

Dos & Don'ts der Datennutzung

Quellen hinterfragen

Um zu prüfen, ob es sich um eine seriöse Quelle handelt, hilft ein Blick auf die Methode der Datenerhebung: Wie groß ist die Stichprobe? Ist sie repräsentativ für die jeweilige Grundgesamtheit? Beantworten die erhobenen Daten wirklich die Fragen, die erforscht werden sollen? Auch ist ein Blick auf den Auftraggeber sinnvoll: Was war die Motivation für die Durchführung der Studie?? Dabei muss eine Untersuchung, die von einem Unternehmen oder einem Interessenverband beauftragt ist, nicht unseriös oder tendenziös sein. Wichtiger ist die methodische Klarheit und Angemessenheit.

Ziele definieren

Schon zu Beginn der Arbeit mit den Daten sollte klar sein, mit welchem Ziel sie genutzt und was geprüft werden soll. Die Daten sollten immer auf eine bestimmte Fragestellung hin untersucht werden. Ein zielloses „Stochern“ hilft nur begrenzt.

Fallzahlen beachten

Die Anzahl der Fälle wird durch den Buchstaben N gekennzeichnet (z. B. „N=2.019 für 2.019 befragte Personen). Diese Fallzahlen sollten Sie immer beachten, auch mit Blick auf Analysen, bei denen nur Teilgruppen berücksichtigt werden, die Zahl der Fälle also kleiner ist als die Gesamtzahl.

Erhebungsmethode berücksichtigen

Ob Befragungsdaten in einem persönlichen Interview, über das Telefon, (postalisch) schriftlich oder online erhoben wurden, hat unterschiedliche Auswirkungen auf die Aussagekraft der Daten, häufig auch auf ihre Qualität. Beispielsweise sind Ergebnisse aus Online-Befragungen im Kontext geringer Literalisierung mit Vorsicht zu betrachten: An solchen Erhebungen können nur Personen teilnehmen, die ausreichend lesen und schreiben können. Zudem bergen diese Verfahren die Gefahr, dass nur solche Personen den Bogen ausfüllen, die sich für ein Thema interessieren – somit aber anders denken als diejenigen, die z. B. bei persönlich-mündlichen Interviews nach anderen (inhaltlichen) Kriterien ausgewählt werden.

Richtige Darstellungsweise wählen

Ergebnisse lassen sich auf unterschiedlichste Art und Weise darstellen. Die Möglichkeiten reichen von Balken- und Tortendiagrammen über Tabellen bis hin zur Ausformulierung von Sachverhalten mit Nennung einzelner Zahlen. Das Format der Darstellung muss zum Inhalt passen und diesen korrekt wiedergeben.

Die Wahl des Diagrammtyps sollte sich nicht an der Schönheit, sondern an der Aussagekraft orientieren.

Kontext beachten

Daten stehen immer in einem Zusammenhang und sollten nicht kontextlos wiedergegeben werden. Beispielsweise haben der Zeitpunkt der Datenerhebung sowie Ort bzw. Reichweite einer Befragung einen erheblichen Einfluss auf die Aussagekraft der Daten und müssen bei der Interpretation berücksichtigt werden (vor oder nach der Corona-Krise; Ort der Erhebung: Bayern, Sylt, Frankreich oder Madagaskar). Ebenso sollte der Kontext der Erstellung beachtet werden: Ist die Studie Ausschnitt einer großen Langzeitstudie oder im Kontext einer universitären Bachelor-Arbeit entstanden? Auch hier sind Reichweite und Kontext nicht unbedingt Qualitätskriterien. Man sollte aber offenlegen und berücksichtigen, aus welchem Zusammenhang die Ergebnisse stammen.

Zusammenhänge (Korrelation) sagen nicht unbedingt etwas über Ursache und Wirkung (Kausalität)

Auch wenn Daten einen Zusammenhang zeigen oder nahelegen, dass also ein Merkmal überzufällig oft zusammen mit einem anderen Merkmal auftritt, heißt das nicht, dass das eine Merkmal die Ursache des anderen ist. Beide Merkmale können sich gegenseitig beeinflussen oder von einem dritten Merkmal verursacht werden. Zum Beispiel korrelieren Übergewicht und Lebenserwartung. Das heißt aber nicht, dass übergewichtige Menschen früher sterben als andere, also Übergewicht die Lebenserwartung kausal bedingt. Eine Korrelation gibt bestenfalls Hinweise.

Auch Nicht-Antworten berücksichtigen

Für die Datenanalyse werden Antworten wie „ich weiß nicht“ oder „keine Angabe“ aus den Datensätzen entfernt. Je nach Kontext und Frage lohnt es sich jedoch auch diese Antworten in den Blick zu nehmen, auch weil sich die Stichprobengröße dadurch von Frage zu Frage unterscheidet.

Skalenniveaus beachten

Ob eine Variable nominal, ordinal oder metrisch skaliert ist, hat Auswirkungen auf die Wahl der Analysemethoden. Nominalskalen sind unsortierte Kategorien (z. B. Geschlecht), wohingegen diese Kategorien bei Ordinalskalen in einer Rangordnung sortiert sind (z. B. Schulnoten). Bei metrischen Skalen sind zudem die Abstände zwischen dieser Rangordnung klar definiert (z. B. Temperatur in Grad Celsius). Je nach Skalenniveau sind andere Maße zur Datenauswertung anwendbar und geeignet.