

Das Verfassen einer empirischen Arbeit

In diesem Artikel finden Sie diverse Hinweise zur Erstellung von empirischen Arbeiten. Damit meinen wir in der Regel Abschlussarbeiten und Forschungsartikel, die eine statistische Auswertung bzw. Anwendung beinhalten. Im Folgenden präsentieren wir Richtlinien zum Verfassen von empirischen Arbeiten, die Sie zum verantwortungsvollen Umgang mit statistischen Ergebnissen anregen soll. Außerdem finden Sie eine *Gliederung* für den empirischen Teil Ihrer Arbeit.

Inhaltsverzeichnis

- [Richtlinien](#)
 - [Reproduzierbarkeit](#)
 - [Nachvollziehbarkeit](#)
 - [Ehrlichkeit](#)
 - [Fokussierung](#)
- [Ziel und Gliederung einer empirischen Arbeit](#)
 - [Methoden](#)
 - [Beschreibung der Daten](#)
 - [Ergebnisteil](#)
 - [Schlussfolgerungen](#)

 fu:stat bietet regelmäßig Schulungen für [Hochschulangehörige](#), sowie für [Unternehmen und weitere Institutionen](#) an. Die Inhalte reichen von Statistikgrundlagen (Deskriptive, Testen, Schätzen, lineare Regression) bis zu Methoden für Big Data. Es werden außerdem Kurse zu verschiedenen Software-Paketen gegeben. Auf Anfrage können wir auch gerne individuelle [Inhouse-Schulungen](#) bei Ihnen anbieten.

Richtlinien

Während Sie Ihre Arbeit verfassen, treffen Sie diverse Entscheidungen darüber, welche Ergebnisse Sie weitergeben. Im Folgenden möchten wir einige Richtlinien präsentieren, um entscheiden zu können, über welche Ergebnisse Sie schreiben, und welche evtl. weniger wichtig sind.

Folgende Frage sollten Sie sich stellen, um das Minimum an Information zu ermitteln:

 Unbekanntes Makro: 'vot'

Reproduzierbarkeit

Stellen Sie sich vor, Sie veröffentlichen Ihre Daten und Ihren Artikel; ist es möglich, alleine mit dieser Information Ihre Ergebnisse zu reproduzieren?

Das bedeutet, alle relevanten Schritte Ihrer Analyse und auch der Datenaufbereitung müssen transparent dargestellt werden. Sie sollten sich damit auf die eindeutige Benennung von statistischen Verfahren einerseits und auf die eindeutige Benennung von Schritten während der Datenaufbereitung andererseits konzentrieren. Dies beinhaltet nicht unbedingt Details der verwendeten Software, obwohl dies hilfreich sein kann, gerade bei spezielleren statistischen Verfahren. Bei den meisten Zeitschriften ist es möglich, Ihr Auswertungsskript (Programm) und Ihre Daten als Zusatzmaterial (Internet Material) zu veröffentlichen. Auch ohne dieses Auswertungsskript sollte es möglich sein, Ihre Datenanalyse nachvollziehen und reproduzieren zu können. Sollten Sie während des Review-Prozesses oder nach der Veröffentlichung oder nach Details im Auswertungsskript gefragt werden, sollten Sie dieses parat haben und auch weitergeben.

Nachvollziehbarkeit

Stellen Sie sicher, dass die Wahl und Präsentation von statistischen Methoden eindeutig und Ihre Schlussfolgerungen nachvollziehbar sind.

Wenn Sie statistische Verfahren anwenden, müssen Sie diese eindeutig benennen. Geben Sie dazu die Quelle an, in der die Methode vorgestellt wurde, oder aber die Quelle, in der eine Methode in Ihrer Forschungsdisziplin zum ersten Mal Anwendung gefunden hat. Wenn eine Methode unüblich in Ihrer Disziplin ist, dann sollten Sie einige Zeit darauf verwenden, in diese einzuführen und zu verdeutlichen, weshalb der existierende 'Goldstandard' innerhalb Ihrer Disziplin nicht ausreichend ist. Viele statistische Verfahren sind mit diversen Annahmen verbunden und die Überprüfung dieser Annahmen kann viel Platz in Anspruch nehmen. Oft ist es ausreichend, im Text über die Eignung einer Methode zu schreiben und das Ergebnis Ihrer Überprüfung darzulegen. Gerade, wenn Sie Probleme dabei feststellen, kann es aber hilfreich sein, im Anhang oder auch in der Arbeit selbst (je nach Fokus und Platz) zusätzliche Information zu liefern. Innerhalb dieser Diskussion sollte deutlich werden, dass Sie den Daten und der Fragestellung entsprechend ein geeignetes Verfahren gewählt haben.

Sichern Sie Ihre Daten!

Häufig wollen andere Wissenschaftler Ihre Ergebnisse replizieren oder mit anderen Methoden analysieren. Wenn ihre Studie öffentlich finanziert wurde und der Datenschutz dies zulässt, haben sie in der Regel die Verpflichtung, anderen Wissenschaftler Zugang zu Ihren Daten zu gewähren. Dazu müssen Sie Ihre Daten sichern. Dies geschieht am besten in einem Datenarchiv, das auch noch existiert, wenn Ihr Vertrag an der Universität schon längst abgelaufen ist. Zu einer ordentlichen Datensicherung gehört natürlich auch eine Beschreibung der Variablen. Dies gilt insbesondere für Variablen, die Sie selber generiert haben. Die FU verfügt über ein eigenes Datenarchiv, das [RE|FUB|IUM](#), in dem Sie neben Ihren Daten auch Ihre Arbeit archivieren können.

Das folgende Video zeigt wie eine junge Wissenschaftlerin ihren Kollegen um einen Datensatz bittet und was das Unternehmen endet. Hier ist der [Link zum Video](#)!

Ehrlichkeit

'Nicht signifikant' ist auch ein Ergebnis!

Aus Sicht der Forschungsdisziplin Statistik ist es weder hilfreich noch empfehlenswert sich an Konzepten wie *statistischer Signifikanz* festzuhalten. In der Forschungspraxis führt dieses zu einem Publikations-Bias. Das heißt, es werden überwiegend nur *signifikante* Ergebnisse publiziert, *nicht signifikante* Ergebnisse werden hingegen verschwiegen. Dadurch ergibt sich bei der Meta-Betrachtung von Forschungsfeldern oft ein verzerrtes Bild: Zu viele fälschlicherweise als signifikant erkannte Effekte. Sollten Sie sich also in der Situation wiederfinden, dass Ihre Auswertung nicht in dem gewünschten Maße Ihre Forschungshypothese unterstützt, versuchen Sie dennoch, Ihre Ergebnisse zu publizieren. Sie liefern damit einen Mehrwert für Ihre Disziplin.

Zudem werden durch die Suche (vielleicht auch Jagd) nach statistisch signifikanten Resultaten statistische Tests zweckentfremdet. Statistische Tests können einen Mehrwert bei der Überprüfung von Hypothesen liefern. Sie sind aber nahezu unbrauchbar, wenn sie dazu genutzt werden, Theorien abzuleiten. Wenn Sie durch Zufall oder unerwartet Ergebnisse finden sollten, tun Sie nicht so, als ob dies immer Ihr Ziel gewesen sei. Stellen Sie dies klar. Sie können diese Ergebnisse dann publizieren und zeigen damit möglicherweise interessante Forschungsfragen für die Zukunft auf.

Das folgende Video zeigt Ihnen einen Dialog zwischen einem frustrierten Wissenschaftler und einer Statistikerin über nicht-signifikante Ergebnisse: Hier der [Link zum Video](#).

Beachten Sie folgende Beiträge zum Thema:

Trafimow, D. & M. Marks (2015): Editorial; in: Basic and Applied Social Psychology, 37:1, 1-2, DOI: [10.1080/01973533.2015.1012991](#)
Wasserstein R. L. & N. A. Lazar (2016): The ASA's statement on p-values: context, process, and purpose; in: The American Statistician; DOI: [10.1080/00031305.2016.1154108](#)

Fokussierung

Achten Sie darauf, dass Ihre Ergebnisse eindeutig auf die Beantwortung Ihrer Forschungsfragen abzielen.

Ob Sie eine konkrete Analyse durchführen oder nicht, sollte sich immer der Frage unterordnen, ob sie helfen kann, Ihre Forschungsfrage zu beantworten. Ergebnisse können dabei sowohl für Ihre Forschungshypothese sprechen als auch dagegen. Oft ist es die Diskussion und Gegenüberstellung verschiedener Ergebnisse, die den größten Erkenntnisgewinn liefern. Behalten Sie dabei im Kopf, dass Ihre Leserschaft nicht unbedingt direkten Zugriff auf Ihre Daten hat, es ist also neben der 'finalen' Analyse auch Ihre Aufgabe, Ihre Daten zu beschreiben (deskriptive Statistik und Datenerhebung).

Ziel und Gliederung einer empirischen Arbeit

Wenn Sie beginnen, den empirischen Teil einer Arbeit zu schreiben, dann sollte der Forschungsprozess schon abgeschlossen sein. Damit ist dieser Teil ein Bericht bzw. eine Zusammenfassung über den Forschungsprozess von der Datengewinnung hin zur Analyse und deren Ergebnisse.

Neben der Aufgabe einer Zusammenfassung sollten Sie dabei folgendes beachten: Eine empirische Arbeit besteht immer aus

1. einem Methodenteil
2. einer Beschreibung der Daten
3. einem Ergebnisteil
4. und Ihren Schlussfolgerungen.

Ihre Aufgabe ist es, die verwendeten Methoden, Daten, und Ergebnisse (1-3) so darzustellen, dass Ihre Schlussfolgerungen (4) nachvollziehbar sind. Je nachdem, welche Schwerpunkte Ihre Arbeit setzt, können diese Punkte unterschiedlich aufwändig sein.

Preece, D.A. (1987): Good Statistical Practice; in Journal of the Royal Statistical Society. Series D (The Statistician), 36:4, 397-408; [Link](#)

Methoden

Wenn Sie eine neue statistische Methode verwenden, oder ein Verfahren verwenden, welches in Ihrer Disziplin noch keine Anwendung gefunden hat, dann ist es Ihre Aufgabe, diese Methode vorzustellen. Gängige statistische Verfahren - zum Beispiel deskriptive Statistik oder Testverfahren - werden oft nicht formal eingeführt. Was *gängige* Verfahren genau beinhaltet, hängt dabei von Ihrer Disziplin ab. Stellen Sie sicher, dass eine informierte Leserin Ihrer Disziplin Ihrer Auswertung folgen kann. Wenn es hilfreich für das Verständnis ist, eine Methode vorzustellen (zum Beispiel weil sie wenig verwendet wird), dann sollten Sie dies tun. Benennen Sie die Quelle, auf die das Verfahren zurückzuführen ist. Bei etablierten Methoden kann es ausreichen, auf Grundlagenwerke zu verweisen.

Sollten Sie die Datenerhebung selbst durchführen, dann ist es auch angebracht, bereits im Rahmen eines Methodenteils das Vorgehen zu beschreiben. Bei eigenen Umfragen kann dies Details zum [Fragebogen](#) und [Stichprobenziehung](#) enthalten. Wenn Sie ein [kontrolliertes Experiment](#) durchführen, zum Beispiel im Rahmen einer Fall-Kontroll-Studie, dann sollten Sie den Versuchsaufbau und das Vorgehen entsprechend beschreiben. Dabei kann es für das Verständnis und Einordnung der Ergebnisse sehr hilfreich sein, auf Probleme bei der Datengewinnung hinzuweisen.

Beschreibung der Daten

Jede statistische Auswertung ist auch mit einer [Beschreibung der zugrundeliegenden Daten](#) verbunden. Ihre Aufgabe besteht darin, Ihrer Leserschaft ausreichend Information bereit zu stellen, um die Datengrundlage bewerten zu können. Dies beinhaltet in jedem Falle die Anzahl der Messungen (Beobachtungen) und Merkmale (Variablen), die Sie interessieren bzw. die Sie untersucht haben. Der Detailgrad hängt -- wie auch bei den Methoden -- davon ab, wie geläufig Ihre Datenquelle bzw. Datengewinnung innerhalb Ihrer Disziplin ist. In diesem Rahmen ist es angebracht, folgende Punkte zu diskutieren:

1. Welche Probleme sind bei der Datengewinnung ([Umfrage](#), [eigene Studie](#), ...) aufgetreten: Messfehler, [fehlende Werte](#), Fehler im [Fragebogen](#), etc.
2. Spiegelt Ihre [Stichprobe](#) ein realistisches Bild der Zielpopulation wieder? Viele Datenanbieter stellen zum Beispiel Gewichtungsfaktoren bereit, wenn die Stichprobe nicht repräsentativ für die Bevölkerung sein sollte. Verwenden Sie eigene Datenquellen, ist es Ihre Aufgabe, ein realistisches Bild zu zeichnen und eine Einordnung zu geben. Dieser Punkt hat maßgeblichen Einfluss darauf, wie Ihre Analyseergebnisse zu bewerten sind.
3. Welche Schritte von den Rohdaten bis hin zum analysierten Datensatz haben Sie durchgeführt? Dies kann zum Beispiel die Einschränkung der ursprünglichen Stichprobe beinhalten, um sie nach (2) auf die interessierende Menge zu reduzieren; allerdings sollten Sie auch beschreiben, wie Sie mit *Ausreißern* und *Messfehlern* umgehen.

Wenn Sie sekundäre Datenquellen verwenden, dann werden Ihnen zu Punkt (1) und (2) zum Teil hilfreiche Informationen bereitgestellt. Sie sollten immer den Herausgeber der Daten zitieren. Wie dies geschehen soll, wird Ihnen im Regelfall vom Datenanbieter mitgeteilt.

Ergebnisteil

Nach der Beschreibung Ihrer Datengrundlage präsentieren Sie die statistische Analyse. Dabei kann es sehr hilfreich sein, zunächst die Analyseergebnisse zu beschreiben (3) und sie erst später (4) zu interpretieren. Die Beschreibung beinhaltet:

1. die (eindeutige) Benennung der verwendeten Methoden
2. eine Motivation für das verwendete Analyseverfahren
3. die Beurteilung der Güte.

Achten Sie bei der Benennung der Verfahren darauf, dass es unabhängig von einer [Statistiksoftware](#) transparent ist, welche Methode Sie verwenden.

Die Motivation für statistische Verfahren lässt sich grob aus zwei Blickrichtungen sehen: entweder Sie folgen dem für eine Fragestellung in Ihrer Disziplin etabliertem Vorgehen, oder Sie halten etablierte Verfahren für unangebracht und präsentieren ein aus Ihrer Sicht verbessertes Vorgehen. In beiden Fällen beziehen Sie sich auf bestehende Forschung und bauen auf bestehendem Wissen auf oder entwickeln es weiter. Dabei kann es vorkommen, dass Sie zunächst mit etablierten Methoden beginnen und an Grenzen stoßen und deshalb auf andere Verfahren zurückgreifen müssen. Dies sollten Sie diskutieren, da es ein wesentliches und interessantes Ergebnis Ihrer Arbeit ist. Ergänzend dazu kann die [Wahl eines Auswertungsverfahrens](#) zunächst allein auf Basis der Eigenschaften der Daten erfolgen. Sie sollten jedoch darauf achten, Ihren Forschungskontext nicht aus den Augen zu verlieren, da oft typische Probleme mit *offensichtlichen* Verfahren bereits in der Literatur diskutiert wurden.

Die [Beurteilung der Güte](#) hängt stark von der verwendeten Methode ab. Sie sollten - soweit dies möglich ist - alle Annahmen, unter dem das verwendete statistische Verfahren formuliert ist, überprüfen. Wenn Sie dabei feststellen, dass das Verfahren nicht optimal für die gegebenen Daten ist, dann kann verwandte Literatur sehr hilfreich sein. Gerade in diesen Fällen bietet es sich allerdings oft an, eine [statistische Beratung in Anspruch zu nehmen](#). Bei der Kommunikation der Modellüberprüfung reicht es oft aus, das Ergebnis dieser Überprüfung verbal zusammenzufassen. Anders ist dies, wenn Sie kein Verfahren finden, das *perfekt* zu Ihren Daten passt. Gerade dann ist es wichtig, die Einschränkungen zu diskutieren, da diese maßgeblich die Aussagekraft der Analyse beeinflussen können. In diesem Zusammenhang sollten Sie sich aber immer den Aphorismus

Wie können Daten zitiert werden:

Hinweise für das Sozioökonomische Panel (SOEP) bereitgestellt vom Deutschen Institut für Wirtschaftsforschung (DIW): [FAQ 3.3](#)

Hinweise der [GESIS](#): [Bibliographisches Zitieren von Forschungsdaten und Dokumenten einer Studie](#)

Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful. – George Box

ins Bewusstsein rufen. Anders als oftmals vermutet, ist ein geeignetes statistisches Verfahren nicht *richtig* oder *falsch*, sondern vielmehr stellen wir uns die Frage: Wie hilfreich oder nützlich ist es? Abweichungen von Modellannahmen sind dabei die Regel. Wenn Sie selbst unsicher sind, wie problematisch die Anwendung in Ihrem konkreten Fall ist, präsentieren Sie die Ergebnisse Forschern mit Erfahrung in der Anwendung des spezifische Auswertungsverfahrens. Aus Ihrem Artikel oder Ihrer Monographie muss hervorgehen, wie Sie das verwendete Verfahren einschätzen.

Schlussfolgerungen

Nachdem Sie nun die verwendeten Methoden, Daten und Ihre Ergebnisse beschrieben haben, können Sie Ihre Schlussfolgerungen daraus ziehen. Gehen Sie dabei wie folgt vor:

1. Wie beantworten Sie Ihre Forschungsfrage anhand der Ergebnisse? Ignorieren Sie zunächst alle Einschränkungen und interpretieren Sie die Ergebnisse.
2. In einem zweiten Schritt ist es dann wichtig, die Ergebnisse der Datenbeschreibung und Modellüberprüfung kurz zusammenzufassen, um dann zu einer Gesamteinschätzung der Aussagekraft Ihrer Ergebnisse zu kommen. Hierbei interpretieren Sie die Ergebnisse nicht neu, sondern geben eine Einschätzung, für wie belastbar bzw. zuverlässig Sie diese halten.
3. In einem finalen Schritt sollten Sie diskutieren, inwiefern Sie vorhandene Forschungsergebnisse (a) bestätigen / reproduzieren konnten, (b) widerlegt haben, bzw. widersprechen und (c) weiterentwickelt, bzw. neue Perspektiven in die Diskussion eingebracht haben.

Bei allen Hinweisen und Ratschlägen, die wir ihnen geben können, ist die Anwendung der Statistik so vielseitig, dass nicht unbedingt alle Punkte in diesem Artikel auf Ihre Auswertung zutreffen mögen. Vergessen Sie deshalb nicht:

Learn the rules, then break them!