis plugin for hydrological analyses. *Hydrological Processes*, 37 (10), e15015. <https://doi.org/10.1002/hyp.15015>
- Neumann, M., Rauschenberger, M., & Schön, E.-M. (2023). „We need to talk about ChatGPT“: The future of AI and higher education. In *2023 IEEE/ACM 5th international workshop on software engineering education for the next generation (SEENG)* (pp. 29–32). <https://doi.org/10.1109/SEENG59157.2023.00010>
- OpenAI (2023a). *Data Analyst*. <https://chat.openai.com/g-gHMNcP6w7d-data-analyst>
- OpenAI (2023b). *Introducing GPTs*. <https://openai.com/blog/introducing-gpts/>
- Peng, L. & Zhao, B. (2024). Navigating the ethical landscape behind ChatGPT. *Big Data & Society*, 11(1), 1–5. <https://doi.org/10.1177/20539517241237488>
- Rasul, T., Nair, S., Kalendra, D., Robin, M., d. Oliveira Santini, F., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning & Teaching*, 6 (1), 41–56. <https://doi.org/10.37074/jalt.2023.6.1.29>
- Rudolph, J., Tan, S., & Aspland, T. (2023). Editorial 6(2): Personal digital assistant or job killer? Generative AI and the teaching profession in higher education. *Journal of Applied Learning & Teaching*, 6 (2), 7–16. <https://doi.org/10.37074/jalt.2023.6.2>
- Stansbury, J. A., Lausch, S., Zahadat, N., & Kelly, D. (2023). *White paper: AI perceptions at the University of Baltimore*. Center for Excellence in Learning, Teaching and Technology, <https://drive.google.com/file/d/1ufdagea0Xm8TpiKsyvbr1Kp-kpez3z6Z/view?pli=1>
- Statista Research Department (2024). *Wofür wird ChatGPT in Ihrem Unternehmen genutzt?* <https://de.statista.com/statistik/daten/studie/1401309/umfrage/chatgtp-nutzung-in-unternehmen/>
- Xing, Y. (2024). Exploring the use of ChatGPT in learning and instructing statistics and data analytics. *Teaching Statistics*, 46 (2), 95–104. <https://doi.org/10.1111/test.12367>

## Anschrift des Verfassers

Joachim Schwarz  
Fachbereich Wirtschaft  
University of Applied Sciences Emden-Leer  
Constantiaplatz 4, 26723 Emden  
[joachim.schwarz@hs-emden-leer](mailto:joachim.schwarz@hs-emden-leer)