



PEARL

Biostatistik in Ophthalmologie und Optometrie: Eine Artikelserie zur Unterstützung evidenzbasierter Entscheidungsprozesse

Oehring, Daniela; Serra, Pedro Miguel

Published in:
Optometry & Contact Lenses

Publication date:
2024

Document version:
Publisher's PDF, also known as Version of record

Link:
[Link to publication in PEARL](#)

Citation for published version (APA):
Oehring, D., & Serra, P. M. (2024). Biostatistik in Ophthalmologie und Optometrie: Eine Artikelserie zur Unterstützung evidenzbasierter Entscheidungsprozesse. *Optometry & Contact Lenses*, 5(1), 1. Article 1. <https://ocli-online.de/de/biostatistik-ophthalmologie-und-optometrie>

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Wherever possible please cite the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.



Biostatistik in Ophthalmologie und Optometrie: Eine Artikelserie zur Unterstützung evidenzbasierter Entscheidungsprozesse

Teil 1: Grundlagen statistischer Methoden in der Augenheilkunde und Optometrie

Daniela Oehring^{1,2}, Pedro Miguel Serra^{1,3}

¹ Ph.D. · ² Faculty of Health, University of Plymouth, UK · ³ Ophthalmology Clinic Vista Sanchez Trancon, Spain

Eingereicht 31. Oktober 2024; angenommen 28. November 2024
Received 31 October 2024; accepted 28 November 2024

COE Multiple-Choice-Fragebogen siehe Seite xx



Zusammenfassung

Zweck. Fortschritte in der Augen- und Sehversorgung basieren auf der Anwendung von Forschungsergebnissen und der präzisen Interpretation klinischer Daten. Dennoch treten in der Augen- und Sehforschung (EVR) häufig Forschungverschwendungen auf, bedingt durch methodische Mängel und unangemessene statistische Analysen. Dies untergräbt die Validität von Studien und führt zu ineffizienter Nutzung erheblicher finanzieller Ressourcen. Dieser Artikel, der erste Teil der Serie „Entwicklung statistischer Kompetenzen in der Augenheilkunde: Eine Reihe zur Unterstützung evidenzbasierter Entscheidungsprozesse“, zielt darauf ab, diese Herausforderungen anzugehen, indem er die statistische Kompetenz von Fachkräften in der Augenheilkunde stärkt.

Material und Methoden. Anhand einer umfassenden narrativen Literaturübersicht und der Erstellung simulierter klinischer Datensätze identifiziert diese Studie wesentliche statistische Konzepte, häufige Fallstricke und bewährte Praktiken, die für die EVR relevant sind. Die Literaturrecherche umfasste Datenbanken wie PubMed, Scopus und Web of Science und konzentrierte sich auf Fachartikel und professionelle Lehrbücher zur statistischen Methodologie. Simulierte Datensätze, die realistische klinische Messungen wie Pupillendurchmesser, Refraktionsfehler, zentrale Hornhautdicke und Augeninnendruck abbilden, wurden mit Python (V3.12.4) erstellt, um zentrale statistische Prinzipien und deren Anwendungen zu veranschaulichen.

Ergebnisse. Das Projekt beschreibt grundlegende statistische Konzepte, einschließlich Datentypen (nominal, ordinal, me-

trisch), Techniken der Datenaufbereitung, den Umgang mit fehlenden Werten und Ausreißern sowie die Anwendung deskriptiver Statistiken. Zudem untersucht es die Eigenschaften von Datenverteilungen, die Überprüfung auf Normalverteilung und Methoden zur Datentransformation, um robuste und zuverlässige statistische Analysen zu gewährleisten. Durch die Verknüpfung von theoretischem Wissen mit praktischen Beispielen soll dieser Teil Fachkräfte in der Augenheilkunde befähigen, Forschung kritisch zu bewerten, evidenzbasierte Praktiken zu integrieren und einen bedeutenden Beitrag zur wissenschaftlichen Gemeinschaft zu leisten.

Fazit. Diese Studie schafft einen grundlegenden Rahmen für die Verbesserung der statistischen Kenntnisse von Augenspezialisten, indem sie wesentliche statistische Konzepte und bewährte Verfahren in der EVR untersucht. Durch die Auseinandersetzung mit häufigen methodischen Fehlern und unsachgemäßen Analysen soll die Verschwendung von Forschungsergebnissen reduziert und die Validität von Studien verbessert werden. Letztendlich soll diese Initiative eine genauere Dateninterpretation, eine bessere klinische Entscheidungsfindung und eine bessere Patientenversorgung im Bereich der Augen- und Sehgesundheit fördern.

Schlüsselwörter

Statistische Grundbildung, Augenpflege, Klinische Entscheidungsfindung, Augen- und Sehforschung, Deskriptive Statistik, Datenanalyseverfahren, Methodologie der Forschung

Advancing Statistical Literacy in Eye Care: A Series for Enhanced Clinical Decision-Making

Part 1: Introduction to Statistical Tools for Eye Care Research

Abstract

Purpose. Advancements in eye and vision care hinge on the rigorous application of research and the precise interpretation of clinical data. However, the field of Eye and Vision Research (EVR) frequently encounters research waste attributed to methodological flaws and improper statistical analyses, undermining the validity of studies and inefficiently utilising substantial financial resources. This paper, the first instalment in the series “Advancing Statistical Literacy in Eye Care: A Series for Enhanced Clinical Decision-Making,” aims to address these challenges by enhancing the statistical literacy of eye care professionals.

Material and Methods. Through a comprehensive narrative literature review and the generation of simulated clinical datasets, this study identifies essential statistical concepts, common pitfalls, and best practices pertinent to EVR. The literature review used multiple databases, including PubMed, Scopus, and Web of Science, focusing on peer-reviewed articles and professional textbooks relevant to statistical methodologies. Simulated datasets reflecting realistic clinical measurements, such as pupil diameter, refractive error, central corneal thickness, and intraocular pressure, were created using Python (V3.12.4) to illustrate key statistical principles and their applications.

Results. The paper explores fundamental statistical concepts, including data types (nominal, ordinal, metric), data preparation techniques, handling missing data and outliers, and applying descriptive statistics. Additionally, it explores data distribution characteristics, normality assessment, and data transformation methods to ensure robust and reliable statistical analyses. By bridging theoretical knowledge with practical examples, this instalment seeks to equip eye care professionals with the tools to critically evaluate research, integrate evidence-based practices, and contribute meaningfully to the scientific community.

Conclusion. This study establishes a foundational framework to enhance statistical literacy among eye care professionals by exploring essential statistical concepts and best practices in EVR. By addressing common methodological flaws and improper analyses, it aims to reduce research waste and improve the validity of studies. Ultimately, this initiative is expected to promote more accurate data interpretation, better clinical decision-making, and improved patient care in the field of eye and vision health.

Keywords

Statistical Literacy, Eye Care, Clinical Decision-Making, Eye and Vision Research, Descriptive Statistics, Data Analysis, Research Methodology

Einleitung

Fortschritte in der Augen- und Sehversorgung setzen gründliche Forschung und präzise Interpretation klinischer Daten voraus. Angesichts eines stetig wachsenden Korpus wissenschaftlicher Literatur ist es für Augenspezialisten unabdingbar, angemessene statistische Methoden zu verstehen und anzuwenden. Eine fundierte statistische Bildung ist entscheidend, um Forschungsergebnisse kritisch zu bewerten, evidenzbasierte Praktiken zu integrieren und zur wissenschaftlichen Gemeinschaft beizutragen.

Evidenzbasierte Praxis ist essenziell für die Augen- und Sehpflege, sie leitet Diagnosen, Behandlungen und die gesamte Patientenversorgung, gestützt auf hochwertige Forschung und solide statistische Methoden.¹⁻⁵ Dennoch sieht sich das Feld erheblichen Herausforderungen gegenüber, da viele Studien durch methodische Mängel und unzulängliche statistische Analysen gekennzeichnet sind, was die Validität

der Forschung untergräbt und zu unnötiger Ressourcenverschwendung führt.³⁻⁵

In der Ophthalmologie ist Forschungsverschwendung ein gravierendes Problem, das zu Ineffizienzen und verschwendeten Ressourcen führt.⁵ Beispielsweise beziehen sich nur 22,4% der Phase-III-Studien in der Augenheilkunde auf systematische Übersichten als Begründung für die Studie, was Chancen verpasst, auf bestehenden Erkenntnissen aufzubauen und somit zu redundanten Forschungen führt.⁶ In chirurgischen Studien wird das Problem durch mangelnde methodische Strenge verschärft, was oft zu unvollständigen oder ungültigen Ergebnissen führt. Schätzungen zufolge könnten bis zu 85% der weltweiten Gesundheitsforschung, die einen erheblichen Teil der jährlichen Ausgaben von 200 Milliarden Dollar ausmachen, durch Nichtveröffentlichung, unklare Berichterstattung und das Fehlen systematischer Überprüfungen im Studiendesign verschwendet werden. In der Augen- und Sehforschung resultiert diese Verschwendung

häufig aus unzureichenden Stichprobengrößen, ungeeigneten statistischen Tests oder Fehlinterpretationen der Ergebnisse.⁷⁻⁹ Eine Überprüfung der augenheilkundlichen Literatur ergab, dass ein erheblicher Teil der veröffentlichten Studien statistische Ungenauigkeiten aufweist, was die Glaubwürdigkeit der Forschung weiter untergräbt, und den Fortschritt evidenzbasierter Praxis verlangsamt.⁸⁻¹⁰

Die Verbesserung der statistischen Bildung unter Forschern und Klinikern ist entscheidend, um diese Probleme anzugehen. Ein solides Verständnis statistischer Prinzipien und deren korrekte Anwendung können die Forschungsqualität verbessern, Verschwendung reduzieren und die evidenzbasierte Patientenversorgung durch zuverlässige und valide Ergebnisse unterstützen.

Methoden

Dieser Artikel ist der erste Teil einer fünfteiligen Serie, die darauf abzielt, Verständnis von Biostatistik in EVR zu verbessern. Um dieses Ziel zu erreichen, wurde ein zweigleisiger methodologischer Ansatz gewählt: eine umfassende Literaturübersicht und die Erstellung simulierter Datensätze, um wichtige statistische Konzepte, die für die Augen- und Sehforschung (EVR) relevant sind, zu veranschaulichen.

Wir führten eine narrative Literaturübersicht durch, um wesentliche statistische Konzepte, häufige Fallstricke und bewährte Verfahren bei der Anwendung statistischer Methoden innerhalb der EVR zu identifizieren. Die Literatursuche wurde in mehreren Datenbanken, darunter PubMed, Scopus und Web of Science, durchgeführt und umfasste Veröffentlichungen bis Oktober 2024.

Suchbegriffe: Die Suche nutzte Kombinationen von Begriffen wie „vision science“, „statistical methods“, „data analysis“, „clinical research“, „biostatistics“, „eye care“, „research methodology“, „data distribution“, „missing data“, „outliers“ und „statistical literacy“.

Einschlusskriterien: Es wurden begutachtete Artikel und Fachbücher eingeschlossen, die sich auf statistische Methoden konzentrierten, die in der Ophthalmologie und Optometrie anwendbar sind und Studien, die häufige statistische Fehler in der EVR hervorheben sowie Leitlinien zu bewährten Praktiken für die Datenanalyse in der klinischen Forschung und Artikel, die die Bedeutung der statistischen Bildung in EVR betonen.

Ausschlusskriterien: Artikel, die nicht direkt mit statistischen Methoden in der Augenpflege zusammenhängen, nicht-englischsprachige Publikationen, Konferenzabstracts ohne Volltexte und Studien, denen methodologische Details fehlen, wurden ausgeschlossen.

Datenextraktion und -synthese: Relevante Artikel wurden zunächst anhand von Titel und Abstract ausgewählt, gefolgt von Volltextüberprüfungen. Zu den extrahierten Schlüsselinformationen gehörten diskutierte statistische Konzepte,

identifizierte häufige Fehler, Empfehlungen für bewährte Verfahren und Implikationen für die klinische Entscheidungsfindung. Die Ergebnisse wurden thematisch synthetisiert, um einen umfassenden Überblick über grundlegende statistische Konzepte, Datentypen, Datenvorbereitung und deskriptive Statistiken zu bieten, die für Augenpflegefachkräfte relevant sind.

Generierung simulierter Daten: Um praktische Beispiele und visuelle Illustrationen der diskutierten statistischen Konzepte zu liefern, wurden simulierten Datensätze mit Python (V3.12.4)¹¹ generiert. Die Datensätze wurden so gestaltet, dass sie realistische klinische Daten widerspiegeln, die in der EVR häufig vorkommen, mit einem Fokus auf Pupillendurchmesser (n=1000), Refraktionsfehler (n=1000) und zentrale Hornhautdicke sowie Augeninnendruck (CCT, IOP, n=1000). Die Daten wurden modelliert, um eine gegebene Verteilung mit Parametern zu folgen (Mittelwert und Standardabweichung), die typische klinische Beobachtungen widerspiegeln (Pupillengröße,¹²⁻¹⁷ Refraktionsfehler,¹⁸⁻²⁴ CCT und IOP.²⁵⁻³⁰

Die Generierung und Analyse der Daten erfolgte mit Python unter Verwendung von Bibliotheken wie NumPy (V2.0.1)³¹ für numerische Berechnungen, pandas³² für Datenmanipulation und matplotlib³³ sowie seaborn³⁴ für die Datenvisualisierung. Die simulierten Datensätze wurden mit statistischen Modellen generiert, die für jede Variable geeignet sind. Zum Beispiel wurden für den Pupillendurchmesser Normalverteilungen verwendet, während die Daten zum Refraktionsfehler so modelliert wurden, dass sie eine Schiefe aufweisen, die ähnlich wie realweltliche klinische Daten ist.

Die simulierten Datensätze sind über das Open Science Framework (OSF) unter DOI: 10.17605/OSF.IO/6EYMW³⁵ zugänglich, was Transparenz gewährleistet und es den Lesern ermöglicht, die präsentierten Analysen zu replizieren.

Da die Studie simulierte Daten und eine Überprüfung vorhandener Literatur umfasste, war keine ethische Genehmigung erforderlich. Die simulierten Daten stellen keine echten Patienteninformationen dar, wodurch Bedenken hinsichtlich der Vertraulichkeit und Privatsphäre von Patienten ausgeräumt werden.

Indem dieser Artikel eine gründliche Literaturübersicht mit der Erstellung simulierter klinischer Datensätze kombiniert, vereint er theoretisches Wissen und praktische Anwendung. Dieser integrative Ansatz soll das Verständnis vertiefen und die Anwendung statistischer Werkzeuge in der klinischen Entscheidungsfindung im Bereich der Augenpflege erleichtern.

Grundlegende statistische Konzepte

Die Interpretation von Kohorten und Populationen beginnt mit der Sammlung einzelner Datenpunkte aus einer Stichprobe, die die Zielbevölkerung repräsentiert. Jeder Datenpunkt, wie beispielsweise der Pupillendurchmesser (**Bild 1**), liefert spezifische Informationen über ein einzelnes Subjekt. Werden diese Datenpunkte aggregiert, entsteht ein Daten-

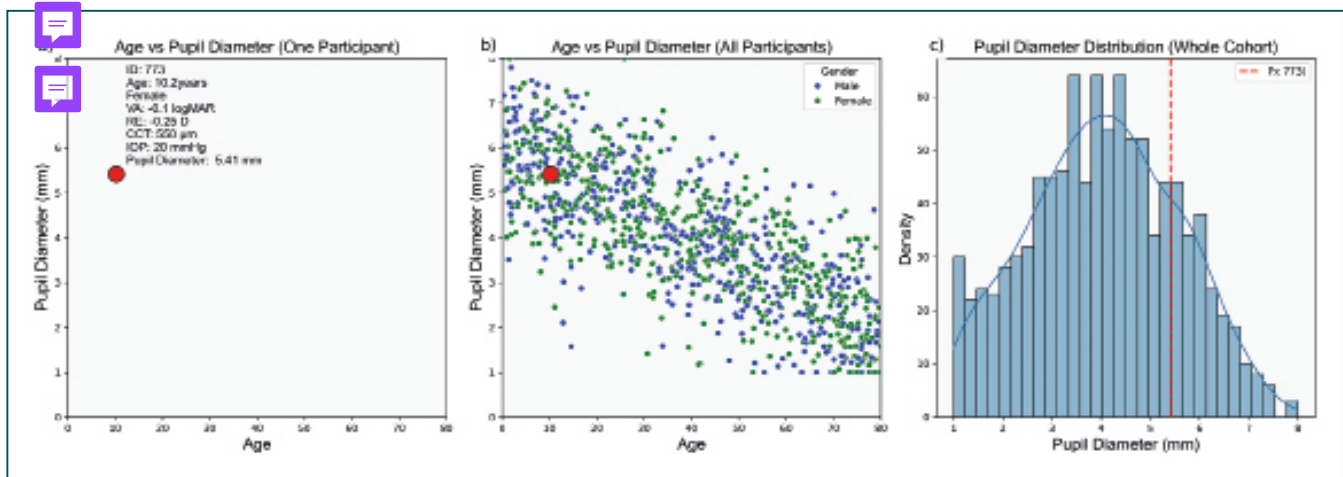


Bild 1: Visualisierung von Alter und Pupillendurchmesser: Einzelner Teilnehmer (a), gesamte Kohorte (b) und ein Histogramm der Pupillendurchmesser (c). Das Histogramm zeigt die Verteilung über 1.000 Teilnehmer, wobei die horizontale Achse den Pupillendurchmesser in Millimetern und die vertikale Achse die Teilnehmeranzahl darstellt. Jeder Balken entspricht der Anzahl an Teilnehmern pro Durchmesserintervall von 0,25 mm. Eine vertikale gestrichelte Linie zeigt die Position eines individuellen Teilnehmers, und eine dünne blaue Linie verdeutlicht die allgemeine Verteilungsform.

satz, der die Eigenschaften und die Variabilität der breiteren Bevölkerung widerspiegelt.

Biologische Messungen sind aufgrund individueller biologischer Unterschiede, Umwelteinflüsse und Messfehler von Natur aus mit Unsicherheit und Variabilität behaftet. Unsicherheit spiegelt das Vertrauen in die Messgenauigkeit und die Schlussfolgerungen wider, während Variabilität natürliche Datenschwankungen betrifft. Daher reicht die bloße Datenerfassung nicht aus. Statistik bietet die Werkzeuge, um Unsicherheit zu quantifizieren und Variabilität zu bewerten, sodass Forscher zwischen zufälligen Schwankungen und bedeutungsvollen Mustern unterscheiden können. Durch die Anwendung statistischer Techniken können Forscher fundierte Schlussfolgerungen über die Bevölkerung ziehen, die Zuverlässigkeit der Ergebnisse bewerten und die Wahrscheinlichkeit abschätzen, dass beobachtete Effekte zufällig sind.

Daten sind nicht monolithisch, sie variieren je nach Art der Messung und der verwendeten Messtechnik. Jeder Datentyp erfordert spezifische analytische Methoden. In der EVR sind Daten primär empirisch und stammen aus Beobachtung und Experimenten. Sie können als nominal (z. B. Art der Katarakt), ordinal (z. B. Sehschärfe) und metrisch (z. B. Achsenlänge) kategorisiert werden, **Tabelle 1**. Darüber hinaus können Daten durch die Werte definiert werden, die eine Variable annehmen kann; kontinuierliche Daten umfassen einen Bereich, der nur durch die Messgenauigkeit begrenzt wird, während diskrete Daten auf spezifische, separate Werte beschränkt sind, die typischerweise Zählungen oder ganze Zahlen darstellen.

Datenvorbereitung

Die Vorbereitung der Daten für die Analyse ist entscheidend für genaue Ergebnisse und umfasst die Schritte Exploration, Transformation und Validierung. Die Datenexploration ermög-

licht es dem Forscher zu verstehen, wie die Daten verwendet werden, und zu bestimmen, wie sie gereinigt, strukturiert und organisiert werden sollen. Die Transformation umfasst das Strukturieren der Daten, das relationale Organisieren und das Normalisieren durch Entfernen von Redundanzen. Die Reinigung adressiert Unregelmäßigkeiten, einschließlich fehlender Werte, Ungenauigkeiten und Ausreißer.³⁶

Die Implementierung von Validierungsregeln, wie Bereichs- und Konsistenzprüfungen während der Dateneingabe, reduziert Fehlerquoten. Regelmäßige Audits identifizieren Anomalien, und die Erkennung von Duplikaten verhindert die doppelte Zählung von Datenpunkten. Ein effektives Datenmanagement, einschließlich standardisierter Verfahren (SOPs) für die Datenerfassung, -eingabe und -speicherung, gewährleistet Konsistenz und Zuverlässigkeit. Die Schulung des Personals und die Nutzung von elektronischen Datenerfassungssystemen (EDC) mit Validierungswerkzeugen verbessern weiterhin die Datenqualität. Darüber hinaus überwachen statistische Qualitätskontrollmaßnahmen, wie Kontrollkarten und Prozessfähigkeitsanalysen, die Datenerfassung über die Zeit, erkennen Verschiebungen oder Trends, die auf Qualitätsprobleme hinweisen könnten, und gewährleisten die Einhaltung von Qualitätsstandards.^{36,37}

Fehlende Daten

Fehlende Daten sind ein häufiges Problem in der Forschung, das zu Verzerrungen führen kann, die statistische Aussagekraft reduziert und die Verallgemeinerbarkeit von Studienergebnissen beeinträchtigt. In der klinischen Forschung können fehlende Daten durch Patientenabbrüche, Nichteinhaltung von Studienprotokollen, technische Probleme bei der Datenerhebung oder Fehler bei der Dateneingabe entstehen. Die Behandlung fehlender Daten erfordert ein Verständnis der zugrunde liegenden Mechanismen und die Anwendung ge-

Tabelle 1: Klassifikation von Datentypen und deren Anwendungen in der ophthalmologischen Forschung

Typ/Skala	Charakteristik	Beispiel/Anwendung
Nominale Daten	Durch diskrete Kategorien oder Labels ohne inhärente Ordnung oder Rang gekennzeichnet, erfassen nominale Daten oft qualitative Unterscheidungen, die für die klinische Praxis entscheidend sind.	Augenfarbe (blau, braun, grün), Art des Katarakts (nuklear, kortikal, subkapsulär posterior), Vorhandensein/Fehlen einer spezifischen okularen Erkrankung. Statistischer Ansatz: Analyse konzentriert sich oft auf Zählungen, Häufigkeiten, Proportionen und Assoziationen zwischen nominalen Variablen und anderen Datentypen.
Ordinal Data	Schließt die Lücke zwischen qualitativen und quantitativen Informationen, indem Kategorien mit einer natürlichen Ordnung oder Rangfolge dargestellt werden.	Sehschärfewerte (1,0 0,5, 0,1), Schweregradskalen für Trockenes Auge (leicht, mittel, schwer), Bewertungsskalen, Likert-Skalen. Statistischer Ansatz: Median, Perzentile und nicht-parametrische Tests werden häufig für die Analyse ordinaler Daten verwendet.
Metric Data	Umfassen numerische Messungen mit bedeutungsvollen und quantifizierbaren Intervallen, die eine breite Palette ophthalmologischer Forschung untermauern. Weitere Klassifikation: Metrische Daten werden weiter unterteilt in: 1. Intervall-Daten: Besitzen eine konsistente Skala, aber keinen wahren Nullpunkt. 2. Verhältnis-Daten: Weisen eine konsistente Skala und einen wahren Nullpunkt auf, was das Vorhandensein des gemessenen Merkmals repräsentiert. Dies ermöglicht sinnvolle Verhältnisse und eine breitere Palette statistischer Analysen.	Intervall-Daten: Temperatur in Celsius oder Fahrenheit, einige psychometrische Skalen, die in der Sehforschung verwendet werden, wie Dezibel bei Sehfeldtests. Die Sehfähigkeit in LogMAR. Verhältnis-Daten: Intraokularer Druck, Achsenlänge, Hornhautdicke. Statistischer Ansatz: Metrische Daten öffnen die Tür zu einer breiten Palette statistischer Techniken, einschließlich Mittelwerte, Standardabweichungen, Korrelationen und parametrische Tests.

eigneter statistischer Methoden.^{38,39} Fehlende Daten lassen sich in drei Kategorien einteilen: Zufällig fehlende Daten (Missing Completely At Random, MCAR), Nicht-zufällig fehlende Daten (Missing At Random, MAR) und Nicht-zufällig und nicht nach Zufall fehlende Daten (Missing Not At Random, MNAR)⁴⁰ (Tabelle 2). Tabelle 2 verdeutlicht, wie entscheidend das Verständnis der Art des Datenfehlers für die Auswahl der richtigen statistischen Methoden ist, um Verzerrungen zu vermeiden und die Integrität der Studienergebnisse zu wahren.

Ausreißer

Ausreißer sind Datenpunkte, die deutlich von der Gesamtheit der Daten abweichen und können durch Messfehler, Eingabefehler oder tatsächlich extreme Werte aufgrund biologischer Variabilität oder seltener Bedingungen entstehen. Das Erkennen und angemessene Verwalten von Ausreißern ist entscheidend, um die Integrität der Daten und die Gültigkeit der statistischen Analyse zu gewährleisten.^{45,46} Tabelle 3 fasst verschiedene Methoden zur Erkennung von Ausreißern zusammen und beschreibt deren Anwendungen und Einschränkungen.

Nachdem Ausreißer erkannt wurden, sollten sie sorgfältig untersucht und behandelt werden. Der erste Schritt ist die Überprüfung, die Überprüfungen auf Dateneingabefehler oder Messungenauigkeiten umfasst, wie Tippfehler oder Kalibrierungsprobleme bei Instrumenten. Es ist auch wichtig zu bestimmen, ob der Ausreißer klinisch plausibel ist oder eine extreme physiologische Bedingung darstellt.^{45,46}

Forscher haben mehrere Optionen, um mit Ausreißern umzugehen. Ihre Beibehaltung ist angemessen, wenn sie gültige Beobachtungen darstellen, obwohl dies die Variabilität erhöhen und die statistische Robustheit beeinträchtigen kann. Im Gegensatz dazu ist das Ausschließen von Ausreißern gerechtfertigt, wenn es sich um Fehler handelt oder sie die untersuchte Population nicht repräsentieren, wobei vordefinierte Ausschlusskriterien zur Vermeidung von Verzerrungen angewendet werden sollten. Alternativ können Datentransformationen (z. B. logarithmische) den Einfluss von Ausreißern ohne deren Entfernung reduzieren, und robuste Methoden wie die Methode der kleinsten absoluten Abweichungen (Least Absolute Deviations, LAD) oder M-Schätzer minimieren den Einfluss von Extremwerten auf die Analyse.⁴⁸

Ausreißer können erhebliche Auswirkungen auf die klinische Forschung und Praxis haben. Manchmal können sie neuartige Befunde darstellen, wie seltene klinische Prä-

Tabelle 2: Arten von fehlenden Daten und ihre Auswirkungen in der klinischen Forschung⁴⁰

Typ	Definition	Gleichung	Implikationen	Klinisches Beispiel
Missing Completely at Random (MCAR) – Vollständig zufällig fehlend	Die Wahrscheinlichkeit des Fehlens ist unabhängig von beobachteten und unbeobachteten Daten.	$P(M = 1 X, Y) = P(M = 1)$	Daten sind eine zufällige Teilmenge der Originaldaten. Analysen bleiben unverzerrt, wenn nur vollständige Fälle verwendet werden. Diese Annahme ermöglicht unverzerrte statistische Schätzungen, obwohl sie in realen Daten oft schwer zu erfüllen ist.	Fehlende Messungen der Sehschärfe aufgrund eines zufälligen Geräteausfalls, der alle Patienten gleichmäßig betrifft.
Missing at Random (MAR) – Zufällig fehlend	Die Wahrscheinlichkeit des Fehlens hängt nur von den beobachteten Daten ab, nicht von den unbeobachteten Daten.	$P(M = 1 X, Y) = P(M = 1 X)$	Das Fehlen kann durch Bedingungen auf beobachtete Variablen berücksichtigt werden. Diese Annahme ist weniger restriktiv als MCAR und oft realitätsnäher, was Forschern erlaubt, gültige Schlüsse zu ziehen, ohne durch das Muster der fehlenden Daten verzerrt zu werden.	Patienten mit schwerer diabetischer Retinopathie kommen weniger wahrscheinlich zu Nachuntersuchungen, und ihre Schwere wurde bei vorherigen Besuchen dokumentiert. ^{41,42}
Missing Not at Random (MNAR) – Nicht zufällig fehlend	Die Wahrscheinlichkeit des Fehlens hängt von den unbeobachteten Daten selbst ab (schwierigstes Szenario).	$P(M = 1 X, Y) = P(M = 1 Y)$	Das Fehlen ist mit den fehlenden Werten selbst verbunden. In MNAR-Situationen sind Standardmethoden wie multiple Imputation oder Maximum-Likelihood oft allein unzureichend, da sie von der Annahme abhängen, dass das Fehlen von beobachteten Daten abhängt. Oft sind zusätzliche Modellierungen oder Sensitivitätsanalysen erforderlich.	Patienten mit schweren Nebenwirkungen (nicht aufgezeichnet, weil sie aus der Studie ausgeschieden sind) brechen eine klinische Studie wahrscheinlicher ab. ^{43,44}

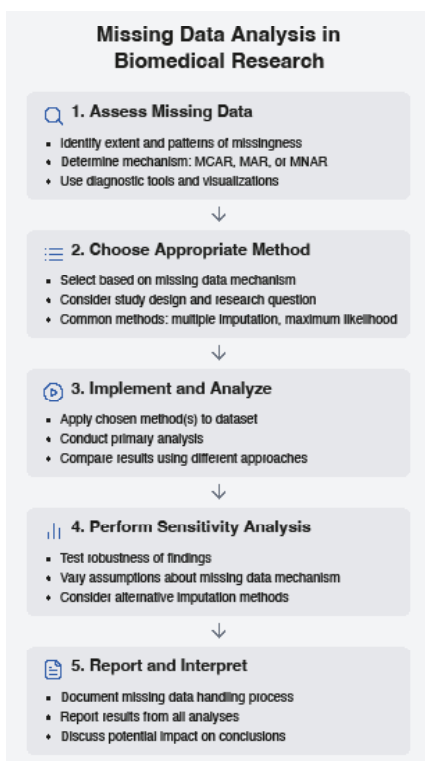


Bild 2: Schrittweiser Ansatz zum Umgang mit fehlenden Daten in der biomedizinischen Forschung (© Daniela Oehring 2024)

sentationen, die weitere Untersuchungen rechtfertigen. Beispielsweise könnte ein ungewöhnlich früher Beginn der altersbedingten Makuladegeneration (AMD) Einblicke in eine atypische Krankheitsprogression oder einzigartige Risikofaktoren bieten. Ausreißer können jedoch auch statistische Analysen verzerren und zu fehlerhaften Schlussfolgerungen führen. Extreme Werte können Parameterschätzungen verzerren und die Gültigkeit der Studie beeinträchtigen. Dies ist besonders relevant bei der Risikostratifizierung, wo das Identifizieren von Ausreißern dazu beitragen kann, Hochrisikopatienten zu erkennen, die möglicherweise spezielle Interventionen benötigen. In klinischen Einstellungen unterstreichen Ausreißer oft wichtige Überlegungen. Zum Beispiel könnte eine Untergruppe von Glaukompatienten mit ungewöhnlich schneller Progression auf eine Nichtbefolgung der Behandlung oder eine aggressivere Krankheitsvariante hinweisen, die eine engere Überwachung erfordert.⁴⁹⁻⁵² Ebenso könnte eine ungewöhnlich lange Achsenlänge bei biometrischen Messungen auf pathologische Myopie hinweisen und weitere Untersuchungen oder Anpassungen der Behandlung erfordern.^{53,54}

Tabelle 3: Gängige Methoden zur Erkennung von Ausreißern⁴⁷

Methode	Charakteristik
<p>Z-Scores oder Standardwerte</p>	<p>Z-Scores messen, wie viele Standardabweichungen ein Datenpunkt vom Mittelwert entfernt ist. Der Z-Score wird berechnet als: $z = (x - \mu) / \sigma$ wobei z der Z-Score des Datenpunktes ist, x der Wert des Datenpunktes, μ der Mittelwert des Datensatzes und σ die Standardabweichung des Datensatzes</p> <p>Interpretation von Z-Scores</p> <ul style="list-style-type: none"> • Ein Z-Score von 0 bedeutet, dass der Datenpunkt genau beim Mittelwert liegt. • Positive Z-Scores zeigen an, dass der Datenpunkt über dem Mittelwert liegt. • Negative Z-Scores zeigen an, dass der Datenpunkt unter dem Mittelwert liegt. • Z-Scores nahe +1 oder -1 bedeuten, dass der Datenpunkt innerhalb einer Standardabweichung vom Mittelwert liegt, während Werte über ± 2 oder ± 3 Standardabweichungen typischerweise als Ausreißer betrachtet werden, je nach Kontext. <p>Diese Methode setzt voraus, dass die Daten einer Normalverteilung folgen, was ihre Anwendbarkeit auf normal verteilte Datensätze einschränkt.</p>
<p>Modifizierter Z-Score</p>	<p>Die mittlere absolute Abweichung (Median Absolute Deviation, MAD) ist eine robustere Alternative für nicht normal verteilte Daten. Der modifizierte Z-Score wird berechnet als:</p> $z_{\text{modified}} = [0,6745 (x_i - \text{median})] / \text{MAD}$ <p>z_{modified} ist der modifizierte Z-Score des Datenpunkts. x_i ist der Wert des Datenpunkts. Median ist der Median des Datensatzes. MAD wird als der Median der absoluten Abweichungen vom Median berechnet: $\text{MAD} = \text{Median}(x_i - \text{Median})$</p> <p>Die Konstante 0,6745 dient als Skalierungsfaktor, der die MAD an die Standardabweichung für große Normalverteilungen anpasst, um sicherzustellen, dass modifizierte Z-Scores den traditionellen Z-Scores in gutartigen Daten näherkommen. Diese Methode ist weniger empfindlich gegenüber extremen Werten und besser geeignet für schiefe Verteilungen.</p> <p>Interpretation von modifizierten Z-Scores</p> <ul style="list-style-type: none"> • Ein modifizierter Z-Score von 0 deutet auf einen Wert hin, der dem Median entspricht. • Höhere absolute modifizierte Z-Scores zeigen Werte an, die weiter vom Median entfernt sind. • Allgemein wird ein modifizierter Z-Score von mehr als 3.5 oft als Schwelle zur Identifizierung potenzieller Ausreißer verwendet. <p>Modifizierte Z-Scores sind besonders nützlich, wenn die Daten Ausreißer enthalten könnten oder die Annahme der Normalität nicht erfüllen. Dieser Ansatz bietet ein stabileres Maß für Variabilität, um ungewöhnliche Beobachtungen in nicht-normalen Verteilungen zu identifizieren.</p>
<p>Boxplot</p>	<p>Boxplots visualisieren Ausreißer als Punkte, die das 1,5-fache des Interquartilsabstands (IQR) von den Quartilen überschreiten. Boxplots sind besonders nützlich in der explorativen Datenanalyse und bieten eine einfache und visuelle Methode zur Erkennung von Ausreißern. Sie haben jedoch ihre Grenzen: Sie basieren auf der 1,5× IQR-Regel, welche Punkte in schiefen oder kleinen Datensätzen fälschlicherweise als Ausreißer kennzeichnen kann und die Schwere der Ausreißer nicht unterscheidet. Dieser Ansatz setzt Symmetrie voraus und könnte multivariate Ausreißer übersehen oder Nuancen in der Verteilungsform ignorieren. Obwohl Boxplots für schnelle explorative Analysen wertvoll sind, sind oft zusätzliche Methoden notwendig, um wahre Ausreißer zuverlässig zu identifizieren, besonders in komplexen Datensätzen.</p>

Fortsetzung nächste Seite

(Fortsetzung) **Tabelle 3:** Gängige Methoden zur Erkennung von Ausreißern⁴⁷

Methode	Charakteristik
Grubbs' test	<p>Der Grubbs-Test kann verwendet werden, um einen einzelnen Ausreißer in einem univariaten Datensatz zu erkennen. Die Teststatistik ist definiert als:</p> $G = (x_i - \bar{x}) / s$ <p>wobei x_i der Wert des vermuteten Ausreißers ist, \bar{x} der Stichprobenmittelwert und s die Stichprobenstandardabweichung.</p> <p>Der Test vergleicht den G-Wert mit einem kritischen Wert aus der Grubbs-Verteilungstabelle bei einem gewählten Signifikanzniveau (häufig 0,05). Wenn G diesen kritischen Wert übersteigt, wird der verdächtige Punkt als Ausreißer betrachtet. Der Grubbs-Test setzt Normalverteilung voraus, was ihn weniger effektiv für nicht normal verteilte Daten macht. Bei Datensätzen mit mehreren Ausreißern ist eine wiederholte Anwendung des Tests erforderlich, was die Genauigkeit reduzieren kann.</p>
Mahalanobis distance	<p>Die Mahalanobis-Distanz ist ein Maß, das verwendet wird, um die Distanz eines Datenpunkts vom Mittelwert einer multivariaten Verteilung zu bestimmen, wobei die Korrelationen zwischen den Variablen berücksichtigt werden. Im Gegensatz zur euklidischen Distanz berücksichtigt die Mahalanobis-Distanz die Kovarianzstruktur der Daten, was sie besonders effektiv für die Identifizierung multivariater Ausreißer in Datensätzen mit korrelierten Variablen macht.</p> <p>Die Mahalanobis-Distanz wird berechnet als $D^2 = (x - \mu)^T \Sigma^{-1} (x - \mu)$ wobei x der Vektor der Datenpunktwerte ist, μ der Mittelwertsvektor des Datensatzes und Σ die Kovarianzmatrix des Datensatzes. $(x - \mu)^T$ ist die Transposition des Abweichungsvektors.</p>

Deskriptive Statistik

Deskriptive Statistiken sind grundlegende Werkzeuge, die Daten zusammenfassen und organisieren, ohne Schlussfolgerungen oder Vorhersagen zu treffen. Sie bieten einen Überblick über die Schlüsselmerkmale eines Datensatzes. In EVR, wo die Daten von biometrischen Messungen bis hin zu Patientenberichten reichen, helfen deskriptive Statistiken, komplexe Informationen in klare, interpretierbare Ergebnisse zu überführen.

Maße der zentralen Tendenz

Maße der zentralen Tendenz sind statistische Kennzahlen, die dazu dienen, den zentralen oder typischen Wert innerhalb einer Datenmenge zu identifizieren. Diese Maße sind entscheidend, um die allgemeine Tendenz einer Verteilung zu verstehen, ohne dass alle Einzelwerte berücksichtigt werden müssen. In der Datenanalyse sind die drei gängigsten Maße der zentralen Tendenz der Mittelwert, der Median und der Modus, **Tabelle 4** und **Bild 3**.

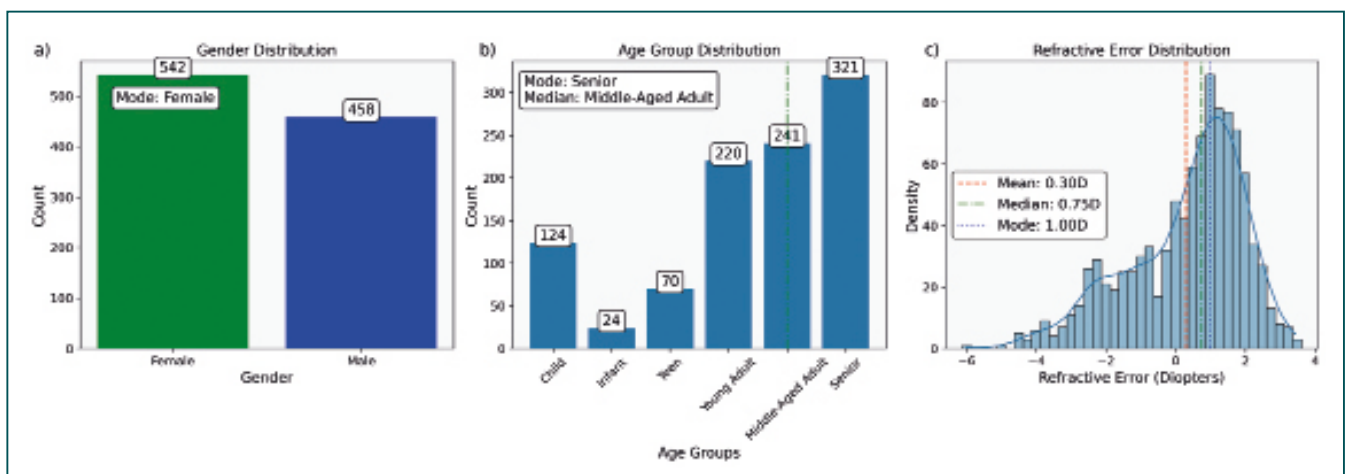


Bild 3: Visuelle Darstellung simulierter Gesundheitsdaten (n = 1000), die die drei zentrale Tendenzmaße (Mittelwert, Median und Modus) vorstellen. Die Diagramme zeigen (a) die Verteilung nach Geschlecht (nominal), (b) Altersgruppen (ordinal) und (c) den Brechungsfehler (metrisch) innerhalb der Kohorte. Das Diagramm zu Brechungsfehlern zeigt die Verteilung im Sample, wobei die Balkenhöhen die Anzahl der Personen mit spezifischen Brechungsfehlern angeben. Der Mittelwert wird durch eine rote gestrichelte Linie dargestellt (ungefähr 0,3 D), der Median bei 0,75 D und der Modus bei 1,0 D).

Tabelle 4: Maße der zentralen Tendenz und ihre Anwendung in der ophthalmologischen Forschung⁵⁵

Maß	Beschreibung	Anwendung
Mittelwert (Arithmetischer Durchschnitt) Anwendbar für metrische Daten mit oder ohne natürlichen Nullpunkt.	Der Mittelwert wird berechnet, indem alle Datenpunkte summiert und durch die Anzahl der Beobachtungen geteilt werden. Er ist empfindlich gegenüber Extremwerten (Ausreißern), die den Mittelwert bei kleinen Stichprobengrößen erheblich beeinflussen können.	Sehschärfe: Bei der Aggregation von Sehschärfemessungen, insbesondere unter Verwendung der logarithmischen Minimalwinkelauflösung (logMAR), bietet der Mittelwert einen präzisen Durchschnitt, der die logarithmische Natur der Sehschärfedaten berücksichtigt. ⁵⁶ Zentrale Hornhautdicke (CCT): Mittlere CCT-Werte sind wichtig zur Bewertung von Risiken für Krankheiten wie Hornhautektasie und Glaukom, bei denen Abweichungen vom Mittelwert klinisch bedeutend sein können. ⁵⁷
Median Anwendbar für metrische Daten mit oder ohne natürlichen Nullpunkt sowie für ordinale Daten.	Der Median ist der mittlere Wert, wenn die Daten der Größe nach geordnet sind. Er ist robust gegen Ausreißer und schiefe Datenverteilungen, was ihn besonders wertvoll in der ophthalmologischen Forschung macht, wo Daten möglicherweise nicht normal verteilt sind.	Intraokularer Druck (IOP): IOP-Messungen zeigen oft eine Schiefe aufgrund der Anwesenheit von Augen mit erhöhtem Druck bei Glaukom. Der Median bietet in solchen Fällen einen repräsentativeren zentralen Wert. ⁵⁸
Mode Anwendbar für alle Datenskalen: metrisch, ordinal, nominal.	Der Modus repräsentiert den am häufigsten vorkommenden Wert in einem Datensatz. Er ist besonders nützlich für kategoriale oder diskrete Daten.	Klassifikation von Refraktionsfehlern: Die Identifizierung des Modus in einer Population kann den häufigsten Refraktionsfehler (z. B. Myopie) innerhalb einer spezifischen demografischen Gruppe hervorheben. ⁵⁹

In multizentrischen Studien oder Metaanalysen, bei denen Daten aus verschiedenen Quellen mit unterschiedlichen Stichprobengrößen stammen, kann der gewichtete Mittelwert diese Unterschiede zwischen den Studien berücksichtigen und so eine genauere Gesamtschätzung liefern. Wenn Daten log-normal verteilt sind oder Wachstumsraten beinhalten, wie zum Beispiel Bakterienzählungen⁶⁰ bei augenmedizinischen Infektionen, bietet der geometrische Mittelwert ein geeigneteres Maß für die zentrale Tendenz.

Maße der Streuung

Maße der Streuung quantifizieren die Variabilität innerhalb eines Datensatzes und ergänzen damit die Maße der zentralen Tendenz, indem sie Einblicke in die Homogenität der Daten geben. Homogenität bezieht sich auf den Grad der Ähnlichkeit oder Gleichförmigkeit zwischen den Datenpunkten.⁶¹ In klinischen Umgebungen ist das Verständnis der Populationsvariabilität entscheidend für die Interpretation individueller Messungen. Beispielsweise neigen Daten zum intraokularen Druck (IOP) von Personen im Alter von 0 bis 80 Jahren dazu, sich um einen altersspezifischen Mittelwert zu gruppieren. (Bild 4) Jedoch erfassen Maße der zentralen Tendenz allein nicht die Variabilität innerhalb der Kohorte. Maße der Streuung sind entscheidend, um den Bereich der Werte abzuschätzen, den ein Individuum aufweisen könnte, beeinflusst durch die Datenskala und die Art der Messungen.

Datenverteilung

Die Datenverteilung ist entscheidend für die Auswahl geeigneter statistischer Methoden und die genaue Interpretation der Ergebnisse. Sie gibt die Wahrscheinlichkeit spezifischer Messungen in einem klinischen Kontext an, leitet Erwartungen und unterstützt die klinische Entscheidungsfindung.⁶⁹ Zwei wichtige Verteilungsmerkmale sind Schiefe und Kurtosis. Die Schiefe misst die Asymmetrie um den Mittelwert und zeigt an, ob Daten stärker auf einer Seite konzentriert sind. Die Kurtosis quantifiziert die Schärfe des Peaks einer Verteilung, was ihre Neigung zur Erzeugung von Ausreißern widerspiegelt. Bild 4 veranschaulicht verschiedene Verteilungsszenarien, und Tabelle 5 fasst die Arten von Schiefe und Kurtosis mit klinischen Beispielen zusammen.

Tabelle 5: Maße der Streuung und ihre Anwendungen in der ophthalmologischen Forschung⁶¹

Maß	Beschreibung	Klinisches Beispiel
Spannweite	Anwendbar für metrische und ordinale Daten. Die Spannweite ist das einfachste Maß und wird als Differenz zwischen den maximalen und minimalen Werten berechnet.	Variationen der Achsenlänge: Die Berichterstattung über die Spannweite der Achsenlängen in einer myopischen Bevölkerung kann die potenzielle Variabilität in den chirurgischen Ergebnissen für Verfahren wie die Implantation von Collamer-Linsen aufzeigen. ^{62,63}
Varianz und Standardabweichung (SD)	<p>Anwendbar für normalverteilte, metrische Daten.</p> <p>Die Varianz repräsentiert die durchschnittliche quadratische Abweichung vom Mittelwert, während die SD die Quadratwurzel der Varianz ist, was die Streuung in denselben Einheiten wie die Daten bietet.</p> <p>Einzelne SD (1σ): Etwa 68 % der Daten liegen innerhalb einer Standardabweichung über und unter dem Mittelwert, etwa 34 % zu jeder Seite.</p> <p>Doppelte SD (2σ): Etwa 95 % der Daten liegen innerhalb von zwei Standardabweichungen über und unter dem Mittelwert, etwa 47,5 % zu jeder Seite.</p> <p>Dreifache SD (3σ): Etwa 99,7 % der Daten liegen innerhalb von drei Standardabweichungen über und unter dem Mittelwert. Dies entspricht etwa 49,85 % zu jeder Seite.</p>	Messungen der Netzhautdicke: Die SD ist entscheidend, um die Variabilität der Dicke der retinalen Nervenfaserschicht (RNFL) zu beurteilen, was die Früherkennung von Glaukomfortschritten unterstützt. ^{64,65}
Variationskoeffizient (CV)	Anwendbar für metrische und normalverteilte Daten. Der CV ist ein normiertes Maß der Streuung, berechnet als SD geteilt durch den Mittelwert, oft als Prozentsatz ausgedrückt.	Konsistenz chirurgischer Ergebnisse: In Studien, die chirurgische Techniken vergleichen, kann der CV aufzeigen, welche Methode konsistentere refraktive Ergebnisse nach der Operation liefert. ⁶⁶
Interquartilsbereich (IQR) Applicable for metrical, ordinal data	Anwendbar für metrische und ordinale Daten. Der IQR umfasst die mittleren 50 % der Daten, zwischen dem 25. (Q1) und dem 75. (Q3) Perzentil, und wird weniger von Ausreißern beeinflusst.	Patientenberichtete Ergebnismaße (PROMs): Der IQR ist nützlich, um PROMs wie Fragebögen zur visuellen Funktion zusammenzufassen, bei denen die Daten ordinal und schief verteilt sein können. ^{67,68}
Fortgeschrittene Überlegungen		
Bootstrapping für Konfidenzintervalle	Für kleine Stichprobengrößen oder nicht normal verteilte Daten können Bootstrapping-Methoden die Variabilität schätzen und Konfidenzintervalle für zentrale Tendenzmaße konstruieren.	

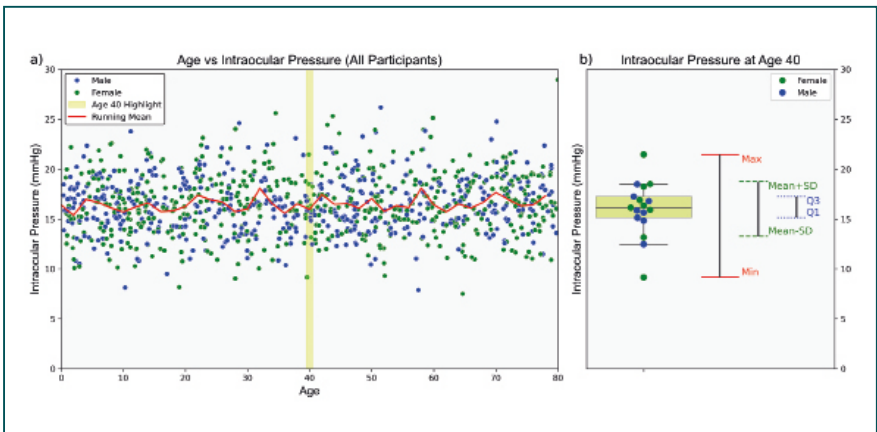


Bild 4: Die Beziehung zwischen Alter und intraokularem Druck (IOP) in einem generierten Datensatz zeigt die Variation des IOP über verschiedene Altersgruppen hinweg für männliche und weibliche Teilnehmer. (a) IOP-Messungen für alle Teilnehmer, wobei Personen im Alter von 40 Jahren hervorgehoben sind; die rote Linie repräsentiert den durchschnittlichen IOP pro Jahr. (b) Ein Boxplot, der den IOP im Alter von 40 Jahren detailliert darstellt und wichtige Streuungsmaße veranschaulicht, einschließlich Spannweite, Interquartilsabstand (IQR) und Standardabweichung (SD).

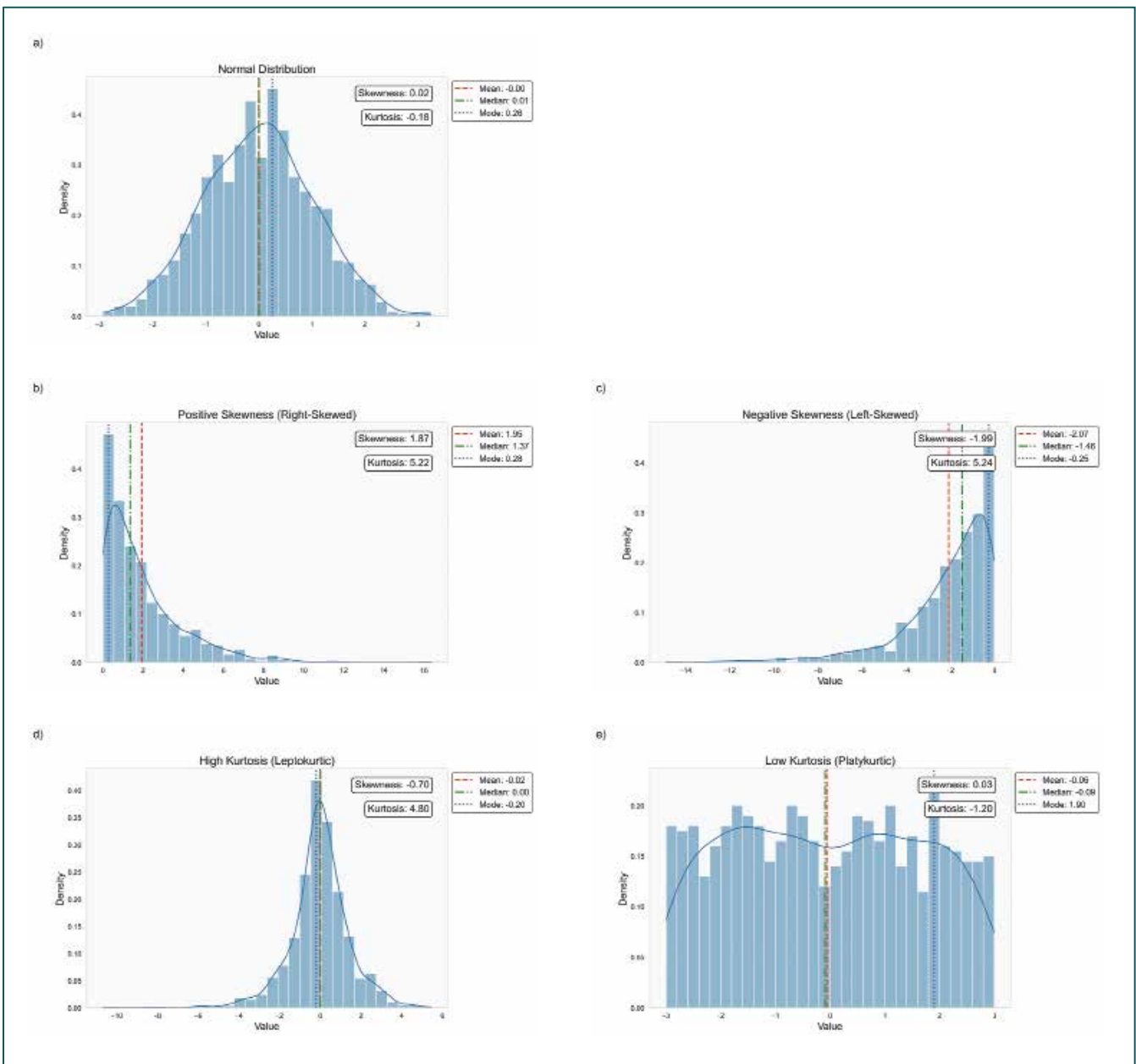


Bild 5: Abbildung verschiedener Verteilungen, die Variationen in der Schiefe und Kurtosis der Verteilung darstellen. (a) Normalverteilung mit einer Schiefe und Kurtosis nahe Null. (b und c) Rechtsschiefe (positive Schiefe) bzw. linksschiefe (negative Schiefe) Verteilungen. (d und e) Verteilungen mit niedriger Kurtosis und hoher Kurtosis.

Tabelle 6: Zusammenfassung der Schiefe und Kurtosis⁷⁰

Schiefe	Merkmal	Klinisches Szenario
Typ		
Schiefe = 0	Zeigt eine symmetrische Verteilung an, bei der die linken und rechten Enden ausgeglichen sind, Bild 5a .	Visusmessungen in einer großen, gesunden Bevölkerung können eine symmetrische Verteilung annähern, da die meisten Werte sich um einen zentralen Punkt gruppieren, mit weniger Extremen auf beiden Seiten. ⁷¹
Positive Schiefe (Rechtsschiefe Verteilung) Schiefe > 0	Der Schwanz der Verteilung erstreckt sich nach rechts, was darauf hindeutet, dass höhere Werte weniger häufig sind, Bild 5b .	IOP-Messungen zeigen oft eine rechtsschiefe Verteilung, da die meisten Patienten normale Druckwerte haben, während einige erhöhte Werte aufweisen. Das Erkennen dieser Tatsache hilft Klinikern, diejenigen mit hohem IOP hinsichtlich potenzieller Risikofaktoren zu überwachen. ⁷²
Negative Schiefe (Linksschiefe Verteilung) Schiefe < 0	Der Schwanz erstreckt sich nach links, was darauf hindeutet, dass niedrigere Werte weniger häufig sind, Bild 5c .	Einige Visuelfeldindizes, bei denen die meisten Patienten hohe Werte erreichen, zeigen eine links-schiefe Verteilung. Seltene niedrige Werte können auf eine zugrunde liegende Pathologie hinweisen und Kliniker anleiten, potenzielle Probleme bei betroffenen Patienten zu untersuchen. ⁷³
Kurtosis		
Mesokurtisch Kurtosis = 3 (oder Exzess-Kurtosis = 0)	Diese Verteilung hat dieselbe Kurtosis wie eine Normalverteilung. Siehe Bild 5a .	—
Leptokurtisch (Hohe Kurtosis) Kurtosis > 3 (oder Exzess-Kurtosis > 0)	Eine Verteilung mit scharfem Gipfel und schweren Enden, was darauf hindeutet, dass Daten eng um den Mittelwert gruppiert sind, jedoch mit einer höheren Wahrscheinlichkeit für extreme Werte. Siehe Bild 5d .	Messungen der Endothelzellzahlen in der Hornhaut können leptokurtische Verteilungen zeigen, wobei die meisten Werte nahe am Mittelwert liegen, aber gelegentlich signifikante Abweichungen auftreten. Hohe Kurtosis alarmiert Forscher hinsichtlich des potenziellen Einflusses von Ausreißern auf statistische Analysen und klinische Interpretationen. ^{74,75}
Platykurtisch (Niedrige Kurtosis) Kurtosis < 3 (oder Exzess-Kurtosis < 0)	Eine Verteilung mit flacherem Gipfel und dünneren Enden, was auf eine gleichmäßigere Datenverteilung mit weniger extremen Werten hindeutet. Siehe Bild 5e .	Eine platykurtische Verteilung könnte in Fällen wie Brechungsfehlermessungen in einer diversen Bevölkerung auf eine breite Palette von gemeinsamen Werten mit weniger Neigung zu Ausreißern hinweisen. ^{76,77}

Schiefe kann basierend auf Schwellenwerten als annähernd symmetrisch (Schiefe zwischen -0,5 und 0,5), mäßig schief (zwischen -1 und -0,5 oder zwischen 0,5 und 1) und stark schief (unter -1 oder über 1) kategorisiert werden. Diese Schwellenwerte können je nach Fachgebiet variieren, daher ist es wichtig, die Schiefe im Kontext der Daten und Analyseziele zu betrachten. Schiefe Daten können die Annahmen parametrischer statistischer Tests verletzen, die in der Regel von normalen (symmetrischen) Verteilungen ausgehen. Datenumwandlungen (z. B. logarithmisch) oder nicht-parametrische statistische Methoden können für gültige Ergebnisse notwendig sein. Die Kurtosis hilft, das Risiko von Ausreißern zu bewerten; beispielsweise kann eine leptokurtische Verteilung in klinischen Studien eine zusätzliche Überprüfung

extremer Werte erforderlich machen, um sicherzustellen, dass es sich um echte Beobachtungen handelt und nicht um Fehler. Sie informiert auch über statistische Entscheidungen, da einige Tests robuster gegenüber Abweichungen der Kurtosis sind.

Normalverteilung

Das Verständnis der Datenverteilung ist in der Statistik entscheidend, da es die Auswahl statistischer Tests leitet und die Gültigkeit der aus den Daten gezogenen Schlussfolgerungen bestimmt.⁷⁰ Die Normal- oder Gauß-Verteilung steht besonders im Zentrum der statistischen Theorie und Praxis,

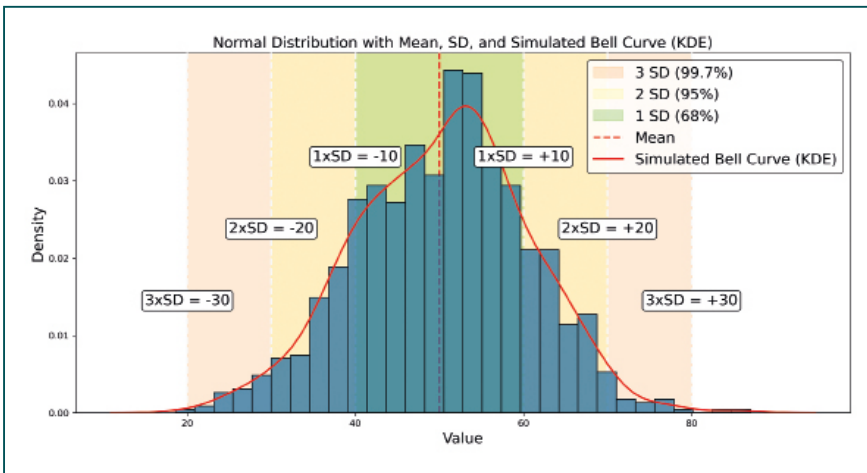


Bild 6: Normalverteilung mit gekennzeichneten Standardabweichungen (1σ , 2σ , 3σ), die den Anteil der Daten innerhalb jedes Bereichs veranschaulicht. Die simulierte Glockenkurve (rot) demonstriert das Konzept der Normalität in der Verteilung klinischer Daten.

insbesondere in der Augen- und Sehforschung (EVR). Diese kontinuierliche Verteilung ist durch eine symmetrische, glockenförmige Kurve gekennzeichnet, die um den Mittelwert mit einer definierten Standardabweichung zentriert ist. (Bild 6)

Die Normalverteilung gilt für kontinuierliche Intervall- oder Verhältnisdaten. In der Augen- und Sehforschung (EVR) sind viele biometrische Messungen, wie die Hornhautdicke, die axiale Länge und die Sehschärfe in logMAR-Einheiten, Verhältnisskalen-kontinuierliche Daten, die unter der Voraussetzung, dass sie Kriterien für Symmetrie und das Fehlen signifikanter Kurtosis erfüllen, mit der Normalverteilung modelliert werden können.⁷⁸ Die Normalverteilung ist grundlegend in der statistischen Analyse und untermauert mehrere fundamentale Methoden, insbesondere parametrische Tests wie T-Tests, Varianzanalyse (ANOVA) und lineare Regression, die normalverteilte Daten oder Residuen voraussetzen. Eine Verletzung dieser Annahme kann zu ungenauen Ergebnissen führen.

Ein verwandtes und kritisches Konzept ist der Zentrale Grenzwertsatz (Central Limit Theorem, CLT), der besagt, dass sich mit zunehmender Stichprobengröße die Verteilung der Stichprobenmittelwerte unabhängig von der ursprünglichen Verteilung der Population der Normalität annähert. Dieses Prinzip ermöglicht es Forschern, normalbasierte Schlussfolgerungen bei großen Stichproben anzuwenden, auch wenn die zugrunde liegenden Daten nicht normal sind.⁷⁹

Die Normalverteilung ist auch für die statistische Inferenz essenziell, insbesondere bei der Berechnung von Konfidenz- und Prognoseintervallen. Diese Intervalle ermöglichen eine genaue Schätzung der Populationsparameter und die Vorhersage zukünftiger Beobachtungen. Sie stützen sich

auf Eigenschaften der Normalverteilung, die zuverlässige Schlussfolgerungen gewährleisten.

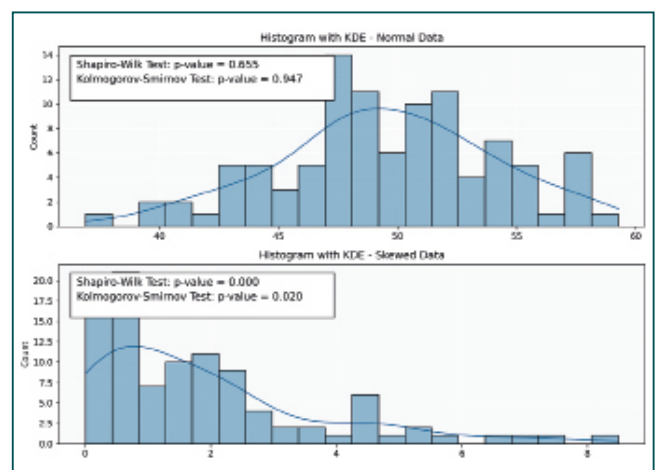
Abschließend erleichtert die Normalverteilung die Standardisierung, bei der Z-Scores (Tabelle 3) den Abstand eines Datenpunktes vom Mittelwert in Einheiten der Standardabweichung messen. Diese Standardisierung ermöglicht Vergleiche über verschiedene Skalen oder Verteilungen hinweg.

Normalverteilung beurteilen

Verschiedene grafische und statistische Methoden können verwendet werden, um die Normalität zu beurteilen, wobei jede Methode spezifische Stärken und Einschränkungen aufweist. Die Kombination grafischer und statistischer Ansätze (Tabelle 5) ermöglicht eine robustere Beurteilung der Daten-Normalität.⁷⁰

Histogramme: Histogramme bieten eine direkte Darstellung der Häufigkeitsverteilungen der Daten. Eine asymmetrische, glockenförmige Kurve deutet im Allgemeinen auf Normalität hin; jedoch kann die Interpretation von Histogrammen subjektiv sein, insbesondere bei kleineren Stichproben, bei denen die Form der Verteilung unklar sein kann. Zusätzlich können Histogramme subtile Abweichungen von der Normalität nicht aufzeigen.

Bild 7: Histogramme zur Beurteilung der Normalität: Histogramme von normal verteilten und schiefen Daten, begleitet von Kerndichteschätzungen (KDE) und p-Werten aus den Shapiro-Wilk- und Kolmogorov-Smirnov-Tests. Diese heben die visuelle und statistische Beurteilung der Normalität in klinischen Daten hervor.



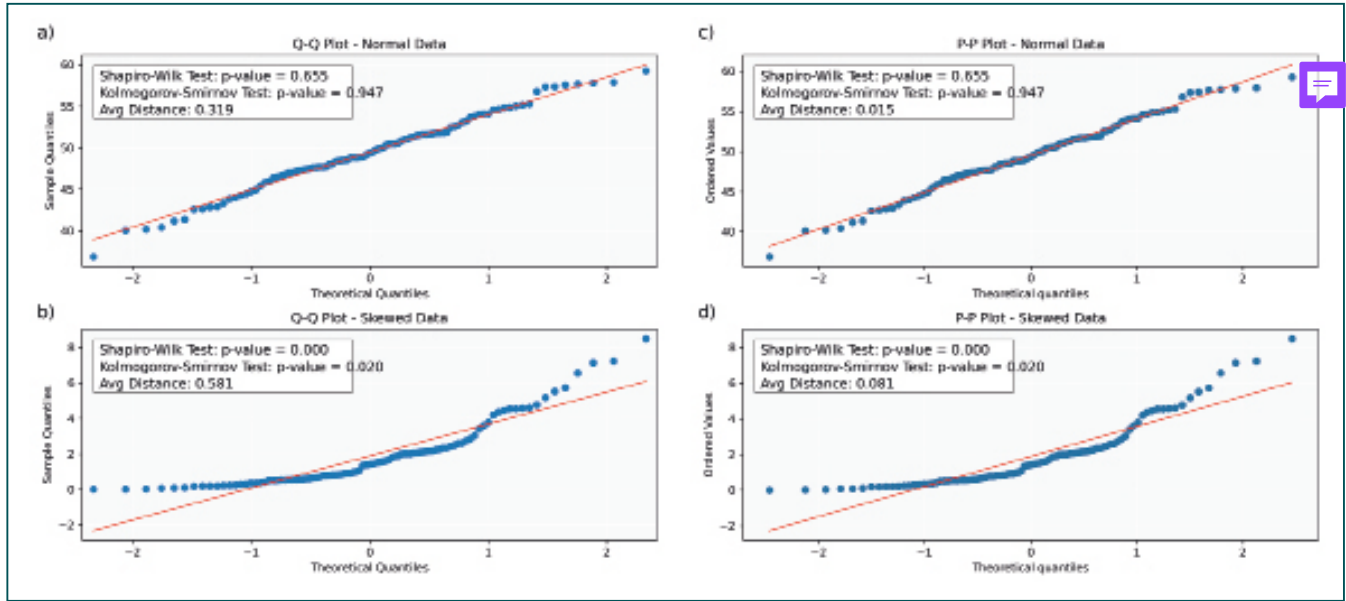


Bild 8: (a und b) Quantil-Quantil-Diagramme für normale und schiefe Daten, (c und d) Wahrscheinlichkeits-Wahrscheinlichkeits-Diagramme für normale und schiefe Daten

Q-Q-Diagramme (Quantil-Quantil-Diagramme): Q-Q-Diagramme vergleichen die Quantile der Stichprobendaten mit denen einer theoretischen Normalverteilung und bieten Einblicke in die Datenverteilung sowie die Identifikation von Ausreißern, Schiefe und Kurtosis. Die x-Achse repräsentiert theoretische Quantile einer Referenzverteilung (z. B. Normalverteilung), während die y-Achse die Quantile der Stichprobendaten zeigt. Jeder Punkt entspricht einem Quantil des Datensatzes, das gegen ein Quantil der Referenzverteilung geplottet ist. Zur Beurteilung der Normalität wird der (euklidische) Abstand zwischen jedem Datenpunkt und dem entsprechenden Punkt auf der Diagonalen (der den idealen Fall einer Normalverteilung darstellt) berechnet in **Bild 8a** und **8b**.

Typische Szenarien für Q-Q-Diagramme umfassen:

- Gerade Linie: Wenn die Punkte eng entlang der Diagonalen ausgerichtet sind, folgen die Stichprobendaten wahrscheinlich einer Normalverteilung.
- Aufwärts- oder Abwärtskrümmung: Deutet auf Rechtsschiefe (aufwärts) beziehungsweise Linksschiefe (abwärts) hin.
- S-förmige Kurve: Weist auf schwere Ausläufer in den Daten hin, was auf extremere Werte hindeutet als in einer Normalverteilung erwartet.
- Ausreißer: Punkte weit von der Linie entfernt repräsentieren Ausreißer oder signifikante Abweichungen von der erwarteten Verteilung.

P-P-Diagramme (Wahrscheinlichkeits-Wahrscheinlichkeits-Diagramme): P-P-Diagramme zeigen die kumulativen Wahrscheinlichkeiten der Stichprobendaten (y-Achse) gegen die einer Normalverteilung (x-Achse). Jeder Datenpunkt in einem P-P-Diagramm stellt die Wahrscheinlichkeit eines Stichprobendatenpunkts dar, geplottet gegen seine entsprechende Wahrscheinlichkeit aus einer theoretischen

Verteilung. Abweichungen von der Linie deuten auf Verletzungen der Normalität hin, ähnlich wie bei Q-Q-Diagrammen beschrieben. Während Q-Q-Diagramme Informationen über die Form der Verteilung liefern, bewerten P-P-Diagramme, wie gut die Gesamtwahrscheinlichkeiten zwischen der theoretischen und der Stichprobenverteilung übereinstimmen. Darüber hinaus sind P-P-Diagramme weniger empfindlich gegenüber Abweichungen in den Ausläufern der Verteilung im Vergleich zu Q-Q-Diagrammen.

Statistische Tests: Während statistische Tests (**Tabelle 7**) wertvolle Einblicke in die Normalität der Daten bieten, haben sie auch Einschränkungen. Eine primäre Herausforderung ist ihre Empfindlichkeit gegenüber der Stichprobengröße; bei kleinen Stichproben können diese Tests möglicherweise keine Nicht-Normalität erkennen (geringe Leistung), während sie bei großen Stichproben triviale Abweichungen als signifikant identifizieren können (Überempfindlichkeit). Daher kann die alleinige Verlass auf statistische Tests irreführend sein, insbesondere wenn sie mit anderen Tests kombiniert werden, die Normalität voraussetzen. Dies zeigt, wie wichtig es ist, grafische und statistische Methoden einzusetzen, um die Datenverteilung umfassend zu verstehen.

In der Praxis führen Abweichungen von der Normalität nicht unbedingt zur Ungültigkeit statistischer Methoden, die Normalität voraussetzen, insbesondere wenn die Stichprobengrößen groß genug sind, dass der Zentrale Grenzwertsatz Anwendung findet. Jedoch können nicht-parametrische Alternativen angemessener sein, wenn signifikante Abweichungen von der Normalität in kleinen bis mittleren Stichproben festgestellt werden.

Klinische Bedeutung und Aussagekraft der Normalität

Die Annahme der Normalität hat erhebliche Auswirkungen auf die klinische Forschung und Praxis und beeinflusst direkt

Table 7: Übersicht über die verschiedenen Tests zur Überprüfung der Normalität in klinischen Daten, ihre Beschreibungen und die damit verbundenen Einschränkungen⁷⁰

Test	Beschreibung	Einschränkungen
Shapiro-Wilk-Test	Der Shapiro-Wilk-Test prüft die Nullhypothese, dass die Daten normal verteilt sind. Er ist besonders effektiv für kleine bis moderate Stichprobengrößen (typischerweise < 50 Proben) und gilt als einer der leistungsfähigsten Tests zur Erkennung von Abweichungen von der Normalverteilung. Ein p-Wert kleiner als das gewählte Signifikanzniveau (z. B. 0,05) zeigt an, dass die Daten signifikant von einer Normalverteilung abweichen. ⁸⁰	Der Shapiro-Wilk-Test kann in großen Datensätzen überempfindlich gegenüber auch geringfügigen Abweichungen von der Normalität sein, was zu falsch positiven Ergebnissen führt, bei denen geringfügige Abweichungen als signifikant markiert werden.
Kolmogorov-Smirnov-Test	Der Kolmogorov-Smirnov-Test (K-S-Test) vergleicht die empirische Verteilungsfunktion der Stichproben mit der kumulativen Verteilungsfunktion einer spezifizierten Verteilung, wie der Normalverteilung. Er ist geeigneter für größere Stichproben.	Wie der Shapiro-Wilk-Test kann der K-S-Test in großen Stichproben zu empfindlich sein und Abweichungen von der Normalität erkennen, die möglicherweise die Ergebnisse parametrischer Tests nicht signifikant beeinflussen. Zusätzlich kann dieser Test bei kleineren Stichproben eine geringe Teststärke aufweisen und vorhandene Nicht-Normalität nicht erkennen.
Anderson-Darling-Test	Dieser Test ist eine Erweiterung des K-S-Tests, legt jedoch mehr Gewicht auf die Enden der Verteilung, was ihn empfindlicher für Abweichungen in den Extremen macht. Er eignet sich zur Erkennung sowohl allgemeiner Abweichungen von der Normalität als auch spezifischer Abweichungen in den Enden.	Wie andere Normalitätstests kann er in großen Stichproben überempfindlich sein und signifikante Ergebnisse selbst für geringfügige Abweichungen liefern, die die Gesamtanalyse nicht beeinträchtigen.
Lilliefors-Test	Eine Variante des K-S-Tests, der Lilliefors-Test ist für Situationen gedacht, in denen Mittelwert und Varianz der Population unbekannt sind. Er wird oft als Alternative zum Shapiro-Wilk-Test verwendet, wenn die Stichprobengröße moderat bis groß ist.	Der Lilliefors-Test teilt ähnliche Einschränkungen mit anderen Tests, insbesondere seine Empfindlichkeit gegenüber kleinen Abweichungen in großen Datensätzen.
D’Agostino-Pearson-Test	Dieser Test bewertet, ob die Schiefe und Kurtosis der Daten signifikant von denen einer normalen Verteilung abweichen. Er kombiniert zwei Tests – einen für Schiefe und einen für Kurtosis – und eignet sich zur Erkennung sowohl von Asymmetrie als auch von schweren oder leichten Enden.	Dieser Test setzt eine angemessen große Stichprobengröße (mindestens 20–50 Datenpunkte) für genaue Ergebnisse voraus und kann wie andere Tests geringfügige Abweichungen in großen Stichproben als signifikant markieren.

die Gültigkeit statistischer Tests. Eine falsche Annahme der Normalität kann zu ungültigen Ergebnissen führen, die letztlich klinische Entscheidungen beeinflussen. Beispielsweise basieren diagnostische Schwellenwerte, wie die Referenzbereiche für die Dicke der retinalen Nervenfaserschicht, auf der Normalverteilung, um pathologische Zustände genau zu identifizieren.^{81,82} Abweichungen von dieser Annahme können zu einer Fehlklassifizierung von Patienten führen, was Diagnose und Behandlung beeinträchtigt.

In der personalisierten Medizin ermöglicht das Verständnis der Verteilung von Biomarkern, wie dem intraokularen Druck oder der zentralen Hornhautdicke, es Klinikern, Interventionen auf individuelle Patientenprofile zuzuschneiden. Ein umfassendes Verständnis der Verteilung dieser Bio-

marker innerhalb von Populationen informiert individuelle Behandlungspläne, berücksichtigt Patientenvariabilität und gewährleistet präzisere, zielgerichtete therapeutische Ansätze.^{83,84}

Die Bedeutung der Sicherstellung der Normalität in klinischen Daten ist besonders evident in Studien zur Behandlungseffektivität. Beispielsweise ist die Bestätigung, dass Messungen der IOP-Reduktion in einer klinischen Studie, die ein neues Glaukommedikament evaluiert, normal verteilt sind, entscheidend für gültige Vergleiche mit parametrischen Tests.⁸⁵ Eine genaue statistische Analyse erleichtert evidenzbasierte Entscheidungen bezüglich der Wirksamkeit der Behandlung, was letztlich deren Übernahme in die klinische Praxis bestimmt.

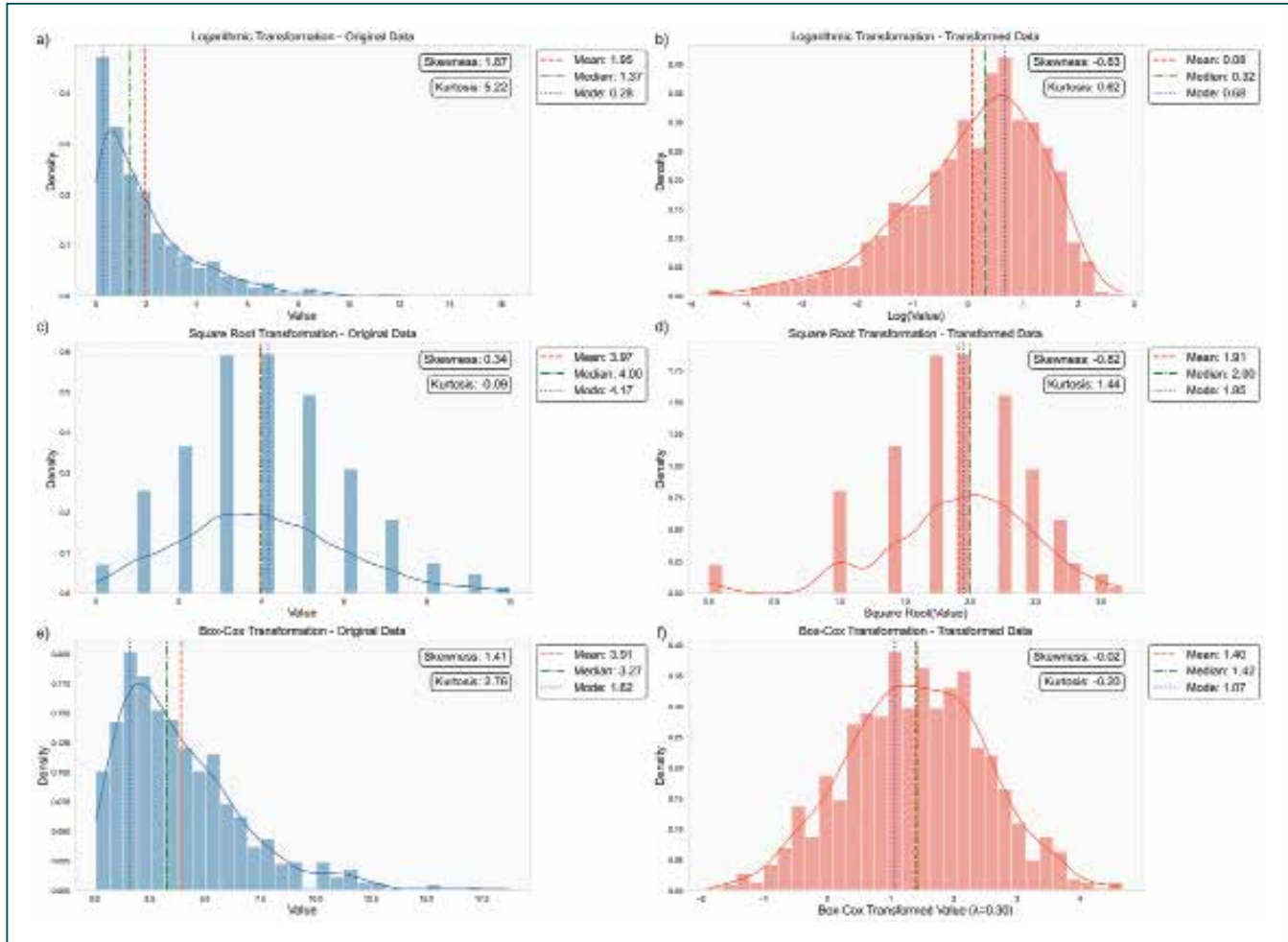


Bild 9: Ursprüngliche und transformierte Datendistributionen. Die Panels a, c und e zeigen die ursprünglichen Exponential-, Poisson- und Gamma-Verteilungen. Die Panels b, d und f zeigen die entsprechenden transformierten Verteilungen mithilfe von logarithmischen, Quadratwurzel- und Box-Cox-Transformationen. Jedes Histogramm umfasst eine Kerneldichteschätzung (KDE) und Annotationen für Mittelwert (rote gestrichelte Linie), Median (grüne gepunktet-gestrichelte Linie), Modus (blaue gepunktete Linie), Schiefe und Kurtosis, die hervorheben, wie jede Transformation die Datenverteilung normalisiert.

Datentransformation

Wenn Daten erheblich von der Normalität abweichen, können fortgeschrittene Techniken dazu beitragen, die Normalität anzunähern und die Genauigkeit der statistischen Analysen zu verbessern. Eine häufig verwendete Strategie ist die Datentransformation. Beispielsweise adressieren logarithmische Transformationen effektiv rechtsschiefe Daten mit positiven Werten, wie die Reduzierung der Schiefe in IOP-Messungen, die hohe Ausreißer enthalten (**Bild 9a**).^{58,72} Ebenso helfen Quadratwurzeltransformationen (Square Root Transformation) bei Zähldaten oder Variablen, die einer Poisson-Verteilung folgen, wie die Anzahl der Mikroaneurysmen in Studien zur diabetischen Retinopathie (**Bild 9b**).^{86,87} Eine weitere leistungsstarke Technik ist die Box-Cox-Transformation, die systematisch die optimale Potenztransformation zur Erreichung der Normalität identifiziert und einen formalisierten Ansatz zur Datennormalisierung bietet (**Bild 9c**).⁸⁸

Einige augenheilkundliche Daten folgen möglicherweise keiner Standardverteilung. Bei gemischten Verteilungen sind spezialisierte Ansätze erforderlich. Beispielsweise können bei

Messungen des Refraktionsfehlers bimodale Verteilungen auftreten, die durch zwei deutliche Peaks gekennzeichnet sind, was separate Peaks für Myopie und Hyperopie widerspiegelt.^{76,77} Ähnlich können multimodale Verteilungen in Populationen mit Untergruppen erscheinen, wie altersbedingte Variationen der Linsentrübung, die Techniken wie endliche Mischmodelle erfordern, um die Daten genau zu modellieren.^{89,90} Wenn Transformationen die Normalität nicht erreichen können, bieten nichtparametrische Methoden eine praktikable Alternative. Verschiedene statistische Tests werden in den nachfolgenden Artikeln dieser Serie diskutiert.⁹¹

Berichterstattung deskriptiver Statistiken in Veröffentlichungen oder Studienberichten

Deskriptive Statistiken fassen die Schlüsselmerkmale eines Datensatzes zusammen und bieten wesentliche Einblicke in die Patientendemografie, klinische Ergebnisse und kritische Variablen. Indem sie Daten prägnant zusammenfassen, helfen

deskriptive Statistiken Klinikern und Forschern, die allgemeinen Eigenschaften ihrer Studienpopulation zu verstehen und signifikante Trends hervorzuheben.

Berichtsstandards für verschiedene Arten von Forschung:

Die Einhaltung etablierter Berichtsstandards gewährleistet Klarheit, Reproduzierbarkeit und Transparenz in verschiedenen Forschungsdisziplinen. Unterschiedliche Studien, wie randomisierte kontrollierte Studien, Beobachtungsstudien, qualitative Forschung und Meta-Analysen, erfordern spezifische Berichtsrichtlinien (z. B. CONSORT für klinische Studien,⁹² STROBE für Beobachtungsstudien,⁹³ PRISMA für systematische Übersichtsarbeiten.⁹⁴ Forscher sollten sich mit den entsprechenden Richtlinien für ihr Studiendesign vertraut machen und diese befolgen, um die Qualität und Glaubwürdigkeit ihrer Berichte zu erhöhen. Um deskriptive Statistiken effektiv zu berichten, sollten Forscher:

- Transparenz in der Datenhandhabung sicherstellen, wie beispielsweise den Umgang mit fehlenden Daten oder die Verwaltung von Ausreißern.
- Die Datentypen klar beschreiben (z. B. nominal, ordinal, kontinuierlich).
- Maße der zentralen Tendenz zusammen mit Variabilität berichten.
- Stichprobengrößen für jede Variable angeben.
- Die Wahl statistischer Maße rechtfertigen (z. B. Verwendung des Medians aufgrund einer schiefen Verteilung).

Die besten Praktiken für die Berichterstattung fehlender Daten beinhalten die Unterscheidung zwischen vollständig zufällig fehlenden (MCAR), zufällig fehlenden (MAR) und nicht zufällig fehlenden (MNAR) Daten. Forscher sollten geeignete statistische Techniken einsetzen (z. B. multiple Imputationen oder Maximum-Likelihood-Schätzung) und das Ausmaß und Muster fehlender Daten berichten sowie Sensitivitätsanalysen durchführen, um die Robustheit der Ergebnisse zu bewerten.

Die Identifikation und das Management von Ausreißern sollten transparent und systematisch sein. Standardmethoden umfassen z-Scores, modifizierte z-Scores (für nicht normale Daten), und grafische Techniken wie Boxplots. Forscher müssen Ausreißer bewerten, um ihre Natur zu bestimmen, ob es sich um Datenfehler, gültige physiologische Extreme oder Messungenauigkeiten handelt. Entscheidungen bezüglich des Umgangs mit Ausreißern sollten klar dokumentiert werden, um eine kritische Bewertung und Reproduktion zu ermöglichen.

Die Datenvisualisierung ist wesentlich, um die Form der Verteilung und Variabilität zu vermitteln und hilft dabei, Probleme wie Schiefe oder Ausreißer zu identifizieren. Gängige Methoden umfassen Histogramme, Boxplots und Q-Q-Plots.

Transparenz in der Berichterstattung ist entscheidend für die Reproduzierbarkeit. Forscher sollten grafische Darstellungen der Datenverteilungen einbeziehen und Annahmen zur Normalität sowie eventuelle Korrekturmaßnahmen diskutieren.

Deskriptive Statistiken sollten mit der klinischen Relevanz verknüpft sein, um ihre Auswirkungen auf klinische Ergebnis-

se und Entscheidungsfindung zu demonstrieren. Beispielsweise kann das Definieren normativer Bereiche für klinische Messungen bei der Diagnose von Zuständen helfen, während die Variabilität in den Behandlungsergebnissen die klinische Wirksamkeit informieren kann.

Schlussfolgerung

Zusammenfassend ist eine strenge und transparente Berichterstattung deskriptiver Statistiken unerlässlich für die Integrität der klinischen Forschung. Durch die Befolgung bewährter Praktiken – wie das Berichten von Maßen der zentralen Tendenz und Streuung, das ordnungsgemäße Management von Ausreißern und fehlenden Daten und das Bereitstellen klarer Visualisierungen – können Forscher die Zuverlässigkeit und klinische Relevanz ihrer Ergebnisse verbessern, was letztlich die Patientenversorgung und Ergebnisse in der ophthalmologischen Forschung verbessert.

Interessenkonflikt

Die Autoren erklären, dass sie keine Verbindungen zu einer Organisation oder Einrichtung haben, die ein finanzielles Interesse an dem behandelten Thema oder den besprochenen Materialien hat.

Finanzierung

Dieser Artikel erhielt keine Mittel von öffentlichen, kommerziellen oder gemeinnützigen Förderorganisationen.

Beitrag der Autoren

DO war die Hauptautorin und Initiatorin des Artikels und verantwortlich für dessen Konzeption, Entwurf und Verfassung. PMS hat den Artikel überprüft und war verantwortlich für die Überarbeitung im Hinblick auf inhaltliche, grammatische und zusammenhängende Aspekte.

Korrespondierende Autorin



Daniela Oehring

E-Mail:
daniela.oehring@plymouth.ac.uk

Literatur

- 1 Ho, K. C., Stapleton, F., Wiles, L., Hibbert, P., Alkhwajah, S., White, A., Jalbert, I. (2019). Systematic review of the appropriateness of eye care delivery in eye care practice. *BMC Health Serv. Res.*19, 646.
- 2 Chalmers, I., Glasziou, P. *BMJ Blogs*. (2016). Is 85% of health research really 'wasted'? *British Medical Journal* <http://blogs.bmj.com/bmj/2016/01/14/paulglasziou-and-ian-chalmers-is-85-of-health-research-really-wasted/> Accessed. 2016;10.



- 3 Bykov, K., Patorno, E., D'Andrea, E., He, M., Lee, H., Graff, J. S., Franklin, J. M. (2022). Prevalence of Avoidable and Bias-Inflicting Methodological Pitfalls in Real-World Studies of Medication Safety and Effectiveness. *Clin. Pharmacol. Ther.*, 111, 209-217.
- 4 Chapman, S. J., Aldaffaa, M., Downey, C. L., Jayne, D. G. (2019). Research waste in surgical randomized controlled trials. *Br. J. Surg.*, 106, 1464-1471.
- 5 Glasziou, P. P., Chalmers, I. (2018). Research waste is still a scandal – an essay by Paul Glasziou and Iain Chalmers. *BMJ*, 363, k, 4645.
- 6 Torgerson, T., Evans, S., Johnson, B. S., Vassar, M. (2020). The use of systematic reviews to justify phase III ophthalmology trials: an analysis. *Eye (Lond)*. 34, 2041-2047.
- 7 Okumura, Y. (2016). Reducing Research Waste Through Good Reporting Practices. *J. Epidemiol.*, 26, 397-398.
- 8 Ramke, J., Kuper, H., Limburg, H., Kinloch, J., Zhu, W., Lansingh, V. C., Congdon, N., Foster, A., Gilbert, C. E. (2018). Avoidable Waste in Ophthalmic Epidemiology: A Review of Blindness Prevalence Surveys in Low and Middle Income Countries 2000-2014. *Ophthalmic Epidemiol.* 25, 13-20.
- 9 Kumar, A. Statistical Errors in Ophthalmic Research. *Indian Journal of Clinical and Experimental Ophthalmology*. 3, 241-242.
- 10 Bunce, C., Czanner, G. (2022). Common Statistical Issues in Ophthalmic Research. *Ophthalmic Epidemiology*: CRC Press; pp. 75-84.
- 11 Van Rossum, G., Drake, Jr. F. L. (1995). Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam.
- 12 Bitsios, P., Prettyman, R., Szabadi, E. (1996). Changes in autonomic function with age: a study of pupillary kinetics in healthy young and old people. *Age Ageing*. 25, 432-438.
- 13 Kiel, M., Grabitz, S. D., Hopf, S., Koeck, T., Wild, P. S., Schmidtman I., Lackner, K. J., Münzel, T., Beutel, M. E., Pfeiffer, N., Schuster, A. K. (2022). Distribution of Pupil Size and Associated Factors: Results from the Population-Based Gutenberg Health Study. *J. Ophthalmol.*, 9520512.
- 14 Kohnen, E. M., Zubcov, A. A., Kohnen, T. (2004). Scotopic pupil size in a normal pediatric population using infrared pupillometry. *Graefes Arch. Clin. Exp. Ophthalmol.*, 242, 18-23.
- 15 Lu, R., Zhang, X., Shi, J. (2021). Tonic pupil size and its variability are associated with fluid intelligence in adolescents aged 11-14 years. *Psych. J.*, 10, 20-32.
- 16 Salati, C., Salvetat, M. L., Zeppieri, M., Brusini, P. (2007). Pupil size influence on the intraocular performance of the multifocal AMO-Array intraocular lens in elderly patients. *Eur. J. Ophthalmol.*, 17, 571-578.
- 17 Winn, B., Whitaker, D., Elliott, D. B., Phillips, N. J. (1994). Factors affecting light-adapted pupil size in normal human subjects. *Invest. Ophthalmol. Vis. Sci.*, 35, 1132-1137.
- 18 Bomotti, S., Lau, B., Klein, B. E. K., Lee, K. E., Klein, R., Duggal, P., Klein, A. P. (2018). Refraction and Change in Refraction Over a 20-Year Period in the Beaver Dam Eye Study. *Invest. Ophthalmol. Vis. Sci.*, 59, 4518-4524.
- 19 Goldblum, D., Brugger, A., Haselhoff, A., Schmickler, S. (2013). Longitudinal change of refraction over at least 5 years in 15,000 patients. *Graefes Arch. Clin. Exp. Ophthalmol.*, 251, 1431-1436.
- 20 Ip, J. M., Kifley, A., Rose, K. A., Mitchell, P. (2007). Refractive findings in children with astigmatic parents: the Sydney Myopia Study. *Am. J. Ophthalmol.*, 144, 304-306.
- 21 Sawada, A., Tomidokoro, A., Araie, M., Iwase, A., Yamamoto, T., Tajimi Study G. (2008). Refractive errors in an elderly Japanese population: the Tajimi study. *Ophthalmology*, 115, 363-370 e3.
- 22 Wesemann W. [Analysis of spectacle lens prescriptions shows no increase of myopia in Germany from 2000 to 2015]. *Ophthalmologie*, 115, 409-417.
- 23 Williams, K. M., Bertelsen, G., Cumberland, P., Wolfram, C., Verhoeven, V. J., Anastasopoulos, E., Buitendijk, G. H., Cougnard-Grégoire, A., Creuzot-Garcher, C., Erke, M. G., Hogg, R., Höhn, R., Hysi, P., Khawaja, A. P., Korobelnik, J. F., Ried, J., Vingerling, J. R., Bron, A., Dartigues, J. F., Fletcher, A., Hofman, A., Kuijpers, R. W., Luben, R. N., Oxele, K., Topouzis, F., van Hanno, T., Mirshahi, A., Foster, P. J., van Duijn, C. M., Pfeiffer, N., Delcourt, C., Klaver, C. C., Rahi, J., Hammond, C. J.; European Eye Epidemiology (E(3)) Consortium. (2015). Increasing Prevalence of Myopia in Europe and the Impact of Education. *Ophthalmology*, 122, 1489-1497.
- 24 Williams, K. M., Verhoeven, V. J., Cumberland, P., Bertelsen, G., Wolfram, C., Buitendijk, G. H., Hofman, A., van Duijn, C. M., Vingerling, J. R., Kuijpers, R. W., Höhn, R., Mirshahi, A., Khawaja, A. P., Luben, R. N., Erke, M. G., van Hanno, T., Mahroo, O., Hogg, R., Gieger, C., Cougnard-Grégoire, A., Anastasopoulos, E., Bron, A., Dartigues, J. F., Korobelnik, J. F., Creuzot-Garcher, C., Topouzis, F., Delcourt, C., Rahi, J., Meitinger, T., Fletcher, A., Foster, P. J., Pfeiffer, N., Klaver, C. C., Hammond, C. J. (2015). Prevalence of refractive error in Europe: the European Eye Epidemiology (E(3)) Consortium. *Eur. J. Epidemiol.*, 30, 305-315.
- 25 Hashemi, H., Heydarian, S., Aghamirsalim, M., Yekta, A., Hashemi, A., Sajadi, M., Khabazkhoob M. (2023). Distribution and associated factors of intraocular pressure in the older population: Tehran Geriatric Eye Study. *Int. J. Ophthalmol.*, 16, 418-426.
- 26 Hoffmann, E. M., Aghayeva, F., Wagner, F. M., Fiess, A., Nagler, M., Munzel, T., Wild, P. S., Beutel, M. E., Schmidtman, I., Lackner, K. J., Pfeiffer, N., Schuster, A. K. (2022). Intraocular Pressure and Its Relation to Ocular Geometry: Results From the Gutenberg Health Study. *Invest. Ophthalmol. Vis. Sci.*, 63, 40.
- 27 Hoffmann, E. M., Lamparter, J., Mirshahi, A., Elflein, H., Hoehn, R., Wolfram, C., Lorenz, K., Adler, M., Wild, P. S., Schulz, A., Mathes, B., Blettner, M., Pfeiffer, N. (2013). Distribution of central corneal thickness and its association with ocular parameters in a large central European cohort: the Gutenberg health study. *PLoS One*, 8, e66158.
- 28 Ma, D., Wei, S., Sun, Y., Li, S. M., An, W. Z., Hu, J. P., Cao, K., Yang, X. H., Lin, C. X., Guo, J. Y., Li, H., Fu, J., Wang, N. (2021). Distribution of IOP and its relationship with refractive error and other factors: the Anyang University Students Eye Study. *Int. J. Ophthalmol.*, 14, 554-559.
- 29 Suzuki, S., Suzuki, Y., Iwase, A., Araie, M. (2005). Corneal thickness in an ophthalmologically normal Japanese population. *Ophthalmology*. 112, 1327-1336.
- 30 Wang, Y. X., Xu, L., Wie, W. B., Jonas, J. B. (2011). Intraocular pressure and its normal range adjusted for ocular and systemic parameters. *The Beijing Eye Study 2011*. *PLoS One*, 13, e0196926.
- 31 Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Del Rio, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T. E. (2015). Array programming with NumPy. *Nature*, 585, 357-362.
- 32 McKinney, W. (2015). Pandas, python data analysis library. URL <http://pandas.pydata.org>, 3-15.
- 33 Barrett, P., Hunter, J., Miller, J. T., Hsu, J. C., (2025). matplotlib – A Portable Python Plotting Package. (Shoppell, L., Britton, M. C., Ebert, eds.) R. Astronomical data analysis software and systems XIV. ASP Conference Series, 347, 91-95.
- 34 Waskom, M., Botvinnik, O., O’Kane, D., Hobson, P., Lukauskas, S., Gempertine, D. C., Augspurger, T., Halchenko, Y., Cole, J. B., Warmenhoven, J., De Ruiter, J., Pye, C., Hoyer, S., Vanderplas, J., Villalba, S., Kunter, G., Quintero, E., Bachant, P., Martin, M., Meyer, K. (2017). mwwaskom/sea-born, v0.8.1, Zenodo.
- 35 Oehring, D., Serra, P. M. (2024). Advancing Statistical Literacy in Eye Care: A Series for Enhanced Clinical Decision-Making. 1 ed. OSF.
- 36 Hellerstein, J. M., Heer, J., Kandel, S. (2018). Self-Service Data Preparation: Research to Practice, *IEEE Data Eng Bull.*, 41, 23-34.
- 37 Kordon, A. K. (2020). Data Preparation. *Applying Data Science*. 221-249.
- 38 Newman, D. A. (2014). Missing Data. *Organizational Research Methods*, 17, 372 - 411.
- 39 Kang, H. (2013). The prevention and handling of the missing data. *Korean J Anesthesiol.*, 64, 402-406.
- 40 Rubin, D. B. (1976). Inference and Missing Data, *Psychometrika*, 63, 3, 581-592.
- 41 Suresh, R., Yu, H. J., Thoverson, A., Swisher, J., Apolinario, M., Zhou, B., Shah, A. R., Fish, R. H., Wykoff, C. C. (2020). Loss to Follow-Up Among Patients With Proliferative Diabetic Retinopathy in Clinical Practice. *Am. J. Ophthalmol.*, 215, 66-71.
- 42 Vengadesan, N., Ahmad, M., Sindal, M. D., Sengupta, S. (2017). Delayed follow-up in patients with diabetic retinopathy in South India: Social factors and impact on disease progression. *Indian. J. Ophthalmol.*, 65, 376-384.
- 43 Elegbede, A., Andrei, A., Holen, K. D. (2009). Reconsenting patients with cancer on clinical trials: Does added risk influence continued participation? *Journal of Clinical Oncology*, 27, 15 suppl., e15610.
- 44 Nantz, E., Liu-Seifert, H., Skljarevski, V. (2009). Predictors of premature discontinuation of treatment in multiple disease states. *Patient Prefer Adherence*, 3, 31-43.
- 45 Hansen, J., Ahern, S., Earnest, A. (2023). Evaluations of statistical methods for outlier detection when benchmarking in clinical registries: a systematic review. *BMJ Open*, 13, e069130.
- 46 Gress, T. W., Denvir, J., Shapiro, J. I. (2018). Effect of removing outliers on statistical inference: implications to interpretation of experimental data in medical research. *Marshall J. Med.*, 4, 9.
- 47 Prasad, S. (2014). *Advanced Statistical Methods*. Publisher, Springer Nature Singapore.

- 48 Deneshkumar, V., Senthamaraikannan, K., Manikandan, M. (2014). Identification of Outliers in Medical Diagnostic System Using Data Mining Techniques. *International journal of statistics and applications*, 4, 241-248.
- 49 Forster, J. E., MaWhinney, S., Ball, E. L., Fairclough, D. (2012). A varying-coefficient method for analyzing longitudinal clinical trials data with nonignorable dropout. *Contemp. Clin. Trials*, 33, 378-385.
- 50 Anderson, A. J., Bedggood, P. A., George Kong, Y. X., Martin, K. R., Vingrys, A. J. (2017). Can Home Monitoring Allow Earlier Detection of Rapid Visual Field Progression in Glaucoma? *Ophthalmology*, 124, 1735-1742.
- 51 Crabb, D. P., Garway-Heath, D. F. (2012). Intervals between visual field tests when monitoring the glaucomatous patient: wait-and-see approach. *Invest. Ophthalmol. Vis. Sci.*, 53, 2770-2776.
- 52 Verma, S., Nongpiur, M. E., Atalay, E., Wie, X., Husain, R., Goh, D., Perera, S. A., Aung, T. (2017). Visual Field Progression in Patients with Primary Angle-Closure Glaucoma Using Pointwise Linear Regression Analysis. *Ophthalmology*, 124, 1065-1071.
- 53 Zhang, S., Chen, Y., Li, Z., Wang, W., Xuan, M., Zhang, J., Hu, Y., Chen, Y., Xiao, O., Yin, Q., Zheng, Y., He, M., Han, X. (2024). Axial Elongation Trajectories in Chinese Children and Adults With High Myopia. *JAMA Ophthalmol.*, 142, 87-94.
- 54 Flores-Moreno, I., Puertas, M., Almazan-Alonso, E., Ruiz-Medrano, J., Garcia-Zamora, M., Vega-Gonzalez, R., Ruiz-Moreno, J. M. (2022). Pathologic myopia and severe pathologic myopia: correlation with axial length. *Graefes Arch. Clin. Exp. Ophthalmol.*, 260, 133-140.
- 55 Adamson, K. A., Prion, S. K. (2013). Making Sense of Methods and Measurement: Measures of Central Tendency. *Clinical Simulation in Nursing*, 9, e617-e618.
- 56 Holladay, J. T. (19997). Proper method for calculating average visual acuity. *J. Refract. Surg.*, 13, 388-391.
- 57 Agarwal, L., Agrawal, N., Badhu, B. P., Lavaju, P. (2019). Central corneal thickness and intraocular pressure in patients of primary open angle glaucoma and normal population in Nepalese population: A hospital based study. *Nepal J. Ophthalmol.*, 11, 46-54.
- 58 Cui, Y., Yang, X., Zhang, G., Guo, H., Zhang, M., Zhang, L., Zeng, J., Liu, Q., Zhang, L., Meng, Q. (2019). Intraocular Pressure in General and Diabetic Populations From Southern China: the Dongguan Eye Study. *Invest Ophthalmol. Vis. Sci.*, 60, 761-769.
- 59 Xu, L., Li, J., Cui, T., Hu, A., Fan, G., Zhang, R., Yang, H., Sun, B., Jonas, J. B. (2005). Refractive error in urban and rural adult Chinese in Beijing. *Ophthalmology*, 112, 1676-1683.
- 60 Robertson, A. H. (1932). Averaging Bacterial Counts. *J. Bacteriol.*, 23, 123-134.
- 61 Rayat, C. S. (2018). Measures of Dispersion. In: *Statistical Methods in Medical Research* (eds. Rayat, C. S.) Springer Nature, pp. 47-52.
- 62 Chen, X., Li, L., Rao, J., Chen, Y. X., Gao, Y., Huang, R. X., Zhou, Q. Z. (2023). Long-term observation on safety and visual quality of implantable collamer lens V4c implantation for myopia correction: a 5-year follow-up. *Int. J. Ophthalmol.*, 16, 1123-1129.
- 63 Liu, H., Li, F. F., Xia, H. J., Zhou, J. (2021). Visual quality after implantation of trifocal intraocular lenses in highly myopic eyes with different axial lengths. *Int J Ophthalmol.*, 14, 371-377.
- 64 Wu, J. H., Moghimi, S., Walker, E., Nishida, T., Liebmann, J. M., Fazio, M. A., Girkin, C. A., Zangwill, L. M., Weinreb, R. N. (2024). Long-term variability of retinal nerve fibre layer thickness measurement in patients with glaucoma of African and European descents. *Br. J. Ophthalmol.*, 108, 1094-100.
- 65 San Pedro, M. J. B., Sosuan, G. M. N., Yap-Veloso, M. I. R. (2024). Correlation of Macular Ganglion Cell Layer + Inner Plexiform Layer (GCL + IPL) and Circumpapillary Retinal Nerve Fiber Layer (cRNFL) Thickness in Glaucoma Suspects and Glaucomatous Eyes. *Clin. Ophthalmol.*, 8, 2313-2325.
- 66 Razmjoo, H., Atarzadeh, H., Kargar, N., Behfarnia, M., Nasrollahi, K., Kamali, A. (2017). The Comparative Study of Refractive Index Variations between Haigis, Srk/T and Hoffer-Q Formulas Used for Preoperative Biometry Calculation in Patients with the Axial Length >25 mm. *Adv. Biomed. Res.*, 6, 78.
- 67 Robertson, A. O., Tadic, V., Cortina-Borja, M., Rahi, J., (2021). Child Vision Pg. Feasibility of using patient-reported outcome measures with visually impaired children/young people attending paediatric ophthalmology clinics. *Arch. Dis. Child.*, 106, 687-692.
- 68 Hepworth, L. R., Rowe, F. J., Burnside, G. (2019). Development of a patient reported outcome measures for measuring the impact of visual impairment following stroke. *BMC Health Serv. Res.*, 19, 348.
- 69 Baek, H. J., Kim, H. S., Kim, N., Choi, Y. J., Kim, Y. J. (2012). Percent change of perfusion skewness and kurtosis: a potential imaging biomarker for early treatment response in patients with newly diagnosed glioblastomas. *Radiology*, 264, 834-843.
- 70 Redmond, C. K., Colton, T. (2001). *Biostatistics in Clinical Trials*. Wiley.
- 71 Cheng, C. Y., Hsu, W. M., Liu, J. H., Tsai, S. Y., Chou, P. (2003). Refractive errors in an elderly Chinese population in Taiwan: the Shihpai Eye Study. *Invest. Ophthalmol. Vis.*, 44, 4630-4638.
- 72 Colton T, Ederer F. The distribution of intraocular pressures in the general population. *Surv. Ophthalmol.*, 25, 123-129.
- 73 Chauhan, B. C., Henson, D. B. (1987). The distribution of visual field scores in a normal population. *Ophthalm. Physiol. Opt.*, 7, 109-115.
- 74 Schimmelpfennig, B. H. (1984). Direct and indirect determination of nonuniform cell density distribution in human corneal endothelium. *Invest. Ophthalmol. Vis. Sci.*, 25, 223-229.
- 75 Ono, T., Mori, Y., Nejima, R., Iwasaki, T., Miyai, T., Miyata, K. (2021). Corneal endothelial cell density and morphology in ophthalmologically healthy young individuals in Japan: An observational study of 16842 eyes. *Sci. Rep.*, 11, 18224.
- 76 Fricke, T. R., Keay, L., Resnikoff, S., Tahhan, N., Koumbo, O., Paudel, P., Ayton, L. N., Britten-Jones, A. C., Kweon, S., Li, J. C. H., Lee, L., Wagner, P., Weng, R., Beranger, B., Olivier, J. (2023). Improving population-level refractive error monitoring via mixture distributions. *Ophthalmic Physiol. Opt.*, 43, 445-453.
- 77 Cumberland, P. M., Bao, Y., Hysi, P. G., Foster, P. J., Hammond, C. J., Rahi, J. S.; UK Biobank Eyes & Vision Consortium. (2015). Frequency and Distribution of Refractive Error in Adult Life: Methodology and Findings of the UK Biobank Study. *PLoS One.*, 10, e0139780.
- 78 Ojaimi, E., Rose, K. A., Morgan, I. G., Smith, W., Martin, F. J., Kifley, A., Robaei, D., Mitchell, P. (2005). Distribution of ocular biometric parameters and refraction in a population-based study of Australian children. *Invest. Ophthalmol. Vis. Sci.* 46, 2748-2754.
- 79 Adams, W. J. (1974). *The life and times of the central limit theorem*. Kaedmon Publishing Company.
- 80 Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika.*, 52, 591-611.
- 81 Lingham, G., Lee, S. S., Charng, J., Clark, A., Chen, F. K., Yazar, S., Macckey, D. A. (2021). Distribution and Classification of Peripapillary Retinal Nerve Fiber Layer Thickness in Healthy Young Adults. *Transl. Vis. Sci. Technol.*, 10, 3.
- 82 Yanni, S. E., Wang, J., Cheng, C. S., Locke, K. I., Wen, Y., Birch, D. G., Birch, E. E. (2013). Normative reference ranges for the retinal nerve fiber layer, macula, and retinal layer thicknesses in children. *Am. J. Ophthalmol.*, 155, 354-360 e1.
- 83 Wolfs, R. C., Klaver, C. C., Vingerling, J. R., Grobbee, D. E., Hofman, A., de Jong, P. T. (1997). Distribution of central corneal thickness and its association with intraocular pressure: The Rotterdam Study. *Am. J. Ophthalmol.*, 123, 767-772.
- 84 Ramm, L., Spoerl, E., Pillunat, L. E., Terai, N. (2020). Is the Corneal Thickness Profile Altered in Diabetes Mellitus? *Curr. Eye Res.*, 45, 1228-1234.
- 85 Heijl, A., Leske, M. C., Bengtsson, B., Hyman, L., Bengtsson, B., Hussein, M. Early Manifest Glaucoma Trial Group. (2002). Reduction of intraocular pressure and glaucoma progression: results from the Early Manifest Glaucoma Trial. *Arch. Ophthalmol.*, 120, 1268-1279.
- 86 Tavakoli, M., Pourreza-Shahri, R., Pourreza, H. R., Mehdizadeh, A., Banaee, T., Toosi, M. T. B. A. (2013). complementary method for automated detection of microaneurysms in fluorescein angiography fundus images to assess diabetic retinopathy. *Pattern Recognit.*, 46, 2740-2753.
- 87 Chen, S. J., Chou, P., Lee, A. F., Lee, F. L., Hsu, W. M., Liu, J. H., Tung, T. H. (2010). Microaneurysm number and distribution in the macula of Chinese type 2 diabetics with early diabetic retinopathy: a population-based study in Kinmen, Taiwan. *Acta Diabetol.*, 47, 35-41.
- 88 Blackie, C. A., Harris, W. F. (1997). Refraction and keratometry: departures from and transformations toward multivariate normality. *Optom. Vis. Sci.*, 74, 452-458.
- 89 Manouchehri, N., Rahmanpour, M., Bouguila, N., Fan, W. (2019). Learning of Multivariate Beta Mixture Models via Entropy-based component splitting. *IEEE Symposium Series on Computational Intelligence (SSCI)*, 2825-2832.
- 90 Quintero, F. O. L., ContrerasReyes, J. E. (2017). Estimation for finite mixture of simplex models: applications to biomedical data. *Statistical Modelling*, 18, 129-148.
- 91 Fife, D. A. (2019). *The Eight Steps of Data Analysis: A Graphical Framework to Promote Sound Statistical Analysis. Perspectives on Psychological Science*, 15, 1054-1075.

92 Moher, D., Schulz, K. F., Altman, D., Group, C. (2001). The CONSORT Statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Lancet*, 357, 1191-1194.

93 von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., Vandenbroucke, J. P.; STROBE-Initiative. (2008). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Internist (Berl)*, 688-693.

94 Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G.; Prisma Group (2010). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int. J. Surg.*, 8, 336-341.

95 van der Valk, R., Schouten, J. S., Webers, C. A., Hendrikse, F., Prins, M. H. (2006). Changes in glaucoma treatment and achieved IOP after introduction of new glaucoma medication. *Graefes Arch. Clin. Exp. Ophthalmol.*, 244, 1267-1272.

96 Bengtsson, B., Leske, M.C., Hyman, L., Heijl, A.; Early Manifest Glaucoma Trial Group (2007). Fluctuation of intraocular pressure and glaucoma progression in the early manifest glaucoma trial. *Ophthalmology*, 114, 205-209.

97 Chevret, S., Seaman, S., Resche-Rigon, M. (2005). Multiple imputation: a mature approach to dealing with missing data. *Intensive Care Med.*, 41, 348-350.

98 Saffari, S. E., Volovici, V., Ong, M. E. H., Goldstein, B. A., Vaughan, R., Dammers, R., Steyerberg, E. W., Liu, N. (2022). Proper Use of Multiple Imputation and Dealing with Missing Covariate Data. *World Neurosurg.*, 161, 284-290.

99 Huque, M. H., Carlin, J. B., Simpson, J. A., Lee, K. J. A (2018). Comparison of multiple imputation methods for missing data in longitudinal studies. *BMC Med. Res. Methodol.*, 18, 168.

100 Rosato, R., Pagano, E., Testa, S., Zola, P., di Cuonzo, D. (2021). Missing data in longitudinal studies: Comparison of multiple imputation methods in a real clinical setting. *J. Eval. Clin. Pract.*, 27, 34-41.

101 Janssen, K. J., Donders, A. R., Harrell, F. E., Jr., Vergouwe, Y., Chen, Q., Grobbee, D. E., Moons, K. G. (2010). Missing covariate data in medical research: to impute is better than to ignore. *J. Clin. Epidemiol.*, 63, 721-727.

102 Cao, Y., Allore, H., Vander, Wyk, B., Gutman, R. (2022). Review and evaluation of imputation methods for multivariate longitudinal data with mixed-type incomplete variables. *Stat. Med.*, 41, 5844-5876.

103 Laird, N. M., Lange, N., Stram, D. O. (1987). Maximum likelihood computations with repeated measures: application of the EM algorithm. *Journal of the American Statistical Association*, 82, 97-105.

104 Anderson, A. J., Johnson, C. A. (2013). How useful is population data for informing visual field progression rate estimation? *Invest. Ophthalmol. Vis. Sci.*, 54, 2198-2206.

105 Daza, E. J., Hudgens, M. G., Herring, A. H. (2017). Estimating inverse-probability weights for longitudinal data with dropout or truncation: The xtrccipw command. *Stata. J.*, 17, 253-278.

106 Shardell, M., Hicks, G. E., Ferrucci, L. (2015). Doubly robust estimation and causal inference in longitudinal studies with dropout and truncation by death. *Biostatistics*, 16, 155-168.

107 Kong, F., Chen, Y. F., Jin K. (2009). A bias correction in testing treatment efficacy under informative dropout in clinical trials. *J. Biopharm. Stat.*, 19, 980-1000.

108 Liu, G. F., Liu, F., Mehrotra, D. V. (2020). Model Averaging Using Likelihoods That Reflect Poor Outcomes for Clinical Trial Dropouts. *Statistics in Biopharmaceutical Research*, 12, 79 - 89.

Appendix

Tabelle A1: Zusammenfassung zum Umgang mit fehlenden Daten

Methoden	Anwendbarkeit	Vorteile	Einschränkungen	Klinische Anwendung
Analyse vollständiger Fälle (listweises oder fallweises Löschen)	Anwendbar bei MCAR	Vereinfacht die Analyse	Verringert die Stichprobengröße; potenzielles Bias, wenn nicht MCAR	Studien mit wirklich zufällig fehlenden Daten, aber ausreichender Stichprobengröße. In Studien mit kleinen Stichproben, wie bei seltenen Augenerkrankungen, kann das Entfernen von Fällen mit fehlenden Daten die Schlussfolgerungen der Studie erheblich beeinträchtigen.
Einfache Imputation	MCAR oder wenn fehlende Werte minimal sind	Leicht umzusetzen	Unterschätzt die Variabilität; potenzielles Bias	Vorläufige Analysen; nicht für die Endanalyse empfohlen
a) Mittelwert- oder Median-Imputation	Ersetzt fehlende Werte durch den Mittelwert (für normalverteilte Daten) oder Median (für schiefe Daten) der beobachteten Werte.	Einfach und leicht umzusetzen	Unterschätzt die Variabilität, führt zu verzerrten Standardfehlern und Konfidenzintervallen und ignoriert Beziehungen zwischen Variablen.	Diese Methode kann wahre Zusammenhänge verschleiern, insbesondere in Studien, die die durchschnittliche Netzhautdicke messen, bei denen Beziehungen zwischen Variablen entscheidend sind.
b) Regressionsimputation	Sagt fehlende Werte basierend auf Regressionsmodellen unter Verwendung anderer beobachteter Variablen vorher.	Integriert Beziehungen zwischen Variablen, verbessert die Qualität der imputierten Werte	Berücksichtigt nicht die Unsicherheit in den Vorhersagen, behandelt imputierte Werte als bekannt und kann zu übermäßigem Vertrauen in Schätzungen führen.	Diese Methode könnte in der klinischen Forschung Bias einführen, wie in Studien zum Glaukomfortschritt, bei denen übermäßiges Vertrauen in imputierte Werte zu Fehlinterpretationen führen kann.

Fortsetzung nächste Seite

(Fortsetzung) **Tabelle A1: Zusammenfassung zum Umgang mit fehlenden Daten**

Methode	Anwendbarkeit	Vorteile	Einschränkungen	Klinische Anwendung
c) Letzte Beobachtung übertragen (LOCF)	Diese Methode ersetzt in Längsschnittstudien fehlende Werte durch die zuletzt beobachtete Messung.	Erhält die Stichprobengröße und ist einfach anzuwenden	Geht von keiner Veränderung über die Zeit aus, was oft unrealistisch ist und die Ergebnisse verzerren kann.	LOCF ist problematisch bei progressiven Erkrankungen wie Glaukom, bei denen sich der Augeninnendruck (IOP) im Laufe der Zeit erheblich ändern kann, was diesen Ansatz unangemessen macht. ^{95,96}
Multiple Imputation	Anwendbar bei MAR	Erhält Beziehungen zwischen Variablen; valide Schlussfolgerungen	Rechenintensiv	Prozess: 1. Imputation: Erstelle mehrere vollständige Datensätze, indem fehlende Werte mehrfach aus ihrer Vorhersageverteilung imputiert werden. 2. Analyse: Führe statistische Analysen auf jedem Datensatz separat durch. 3. Pooling: Kombiniere die Ergebnisse mit Rubin's Regeln, um Gesamtschätzungen und Standardfehler abzuleiten. Klinische Anwendung: Imputieren fehlender OCT-Messungen basierend auf beobachteten Patienteneigenschaften. ⁹⁷⁻¹⁰²
Maximum-Likelihood (EM)	Anwendbar bei MAR	Effiziente Schätzungen bei korrektem Modell	Erfordert korrekte Modellspezifikation	Schätzt Modellparameter durch Maximierung der Likelihood-Funktion unter Verwendung aller verfügbaren Daten und berücksichtigt die Struktur der fehlenden Daten. Klinische Anwendung: In Glaukomstudien mit intermittierenden Follow-up-Daten kann der EM-Algorithmus verwendet werden, um Fortschrittsraten zu schätzen und genaue Ergebnisse selbst bei unvollständigen Datensätzen zu liefern. ^{103,104}
Gewichtungsmethoden (IPW)	Anwendbar bei MAR	Nutzt alle Daten; reduziert Bias	Erfordert Modellierung der Wahrscheinlichkeiten für fehlende Daten	Gewichtungsmethoden wie die inverse Wahrscheinlichkeit (IPW) passen die Analyse an fehlende Daten an, indem sie beobachteten Fällen Gewichte zuweisen basierend auf der Wahrscheinlichkeit, beobachtet zu werden. Klinische Anwendung: In Längsschnittstudien zum Komfort von Kontaktlinsen, bei denen Abbruchraten von vorherigen Unbehaglichkeitsstufen abhängen können, kann IPW den differenziellen Abbruch anpassen und die Genauigkeit der Ergebnisse verbessern. ^{105,106}
Modellbasierte Ansätze	Anwendbar bei MNAR	Modelliert den Mechanismus fehlender Daten explizit	Erfordert starke Annahmen über den Mechanismus fehlender Daten, die oft schwer zu verifizieren sind; Fehlerspezifikation kann zu verzerrten Ergebnissen führen.	Umgang mit Abbruch aufgrund nicht beobachteter Nebenwirkungen in Studien zu Augenmedikamenten. ^{49,107,108}

COE Multiple-Choice-Fragebogen



Die Publikation „Biostatistik in Ophthalmologie und Optometrie: Eine Artikelserie zur Unterstützung evidenzbasierter Entscheidungsprozesse – Teil 1: Grundlagen statistischer Methoden in der Augenheilkunde und Optometrie“ wurde von der Gütegemeinschaft Optometrische Leistungen (GOL) als COE Fortbildungsartikel anerkannt. Die unten stehende Reihenfolge entspricht nicht der Online-Version. Pro Frage ist nur eine Antwort richtig. Eine erfolgreiche Teilnahme setzt die Beantwortung von vier der sechs Fragen voraus.

Letzter Termin, um die Fragen online auf der Website www.ocl-online.de zu beantworten, ist der 1. Januar 2026

Frage 1: Welche der folgenden Aussagen bezüglich der Thematik „Forschungverschwendung in der Ophthalmologie“ ist falsch?

- a) Nur 22,4% der Phase-III-Studien in der Augenheilkunde beziehen sich auf systematische Übersichten als Begründung für die Studie, was Chancen verpasst, auf bestehenden Erkenntnissen aufzubauen und somit zu redundanten Forschungen führt.
- b) Schätzungen zufolge könnten bis zu 85% der weltweiten Gesundheitsforschung, die einen erheblichen Teil der jährlichen Ausgaben von 200 Milliarden Dollar ausmachen, durch Nichtveröffentlichung, unklare Berichterstattung und das Fehlen systematischer Überprüfungen im Studiendesign verschwendet werden.
- c) In der Augen- und Sehforschung resultiert die Verschwendung häufig aus unzureichenden Stichprobengrößen, ungeeigneten statistischen Tests oder Fehlinterpretationen der Ergebnisse.
- d) Eine Überprüfung der augenheilkundlichen Literatur ergab, dass nur ein geringer Teil der veröffentlichten Studien statistische Ungenauigkeiten aufweist, was die Glaubwürdigkeit der Forschung nicht beeinträchtigt und keinen Einfluss auf den Fortschritt evidenzbasierter Praxis hat.

Frage 2: Welche der folgenden Aussagen bezüglich statistischer Konzepte ist richtig?

- a) Biologische Messungen sind aufgrund nicht vorhandener individueller biologischer Unterschiede, Umwelteinflüsse und Messfehler von Natur nicht mit Unsicherheit und Variabilität behaftet.
- b) Daten sind nicht monolithisch, sie variieren je nach Art der Messung und der verwendeten Messtechnik.
- c) Kontinuierliche Daten sind auf spezifische separate Werte beschränkt, die typischerweise Zählungen oder ganze Zahlen darstellen, während diskrete Daten einen Bereich, der nur durch die Messgenauigkeit begrenzt wird, darstellt.
- d) Durch die Anwendung statistischer Techniken können Forscher zwar fundierte Schlussfolgerungen über die Bevölkerung ziehen, aber nicht die Zuverlässigkeit der Ergebnisse bewerten und die Wahrscheinlichkeit abschätzen, dass beobachtete Effekte zufällig sind.

Frage 3: Welche der folgenden Aussagen bezüglich fehlender Daten in der Forschung ist falsch?

- a) In der klinischen Forschung können fehlende Daten durch Patientenabbrüche, Nichteinhaltung von Studienprotokollen, technische Probleme bei der Datenerhebung oder Fehler bei der Dateneingabe entstehen.
- b) Die Behandlung fehlender Daten erfordert ein Verständnis der zugrunde liegenden Mechanismen und die Anwendung geeigneter statistischer Methoden.
- c) Fehlende Daten sind ein häufiges Problem in der Forschung, das zu Verzerrungen führen kann, die statistische Aussagekraft reduziert und die Verallgemeinerbarkeit von Studienergebnissen beeinträchtigt.
- d) Fehlende Daten lassen sich in zwei Kategorien einteilen: Zufällig fehlende Daten (Missing Completely At Random, MCAR) und Nicht-zufällig fehlende Daten (Missing At Random, MAR).

Frage 4: Welche der folgenden Aussagen bezüglich der Datenverteilung für die Auswahl geeigneter statistischer Methoden und die genaue Interpretation der Ergebnisse ist falsch?

- a) Die Datenverteilung ist entscheidend für die Auswahl geeigneter statistischer Methoden und die genaue Interpretation der Ergebnisse.
- b) Die Datenverteilung gibt die Wahrscheinlichkeit spezifischer Messungen in einem klinischen Kontext an, leitet Erwartungen und unterstützt die klinische Entscheidungsfindung.
- c) In der Datenverteilung kennen wir mit der Schiefe, der Kurtosis und der Crysiss drei wichtige Verteilungsmerkmale.
- d) Ausreißer sind Datenpunkte, die deutlich von der Gesamtheit der Daten abweichen und können durch Messfehler, Eingabefehler oder tatsächlich extreme Werte aufgrund biologischer Variabilität oder seltener Bedingungen entstehen.

Frage 5: Welche der folgenden Aussagen bezüglich der Normalverteilung in der Statistik ist falsch?

- a) In der Augen- und Sehforschung (EVR) sind viele biometrische Messungen, wie die Hornhautdicke, die axiale Länge und die Sehschärfe in logMAR-Einheiten, Verhältnisskalen-kontinuierliche Daten, die unter der Voraussetzung, dass sie Kriterien für Symmetrie und das Fehlen signifikanter Kurtosis erfüllen, mit der Normalverteilung modelliert werden können.
- b) Die Normalverteilung ist grundlegend in der statistischen Analyse und untermauert mehrere fundamentale Methoden, insbesondere parametrische Tests wie T-Tests, Varianzanalyse (ANOVA) und lineare Regression, die normalverteilte Daten oder Residuen voraussetzen.
- c) Die Normalverteilung ist für die statistische Inferenz nicht essenziell, insbesondere nicht bei der Berechnung von Konfidenz- und Prognoseintervallen.
- d) Die Normalverteilung erleichtert die die Standardisierung, bei der Z-Scores, den Abstand eines Datenpunktes vom Mittelwert in Einheiten der Standardabweichung zu messen.

Frage 6: Welche der folgenden Aussagen bezüglich P-P-Diagramme (Wahrscheinlichkeits-Wahrscheinlichkeits-Diagramme) ist richtig?

- a) P-P-Diagramme zeigen die distinktive Wahrscheinlichkeiten der Stichprobendaten (y-Achse) gegen die einer Normalverteilung (x-Achse).
- b) Jeder Datenpunkt in einem P-P-Diagramm stellt die Wahrscheinlichkeit eines Stichprobendatenpunkts dar, geplottet gegen seine entsprechende Wahrscheinlichkeit aus einer empirischen Verteilung.
- c) Während Q-Q-Diagramme Informationen über die Form der Verteilung liefern, bewerten P-P-Diagramme, wie gut die Gesamtwahrscheinlichkeiten zwischen der theoretischen und der Stichprobenverteilung übereinstimmen.
- d) P-P-Diagramme sind sehr empfindlich gegenüber Abweichungen in den Ausläufern der Verteilung im Vergleich zu Q-Q-Diagrammen.