

Als-ob-Statistik

Stefan Huschens

Fassung vom 29. Februar 2020

Zusammenfassung

Das Konzept der Als-ob-Statistik wird als Bezeichnung bestimmter missbräuchlicher Anwendungen statistischer Methodik eingeführt. Diese Anwendungen sind dadurch gekennzeichnet, dass – überwiegend im sozialwissenschaftlichen Bereich – statistische Methoden formal so verwendet werden, **als ob** Voraussetzungen der klassischen Statistik, wie z. B. metrische oder ordinale Skalierung der betrachteten Variablen oder Vorliegen einer Zufallsstichprobe, erfüllt sind.¹

1 Klassische Statistik

Der traditionelle Gegenstand der Statistik ist die Analyse beobachtbarer Variablen. Diese Variablen entsprechen messbaren (im Sinn der Längenmessung, Gewichtsmessung, Zählung usw.) Merkmalen statistischer Einheiten. Dieser Bereich der statistischen Methodik entspricht der wirtschaftsstatistischen, biometrischen, ökonometrischen und naturwissenschaftlichen Tradition der Statistik. Mindestens 95% des Methodenspektrums der Statistik dürften diesem Bereich zuzuordnen sein.

Das typische Vorgehen bei inferenzstatistischen Verfahren (statistische Schätz- und Testverfahren) ist die Erstellung eines statistischen Modells, das aus einer Menge zugelassener Wahrscheinlichkeitsverteilungen besteht, die mit den beobachteten Variablenwerten verträglich sein könnten. Die beobachteten Variablenwerte (Daten) sind dabei das Ergebnis einer Zufallsstichprobe. Es ist nicht notwendig vorausgesetzt, dass dem Zufallsmechanismus, der die Daten generiert, exakt eine der durch das statistische Modell zugelassenen Wahrscheinlichkeitsverteilungen zugrunde liegt. Dies wird schon durch den Begriff „Modell“ deutlich. Manchmal vielleicht hilfreich, aber grundsätzlich problematisch und irreführend, sind daher Vorstellungen und Sprechweisen vom „wahren Modell“, vom „wahren Parameter“ usw.

Mithilfe statistischer Schätzverfahren wird eine der durch das Modell zugelassenen Wahrscheinlichkeitsverteilungen, häufig über die Schätzung eines Parameters, als die am besten mit den Daten verträgliche Verteilung innerhalb des statistischen Modells identifiziert. Mithilfe statistischer Testverfahren werden im Rahmen des vorgegebenen statistischen Modells Hypothesen über in der Regel zwei Teilmengen von Wahrscheinlichkeitsverteilungen getestet. Im einfachsten Fall enthält die Nullhypothese genau eine Wahrscheinlichkeitsverteilung oder einen Parameterwert, dessen Verträglichkeit mit den beobachteten Variablenwerten überprüft wird.

¹Dem Autor ist bewusst, dass im durch „als ob“ eingeleiteten Nebensatz grammatikgerecht der Konjunktiv II, seltener der Konjunktiv I, steht. Dagegen wird umgangssprachlich und in der gesprochenen Sprache zunehmend der Indikativ verwendet. Die Wahl des Indikativ erfolgt in diesem Aufsatz bewusst.

Statistische Inferenzverfahren sind immer relativ zum verwendeten und postulierten statistischen Modell zu sehen. Bei statistischen Testverfahren tritt daher neben die üblichen Fehler erster und zweiter Art ein Fehler dritter Art, der als Modellfehler dann auftritt, wenn der gewählte Modellrahmen des Testverfahrens zu eng ist.

2 Als-ob-Statistik

In mehreren Anwendungsbereichen der Statistik gibt es methodische Vorgehensweisen mit einer oft problematischen Verwendung inferenzstatistischer Verfahren und gewagten Als-ob-Unterstellungen. Im folgenden werden drei Problemkreise angesprochen:

1. Variablen mit vagem oder undefiniertem Messniveau: dabei werden Meinungsdaten und Meinungsäußerungen von Personen werden behandelt, **als ob** es sich um Messungen eines ordinalen Merkmals an verschiedenen statistischen Einheiten handelt. Daten, die im günstigsten Fall als Messungen ordinaler Variablen aufgefasst werden könnten, bearbeitet, **als ob** es sich um metrische Variablen handelt.
2. Inferenzstatistik ohne stochastische Grundlage: dabei werden statistische Inferenzverfahren auf Daten angewendet, **als ob** es sich um die Ergebnisse einer Zufallsstichprobe im Sinn der Statistik handelt. Daten aus methodisch verzerrten Stichproben werden – in der Regel mit Verweis auf die angebliche Repräsentativität – behandelt, **als ob** es sich um Ergebnisse einer Zufallsstichprobe im Sinn der klassischen Statistik handelt und statistischen Inferenzverfahren unterworfen.
3. Modelle mit unbeobachtbaren Konstruktvariablen und nichtidentifizierbarer Parameterstruktur, bei der eine willkürliche, an Daten angepasste Parameterstruktur so interpretiert wird, **als ob** es sich um ein statistisches Inferenzverfahren handelt.

2.1 Variablen mit vagem oder undefiniertem Messniveau

Variablen mit vagem oder undefiniertem Messniveau sind häufig das Ergebnis von Befragungen, bei denen Personen auf meist fünf- oder siebenstufigen Skalen Meinungen, Einstellungen, Befindlichkeiten usw. spezifizieren. Sowohl das Zustandekommen der befragten Einheiten als auch die so erhobenen Daten schließen in der Regel die Anwendung der üblichen inferenzstatistischen Methoden aus.

Ein Beispiel soll die Problematik verdeutlichen. Nehmen wir an, das Merkmal Körpergröße würde in einer Befragung erhoben, indem die Befragten auf einer fünfstufigen Skala „sehr klein / klein / mittel / groß / sehr groß“ ihre Körpergröße angeben, wobei die Skala mit „sehr klein = 155 cm und weniger“, „klein = 156 bis 165 cm“, „mittel = 166 bis 175 cm“, „groß = 176 bis 185 cm“, „sehr groß = mehr als 185 cm“ präzisiert. In diesem Fall ergibt sich zwar ein Informationsverlust durch den Übergang von einem metrischen zu einem ordinalen Messniveau, aber bestimmte Rückschlüsse aus einer Stichprobe auf die Grundgesamtheit wären noch möglich. Eine Kodierung der fünf Skalenwerte mit den Zahlen 1 bis 5 bildet den Informationsgehalt der ordinalen Messung ab, so lange nur die durch $<$ -Relation induzierte Ordnung verwendet wird. Jede streng monoton wachsende Transformation ist informationserhaltend.

Als zweiten Fall stelle man sich vor, die Erläuterungen der Skala würden weggelassen und jeder Befragte habe sich selbst auf der Skala „sehr klein / klein / mittel / groß / sehr

groß“ einzuschätzen. Dies bedeutet erstens, dass das Merkmal Körpergröße nur noch ordinal gemessen wird, zweitens aber auch, dass jedes Individuum seine eigene Ordinalskala hat oder haben kann. Es ist nicht klar, ob das, was Person 1 als „groß“ einstuft, nicht von Individuum 2 als „sehr groß“ oder als „mittel“ eingestuft wird usw. Nach einer Kodierung der fünf Stufen mit den Zahlen 1 bis 5 liegen zwar formal beobachtete Werte einer variablen Größe vor, aber es liegen noch nicht einmal Werte einer Variablen mit ordinalem Messniveau im Sinn der statistischen Methodik vor. Vielmehr sind die verwendeten Skalen subjektiv und individuenpezifisch, so dass mit dem Zusammenwerfen solcher Werte nicht das Messniveau eine Ordinalskala erreicht wird.

Die Problematik der individuenpezifischen und nicht objektivierbaren Messskalen verschärft sich, wenn anders als im obigen Beispiel der Körpergröße überhaupt kein zugrundeliegendes objektiv messbares Merkmal existiert, sondern nur die Stärke einer Meinung oder Befindlichkeit von Personen ausgedrückt werden soll, so dass noch eine zusätzliche Unschärfe bezüglich des Gegenstandes der Messung tritt. Wenn im Bereich der empirischen Sozialforschung auf solche Daten Verfahren angewendet werden, die z. B. im Bereich der Biometrie für metrisch skalierte Experimentaldaten entwickelt wurden, dann entsteht nur etwas wenig mehr als Unsinn. Dieses „wenig mehr“ ist aber auch in wiederholten Studien nicht quantifizierbar. Es kann zufällig sein, dass im Beispiel der Körpergrößenmessung die Befragten unter den vorgegebenen Kategorien etwa dasselbe verstehen und es kann sein, dass eine Methode, die für metrische Daten bestimmt ist, auch bei ordinalen Daten Ergebnisse liefert, die in die richtige Richtung zeigen. Der Hinweis darauf, dass dieses Vorgehen in bestimmten Anwendungsbereichen weitgehend üblich ist und dort als unkritisch angesehen wird, ist nicht beruhigend, im Gegenteil. Anstatt sich der mühsamen Forschungsaufgabe zu stellen, für die vorliegenden schwachen Messniveaus geeignete Methoden zu entwickeln, werden Methoden angewendet, die für völlig andere Datenkonstellationen entwickelt wurden und geeignet sind.

Häufig sind die folgenden drei Schritte in die Vagheit des Als-ob festzustellen.

1. Subjektiv-individuenpezifische ordinale Messungen werden so behandelt, **als ob** objektive Messungen eines ordinalen Merkmals vorliegen.
2. Wenn ein ordinales Merkmal durch Zahlen, z. B. durch die Zahlen 1 bis 5, codiert ist, wird es so behandelt, **als ob** Messungen eines metrischen Merkmals vorliegen.
3. Wenn ein metrisches Merkmal mit den möglichen Werten 1 bis 5 vorliegt, wird es so behandelt, **als ob** Werte einer normalverteilten Zufallsvariable vorliegen.

2.2 Inferenzstatistik ohne stochastische Grundlage

Inferenzstatistische Methoden beruhen auf stochastischen Modellen, bei denen der Zufallsmechanismus entweder auf einem wohldefinierten Ziehungsschema für Stichproben aus Grundgesamtheiten oder auf künstlicher Randomisierung einer Gruppe von Untersuchungseinheiten beruht. Dagegen werden in bestimmten Anwendungsbereichen inferenzstatistische Methoden auch dann verwendet, wenn das Paradigma der Zufallsstichprobe aus einer Grundgesamtheit nicht zutreffend ist und eigentlich nur ein mehr oder weniger willkürlich zustande gekommene Menge von statistischen Einheiten vorliegt. Teilweise werden auf dem Stichprobenkonzept beruhende inferenzstatistische Verfahren sogar auf Grundgesamtheiten angewendet.

Die Anwendung inferenzstatistischer Verfahren mit dem Ziel, Signifikanzsternchen zu produzieren, ist oft problematisch, da die unkritisch verwendeten inferenzstatistischen Methoden z. B. auf multivariat normalverteilten Variablen in der Grundgesamtheit und auf unabhängig und identisch verteilten Beobachtungen beruhen. Bei der Interpretation ist die Verwechslung von statistischer Signifikanz mit Relevanz eher der Normal- als der Ausnahmefall. Mit Hilfe geduldiger Software und gewagter Konstruktionen werden klassische Methoden der Inferenzstatistik, die in der Regel für Experimentaldaten entwickelt wurden, angewendet.

Als qualitative Statistik im Unterschied zur quantitativen Statistik können Ansätze charakterisiert werden, die ohne Bezugnahme auf konkrete Begriffe der wissenschaftlichen Statistik sehr viel über Statistik reden und sich dabei teilweise eine eigene vage Begriffswelt geschaffen haben. Typischerweise erfolgt kein Bezug auf konkrete statistische Maßzahlen, auf Definitionen oder auf Gleichungen. Das Reden über Statistik soll die mühsame Auseinandersetzung mit den formalen und mathematisch nicht ganz leichten Konzepten der Statistik ersetzen. Ein typisches Beispiel ist der unter Anwendern und denjenigen, die statistische Auswertungen verkaufen, überaus beliebte, blumige Begriff der Repräsentativität, der kein Fachbegriff der Statistik ist und in der quantitativen Statistik eine völlig unbedeutende Rolle spielt². Insbesondere gibt es keine Maßzahl für Repräsentativität.

2.3 Modelle mit unbeobachtbaren Konstruktvariablen und nicht-identifizierbaren Parametern

Eine unbeobachtbare Konstruktvariable beschreibt eine abstrakte Eigenschaft einer statistischen Einheit. Diese Eigenschaft entzieht sich der direkten Beobachtung und kann nur indirekt über Hilfsvariablen messbar gemacht werden kann. In der Psychologie führte die Frage nach der Messung der menschlichen Intelligenz und ihrer Komponenten zur Entwicklung des Methodenkomplexes der Faktorenanalyse. Ähnlich abstrakte Eigenschaften, die sich einer direkten Messung entziehen, sind Glück, Zufriedenheit und politische Grundeinstellung von Personen; Innovationsfreude, Korruptionsresistenz und Umweltfreundlichkeit von Unternehmen usw. In Teilen der empirischen Sozialforschung, in der Marktforschung, im Marketing und in einigen weiteren Bereichen der Betriebswirtschaftslehre interessieren Einstellungen von Personen oder abstrakte Eigenschaften von Institutionen der Fall und erfreuen sich Modelle mit unbeobachtbaren Konstruktvariablen und die dafür entwickelten Methoden (Faktorenanalyse, Strukturgleichungsmodelle usw.) großer Beliebtheit.

Die Konzepte der Faktorenanalyse basieren auf Modellen mit nichtidentifizierbaren Parameterstrukturen. Es gibt also im Rahmen des postulierten Modells grundsätzlich unendlich viele Parameterwerte, die beobachtungsäquivalent sind, also dieselbe Wahrscheinlichkeitsverteilung für die beobachtbaren Variablen implizieren. Dadurch ist es prinzipiell unmöglich, von den beobachteten Werten der beobachtbaren Variablen auf Parameter zurückzuschließen. Diese Nichtidentifizierbarkeit der Parameter ist nicht auf die endliche Anzahl der Beobachtungen zurückzuführen. Auch dann, wenn unendlich viele Beobachtungen vorlägen, könnte nicht zwischen beobachtungsäquivalenten Parametern unterschieden werden.

In der klassischen Statistik gilt die Nichtidentifizierbarkeit der Parameter eines Modells als schwerer Defekt eines Modells. Da die Unbestimmtheit der Parameter grundsätz-

²http://www.von-der-lippe.org/dokumente/Repräsentativität_4.pdf, 0.9 MB

lich nicht behebbar ist, ist das Modell für alle Zwecke ungeeignet, bei denen es auf die Interpretation einzelner Parameterwerte ankommt. Es gibt allerdings Verwendungszwecke eines Modells, bei der die Nichtidentifizierbarkeit von Parametern unschädlich sein kann. Wird beispielsweise ein lineares Regressionsmodell für Zwecke der Prognose eingesetzt, ohne dass die einzelnen Regressionsparameter interpretiert werden müssen, so kann die Nichtidentifizierbarkeit der Regressionsparameter für diesen Zweck unschädlich sein.

In bestimmten Anwendungsgebieten wird die Eigenschaft der nichtidentifizierbaren Parameterstruktur mit unendlich vielen beobachtungsäquivalenten Parameterwerten kreativ genutzt, um durch geeignete Manipulationen, die so genannte Faktorrotation, „schöne“ oder in irgendeinem Sinn zweckmäßige oder gut interpretierbare Parameter auszuwählen. Dann wird so getan, als seien die so willkürlich ausgewählten Parameterwerte das Ergebnis einer statistischen Inferenzprozedur. Ein anderer zu beobachtender Umgang mit der nichtidentifizierbaren Parameterstruktur faktorenanalytischer Modelle besteht darin, einfach die Voreinstellungen der Statistiksoftware zu verwenden, wodurch eine der unendlich vielen, denkbaren Faktorrotationen ausgeführt und eine berechnete Parameterstruktur ausgegeben wird.

Bei den ursprünglichen Entwicklungen der Faktorenanalyse im Bereich der Psychologie waren die beobachtbaren Variablen, aus deren Kovarianz- oder Korrelationsmatrix die unbeobachtbaren Faktoren erzeugt („extrahiert“) werden, intersubjektiv vergleichbare Messergebnisse von Teilaufgaben zur Intelligenzmessung. Dagegen werden in späteren Anwendungen im Bereich der empirischen Sozialforschung als beobachtbare Variablen vage Variablen im Sinn des ersten Abschnitts verwendet, zugleich aber die von der Software angebotenen inferenzstatistischen Maßzahlen genutzt, so dass sich ein Kombination der in den Abschnitten 2.1 und 2.2 beschriebenen **Als-ob**-Konstruktionen ergibt.