

Prüfmethoden für Daten, Tabellen und statistischen Auswertungen

Rohdaten und Dokumentation prüfen

Anfordern oder prüfen:

- Rohdatensatz
- Codebook oder Variablenerklärung
- Erhebungsinstrument
- Ein- und Ausschlusskriterien
- Datenerhebungszeitraum
- Dokumentation der Datenbereinigung
- Analysecode oder Auswertungsschritte
- Versionen von Tabellen und Auswertungen
- Ethikvotum, Datenschutzfreigabe oder sonstige Genehmigungen, soweit erforderlich

Konsistenzprüfung

Prüfen:

- Stimmen Fallzahlen in Text, Tabellen und Abbildungen überein?
- Passen Prozentwerte zu den angegebenen absoluten Zahlen?
- Sind Mittelwerte, Standardabweichungen, Konfidenzintervalle und p-Werte plausibel?
- Stimmen Tabellenüberschriften, Variablen und Legenden?
- Werden dieselben Gruppenbezeichnungen konsistent verwendet?
- Sind Rundungen nachvollziehbar?
- Gibt es Ergebnisse ohne vorher beschriebene Analyse?
- Werden Analysen beschrieben, die in den Ergebnissen nicht auftauchen?

Warnsignale:

- wechselnde Fallzahlen ohne Erklärung
- Prozentwerte, die sich nicht aus den Fallzahlen ergeben
- identische Werte an ungewöhnlich vielen Stellen
- zu perfekte Verteilungen
- fehlende Ausreißer bei realistisch streuenden Daten
- Korrelationen oder Effekte, die methodisch nicht plausibel sind
- Tabellen, die wie generische Musterbeispiele wirken

Reproduzierbarkeit prüfen

Prüfen:

- Kann die Auswertung anhand der Rohdaten nachvollzogen werden?
- Können zentrale Ergebnisse mit einfachen Nachrechnungen bestätigt werden?
- Gibt es eine nachvollziehbare Verbindung zwischen Rohdaten, Analyse und Ergebnisdarstellung?
- Sind Transformationen, Ausschlüsse und Bereinigungen dokumentiert?
- Können Studierende erklären, warum bestimmte statistische Verfahren gewählt wurden?

Mögliche Maßnahmen:

- einzelne Berechnungen gemeinsam durchgehen
- Studierende zentrale Ergebnisse nachrechnen lassen
- Auswertungsschritte erläutern lassen
- Analysecode oder Tabellenformeln prüfen
- Abgleich mit Betreuungsprotokollen, Laborbuch oder Forschungstagebuch

Spezifische Warnsignale bei KI-generierten oder erfundenen Daten

Warnsignale:

- Daten wirken vollständig, sauber und widerspruchsfrei, obwohl das Erhebungsdesign Ausfälle erwarten lässt
- keine fehlenden Werte, obwohl die Methode solche erwarten lässt
- unplausible Zeitstempel oder Erhebungsabstände
- doppelte oder fast identische Fälle
- auffällige Wiederholungen in Antwortmustern
- Variablen enthalten Werte außerhalb des zulässigen Bereichs
- beschriebene Stichprobe passt nicht zum Datensatz
- Ergebnisdarstellung ist elaborierter als die verfügbare Datenbasis
- Datenstruktur passt nicht zum beschriebenen Instrument

Revision #1

Created 2026-08-19 12:33:18 UTC by thomas.caspari

Updated 2026-08-19 12:34:39 UTC by thomas.caspari