

Einführung: Wissenschaftliche Gütekriterien

Dieser Massive Open Online Course (MOOC) bietet allen Teilnehmerinnen und Teilnehmern eine Grundlage für die eigenständige Datenanalyse mit R und somit für seriöse Forschung. Dabei setzt *seriöse Forschung* die Berücksichtigung von Gütekriterien voraus, weist diese in der Regel an entsprechender Stelle aus und legt somit die *möglichen Schwachstellen* des Datensatzes, der Erhebung oder der Forscherinnen und Forscher offen, um diese bei der Interpretation der Ergebnisse angemessen berücksichtigen zu können. In den quantitativen Forschungsmethoden und im Machine Learning sind insbesondere die Objektivität, die Reliabilität sowie die Validität zu berücksichtigen, welche nachfolgend im Detail vorgestellt werden. Während sich die Objektivität und Reliabilität relativ prägnant darstellen lassen, sind insbesondere bei der Validität verschiedene Bezüge zu berücksichtigen, welche gleichermaßen herausstellen, dass valide Forschung als Voraussetzung zugleich immer auch objektiv und reliabel sein muss.

(1) Objektivität

Mit der Objektivität soll nachgewiesen werden, dass die beteiligten Forscherinnen und Forscher sowie die von diesen zur Anwendung gebrachten Erhebungsverfahren *unabhängig von den ausgewiesenen Ergebnissen* sind. Damit dies gelingt, müssen sowohl die verwendeten Fragestellungen als auch die verwendeten Hypothesen möglichst präzise und eindeutig aufgestellt werden, um mögliche Ungenauigkeiten in der Interpretation der Ergebnisse auszuschließen. Nur so kann sichergestellt werden, dass *unterschiedliche Forscherinnen und Forscher unter Anwendung entsprechend einschlägiger Erhebungsverfahren zu vergleichbaren Ergebnissen gelangen* und sich etwas gegebenenfalls als theoretische Grundlage bewähren kann.

Beispielsweise wäre Forschung alles andere als objektiv, wenn die beteiligten Forscherinnen und Forscher ein bestimmtes Ergebnis provozieren möchten und dafür entweder die angewendeten Erhebungsverfahren *manipulieren* oder aber den daraus resultierenden Datensatz sowie dessen Interpretation zur Herleitung dieses bestimmten Ergebnisses beeinflussen. In einem solchen Szenario ginge es nicht mehr um die Suche nach neuen Erkenntnissen, deren Dokumentation und Veröffentlichung, sondern vielmehr um eine bewusste Einflussnahme auf Dritte. Insbesondere wenn Forscherinnen und Forscher durch Dritte finanziert werden, gilt es stets zu hinterfragen, ob diese sich objektiv oder doch eher subjektiv mit einem realen Phänomen auseinandersetzen.

(2) Reliabilität

Die Reliabilität bezieht sich auf das Erhebungsinstrument der Forscherinnen und Forscher, welches im Rahmen des Erhebungsverfahrens zur Anwendung kommt. Dabei wird der Frage nachgegangen, ob das Erhebungsinstrument geeignet ist, ein reales Phänomen zuverlässig zu messen. Somit geht es bei der Reliabilität zum einen um die *Stabilität und Genauigkeit der Messung* bei wiederholtem Einsatz des Erhebungsinstruments und zum anderen aber immer auch um die *Rahmenbedingungen der Messung*

selbst. Viele reale Phänomene unterliegen nämlich weiteren und diese somit beeinflussenden Rahmenbedingungen, welche als weitere Variablen nicht nur die Beziehungen der relevanten Variablen untereinander zu beeinflussen vermögen, sondern gegebenenfalls auch die tatsächliche Merkmalsausprägung einer Variable selbst. Dadurch, dass bei der Reliabilität das Erhebungsinstrument im Fokus der Betrachtung liegt, bezieht diese sich immer auch auf die Durchführung der Erhebung, den daraus resultierenden Datensatz, deren Auswertung und Interpretation. Durch Beachtung der Reliabilität eines Erhebungsinstrumentes kann insgesamt die Reproduzierbarkeit der Forschung im Hochschulstudium und im wissenschaftlichen Kontext sichergestellt und ausgewiesen werden.

(3) Validität

Bei der Validität rückt die *Präzision* des Erhebungsinstruments in den Fokus. Dabei geht es darum, ob das Erhebungsinstrument auch tatsächlich das zur Forschungsfrage und den Hypothesen dazugehörige reale Phänomen misst, oder möglicherweise ein ganz anderes und gegebenenfalls ebenfalls reales Phänomen, welches nur ähnlich in Erscheinung tritt. Einige der gängigsten *Unterarten von Validität* sind die Konstruktvalidität, die Kriteriumsvalidität und die prognostische Validität.

Bei der *Konstruktvalidität* ist ein Detailblick auf die Frage *womit ich messe* geboten, denn bei vielen realen Phänomenen handelt es sich um sogenannte Konstrukte, die sich aus mehreren latenten Variablen zusammensetzen können. Das Erhebungsinstrument muss schließlich in der Lage sein, dieses Konstrukt, obschon der mit den latenten Variablen einhergehenden Herausforderungen, möglichst präzise messen zu können. Anders - aber keineswegs ausschließend - verhält es sich bei der *Kriteriumsvalidität*, welche die Frage nach dem *was ich messe* in den Vordergrund stellt. Um zu überprüfen, ob das Erhebungsinstrument auch tatsächlich das in Frage kommende reale Phänomen misst, können die daraus hervorgehenden Variablen im Anschluss an das Erhebungsverfahren auf ihren Zusammenhang mit weiteren und bereits bewährten Instrumenten und den daraus hervorgehenden Variablen überprüft werden. Dieses Vorgehen wird im Kontext der Forschungsmethoden oftmals auch als Übereinstimmungsvalidität bezeichnet. Wenn die Kriteriumsvalidität gegeben ist, dann kann dem Erhebungsinstrument unter Voraussetzung der Reliabilität ebenfalls eine *prognostische Validität* unterstellt werden. Bei gegebener prognostischer Validität sind die über das Erhebungsinstrument ermittelten Variablen bis zu einem gewissen Grad in der Lage, eine Vorhersage für nicht mit dem Erhebungszeitpunkt übereinstimmende reale Phänomene zu treffen. Dies ist beispielsweise dann der Fall, wenn im Rahmen eines Assessment Centers die zukünftige berufliche Performance der Bewerberinnen und Bewerber über ein einschlägiges Erhebungsinstrument antizipiert werden soll.

Eine weitere und ganz zentrale Unterscheidung im Kontext der Validität bezieht sich auf die Rahmenbedingungen des Erhebungsverfahrens. Wenn die Rahmenbedingungen von den beteiligten Forscherinnen und Forschern kontrolliert werden können, spricht man von *interner Validität*. Diese ist beispielsweise unter Laborbedingungen oder im Rahmen von experimentellen Studien außerhalb des

gewöhnlichen Umfeldes der Teilnehmerinnen und Teilnehmer - also innerhalb eines kontrollierbaren Umfeldes - gegeben. Interne Validität ist insbesondere dann vorhanden, wenn alle über die Hypothesen herausgestellten relevanten Variablen ohne Einflussnahme weiterer externer Variablen gemessen werden können. So werden beispielsweise Medikamente im Rahmen klinischer Studien unter eben diesen kontrollierten Rahmenbedingungen untersucht, um deren Wirksamkeit zu belegen. Dabei ist jedoch zu berücksichtigen, dass sich Menschen außerhalb dieser kontrollierten Rahmenbedingungen und in ihrem gewöhnlichen Umfeld anders verhalten könnten, so dass stets weitere Variablen zu berücksichtigen sind. Dadurch könnten beispielsweise die Teilnehmerinnen und Teilnehmer unter Vorlage der internen Validität im Rahmen einer klinischen Studie vollständig genesen, während es durchaus sein kann, dass einige Personen im tatsächlichen Anwendungsfall des Medikamentes tatsächlich nicht genesen werden, weil sie möglicherweise nicht mit dem Medikament verträgliche Lebensmittel zu sich genommen haben. Die Einnahme, der nicht verträglichen Lebensmittel, wäre unter kontrollierten Rahmenbedingungen von den Forscherinnen und Forschern unterbunden worden. Unter Vorlage des gewöhnlichen Umfeldes geschieht dies jedoch in der Regel nicht. Wenn also weitere nicht kontrollierbare Variablen zu berücksichtigen sind, spricht man im Kontext der Forschungsmethoden von *externer Validität*. In der Regel bezieht sich die externe Validität auf ein realistischeres Abbild realer Phänomene, ist jedoch in Folge dessen auch schwerer nachzuweisen.