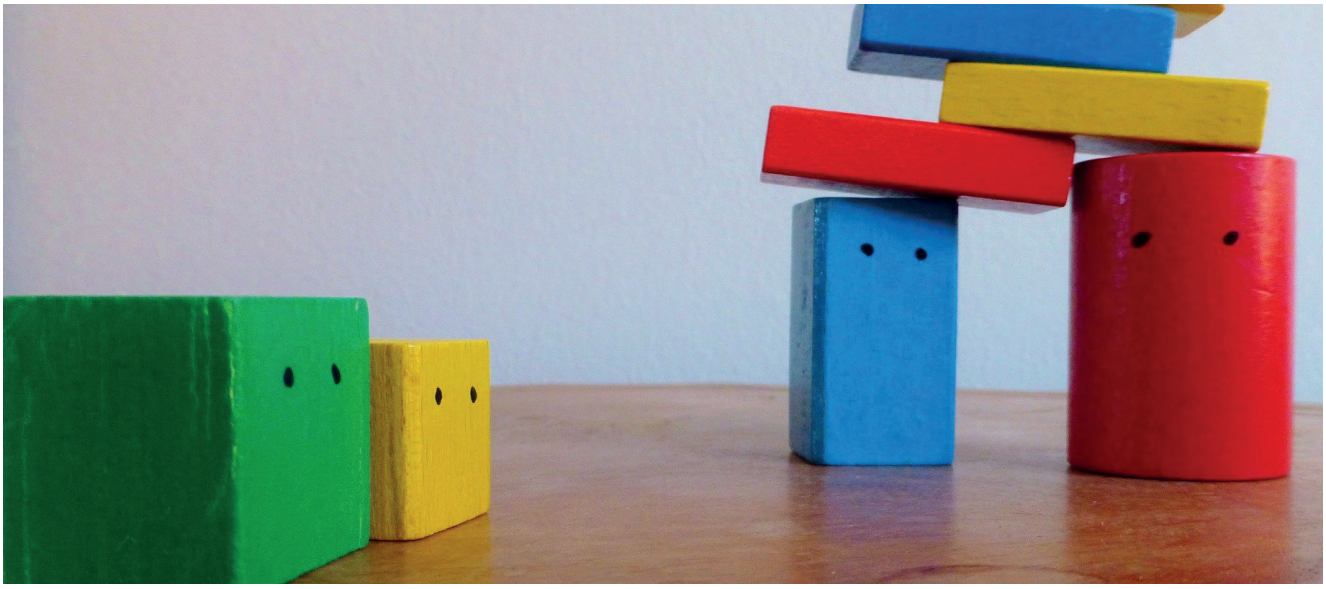


© pixabay 2020, Foto: Ulrike Mair

Die Bausteine der statistischen Datenanalyse

von Daniela Keller

Jede statistische Analyse besteht aus verschiedenen Elementen, die unterschiedliche Zwecke erfüllen, mit unterschiedlichen Methoden umgesetzt werden und meist in einer bestimmten Reihenfolge durchgeführt werden. In diesem Artikel stelle ich Ihnen diese wichtigen Bausteine und ihre Bedeutung vor.

Datencheck

Der Datencheck wird vor der eigentlichen Datenanalyse durchgeführt. Hier untersuchen Sie Ihren Datensatz deskriptiv. Sie berechnen für jede Variable einzeln Kennwerte wie Fallzahl, Häufigkeitsverteilung, Minimum und Maximum. Damit stellen Sie sicher, dass Sie keine groben Fehler in der Datenerhebung haben. Es würde Ihnen hier zum Beispiel auffallen, wenn Sie sehr viele fehlende Werte hätten oder wenn Tippfehler bei der Dateneingabe aufgetreten sind. Diese groben Fehler können Sie so vor der eigentlichen Analyse noch beheben.

Datenbearbeitung

Wenn die Daten nach dem Datencheck soweit in Ordnung sind, kommt nun der Schritt der Datenbearbeitung. Alle Variablen, die Sie für ihre Analyse brauchen, aber nicht direkt gemessen haben, werden Sie jetzt hier als neue Variablen im Datensatz erstellen. Sie bilden Gruppenvariablen, wenn Sie zum Beispiel die Altersvariable als metrische Variable erhoben haben und zusätzlich Altersgruppen berichten möchten. Oder Sie haben Datum- und Uhrzeitangaben erhoben und benötigen für die Analyse Zeitspannen, die Sie nun als neue Variablen berechnen.

Stichprobenbeschreibung

Nachdem die Daten vorbereitet sind, starten Sie nun mit der eigentlichen Datenanalyse. Hier beschreiben Sie im ersten Schritt die Stichprobe. Dazu nehmen Sie die für die Beschreibung der Stichprobe inhaltlich sinnvollen Parameter her. Bei Personendaten kann das zum Beispiel Alter, Geschlecht, Bildungsstand und Familienstand sein. Diese Variablen untersuchen Sie deskriptiv mittels Häufigkeiten, Mittelwert, Standardabweichung, Minimum und Maximum. Falls Sie mit verschiedenen Stichproben arbeiten, zum Beispiel mit einer Interventions- und einer

Kontrollgruppe, führen Sie diese Analyse getrennt für die Stichproben durch.

Deskriptive Statistik und Abbildungen für die Forschungsfrage

Im Anschluss an die allgemeine Beschreibung der Stichprobe starten Sie nun mit der Untersuchung Ihrer Forschungsfrage. Dazu beginnen Sie mit der deskriptiven Auswertung. Das heißt, Sie führen die für die Beantwortung Ihrer Forschungsfrage passenden deskriptiven Analysen durch.

Wenn Sie einen Zusammenhang zwischen zwei metrischen oder ordinalen Parametern untersuchen, ist das zum Beispiel die Korrelation. Wenn Sie eine Veränderung eines metrischen Parameters über die Zeit untersuchen, ist das die Beschreibung des Verlaufs anhand von Mittelwerten und Standardabweichungen für jeden Messzeitpunkt.

Zusätzlich zu diesen beschreibenden Zahlen erstellen Sie geeignete Abbildungen. Für die oben genannte Korrelation wäre das ein Streudiagramm. Für den Zeitverlauf ein Mittelwertdiagramm mit Fehlerbalken.

Durch diesen Schritt bekommen Sie ein Bild davon, was sich in Ihren Daten zeigt. Sie sehen, ob es die zu untersuchenden Zusammenhänge und Unterschiede gibt, wie groß sie sind und in welche Richtung sie gehen. Das hilft Ihnen später, die Ergebnisse der Signifikanztests schneller zu verstehen und wird Ihnen die Interpretation und das Berichten der Ergebnisse erleichtern.

Voraussetzungsprüfung

Wenn Sie – wie in den meisten Studien – auch Signifikanztests für Ihre Analyse geplant haben, dann ist jetzt der richtige Zeitpunkt, die Voraussetzungen der geplanten Analysen zu überprüfen.

Je nach geplanter schließender Methode suchen Sie sich also die zu prüfenden Voraussetzungen heraus. Geprüft werden die Voraussetzungen mit unterschiedlichen Methoden. Teilweise wird mit deskriptiven Methoden (z. B. Zellbesetzung) und grafischen Analysen (z. B. Überprüfung der Normalverteilung mit Q-Q-Diagrammen) gearbeitet. Manchmal werden auch Signifikanztests zur Voraussetzungsprüfung eingesetzt.

Zur Überprüfung der Voraussetzungen gehört auch, gegebenenfalls Maßnahmen zu ergreifen, die bei nicht erfüllten Voraussetzungen notwendig sind. Auch dafür ist jetzt der richtige Zeitpunkt. Das heißt, Sie entscheiden sich an dieser Stelle vielleicht für die Verwendung von nicht-parametrischen Methoden oder für eine Transformation der Daten.

Manche Voraussetzungen lassen sich erst nach der Rechnung der eigentlichen Analyse untersuchen. Das ist zum



© pixabay 2020, Foto: Bernd Hildebrandt

Beispiel für die Normalverteilung der Residuen bei der linearen Regression der Fall. Hier können Sie zum jetzigen Zeitpunkt noch nicht abschließend sagen, ob die Voraussetzungen gegeben sind, sondern führen die finale Voraussetzungsprüfung erst im Anschluss an die Rechnung des Modells durch.

Signifikanztests und Modelle

Wenn Sie nun – soweit vorab möglich – die Voraussetzungen für die gewünschte schließende Statistik überprüft haben und diese erfüllen, dürfen Sie jetzt den geplanten Signifikanztest bzw. das geplante Modell rechnen. Hier untersuchen Sie nun die aus Ihren Forschungsfragen generierten Hypothesen auf Signifikanz. Sie untersuchen die Effekte, Zusammenhänge oder Unterschiede, die Sie vorab in den deskriptiven Analysen beschrieben und visualisiert haben. Anschließend können Sie die Nullhypothese bei signifikantem p-Wert verwerfen und haben damit den Unterschied, Effekt oder Zusammenhang als signifikant nachgewiesen. Bei nicht signifikantem p-Wert können Sie keinen signifikanten Unterschied, Effekt oder Zusammenhang nachweisen.

Nun haben Sie einen Überblick über die Bausteine der statistischen Datenanalyse. Sie kennen ihren Zweck und wissen, wie sie aufeinander aufbauen. Mit diesem Wissen können Sie Ihre nächste Datenanalyse zielgerichtet angehen und haben immer den nächsten Schritt vor Augen.



© privat

Die Autorin

Daniela Keller ist leidenschaftliche Statistik-Expertin und berät Studierende und Wissenschaftler*innen zu allen Themen der statistischen Datenanalyse. Während ihres Studiums der Diplom-Mathematik gründete sie mit Kommilitonen eine studentische statistische Beratung und arbeitete anschließend selbstständig in diesem Feld. Neben Einzelberatungen und Workshops unterstützt sie Ihre Kund*innen seit 2019 mit der Statistik-Akademie, ihrem Online-Mitgliederbereich für alle, die Statistik verstehen und selbstständig anwenden wollen. Ihr Blog (www.statistik-und-beratung.de/blog) und ihr YouTube-Kanal sind Fundgruben für leicht verständlich aufbereitetes Statistikwissen für die Praxis.