

Entgegnung zu Beitrag von Langeheine

Jagodzinski, Wolfgang

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Jagodzinski, W. (1987). Entgegnung zu Beitrag von Langeheine. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 21, 77-81. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-205290>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.



Entgegnung₂ zu Beitrag₂ von Langeheine
von Wolfgang Jagodzinski

Rolf LANGEHEINE scheint meine beiden Thesen im wesentlichen zu akzeptieren (1). Was die Anwendung von L^2 als Anpassungsmaß anbelangt (These 1), so ist mir LANGEHEINES Position allerdings nicht völlig klar. Grundsätzlich will er dem L^2 bei Separierbarkeit der Klassenprofile wohl trauen, aber eben doch nicht ganz, weil er jetzt zusätzlich PEARSONS X^2 berechnet. Dabei zeigt sich in der konkreten Analyse prompt, was häufig als Indiz für die Inadäquatheit der χ^2 - und L^2 -Teststatistik gewertet wird, daß nämlich das modifizierte Black & White-Modell M3 nach der einen Prüfgröße (PEARSONS X^2) mit den Daten verträglich ist, nach der anderen (L^2) aber nicht. Meine Ansicht zur Aussagekraft sowohl der Teststatistiken wie auch der Simulationsstudien habe ich in der Replik 1 darzulegen versucht und ich will mich insoweit nicht wiederholen.

Neu in unserer Diskussion und interessant ist LANGEHEINES Vorschlag, zum Vergleich der Modelle M3 und M4 nicht die einfachen L^2 -Werte, sondern deren Differenz heranzuziehen, die bei geschachtelten Modellen unter allgemeineren Voraussetzungen chiquadratverteilt ist als das einfache L^2 . Nun bedürften die Anwendungsvoraussetzungen dieses L^2 -Differenz-Tests sicherlich einer gründlicheren Analyse als sie an dieser Stelle möglich ist. Es muß hier genügen, auf drei grundsätzliche Probleme hinzuweisen. Zunächst einmal wird das weniger restriktive Modell meist nicht a priori spezifiziert, sondern nach mehr oder minder ausgedehnten explorativen Studien gefunden. Man modifiziert mit anderen Worten die Modellannahmen solange, bis man auf ein möglichst gut an die Daten angepaßtes Modell stößt. Wenn man anschließend das L^2 des restriktiven Ausgangsmodells mit dem L^2 des nachträglich angepaßten Modells vergleicht, so testet man nicht mehr im strengen Sinne, sondern demonstriert nur - um es etwas überspitzt zu formulieren - seine Talente als 'Modellbauer'. Die explorative Strategie birgt zudem die Gefahr in sich, datensatzspezifische, nicht replizierbare Modelle zu erzeugen.

Ist dies ein allgemeines Problem des Likelihood-Ratio-Tests für hierarchisch geschachtelte Modelle, so betreffen die weiteren Komplikationen speziell unsere Modelle M3 und M4. Sie sind, worauf LANGEHEINE selbst hinweist, genaugenommen nicht ineinander geschachtelt. Erklärt man nämlich in M3 die Antwortwahrscheinlichkeiten von Materialisten, Postmaterialisten und



Mischtypen in der zweiten und dritten Welle zu freien Parametern, so entsteht ein Modell M4', das neben den drei Werttypen auch noch eine Klasse von Meinungslosen vorsieht. Es ist ein glücklicher Zufall, daß für diese letzte Klasse in diesem Datensatz und auf diesem Aggregationsniveau ein Anteil von Null geschätzt wird, also die gleichen Parameterwerte ermittelt werden wie in dem von LANGEHEINE spezifizierten Modell M4².

Drittens führt Rolf LANGEHEINE den Test auf zu hohem Aggregationsniveau durch. Denn wie seine Ausführungen zu meiner These 2 zeigen, ist ja auch er der Auffassung, daß in M4 der Anteil der zeitstabilen Mischtypantworten erheblich überschätzt wird. Welchen Sinn hat es aber, die bessere Anpassung von M4 an die 3³-Tabelle nachzuweisen, wenn man genau weiß, daß dieses Modell in einem entscheidenden Punkt ganz und gar nicht zu den Daten paßt! Man muß also die Modelle M3" und M4" miteinander vergleichen, nicht aber M3' und M4'.

Nehmen wir an, auch in einem mit M3" geschachtelten Modell würde ein Anteil von null Meinungslosen geschätzt und die Koeffizienten von M4" stimmen exakt mit jenem wirklich geschachtelten Modell überein. Da die Chi-Quadratdifferenz zwischen M3" und M4" im Vergleich zur Differenz der Freiheitsgrade beträchtlich ist, spräche auf den ersten Blick alles für die Verwerfung des Black & White-Modells. Dennoch wäre es m.E. verkehrt, M4" jenem Modell vorzuziehen.

Der Grund ist Tabelle 1 zu entnehmen, wo die Schätzwerte für Antwortwahrscheinlichkeiten und Klassenanteile von Modell M4" berichtet sind³. Im Kopf der Tabelle stehen die Klassennamen, wobei die ersten beiden Klassen als Materialisten und Postmaterialisten bezeichnet werden, weil ihre Mitglieder, die entsprechenden Antwortkombinationen mit großer Wahrscheinlichkeit wählen. Um die Zuordnung auch bei den Mischtypen zu erleichtern, wurde die größte Antwortwahrscheinlichkeit zu jedem Zeitpunkt unterstrichen. Der erste Mischtyp (Mi1) favorisiert also schwach die Antwortkombinationen Ruhe und Ordnung und Partizipation, der zweite (Mi2) kombiniert relativ oft Ruhe und Ordnung mit Meinungsfreiheit usw. Die beiden ersten Mischtypklassen zusammen umfassen knapp 40% der Stichprobe, während die übrigen beiden nur ca. 15% der Befragten einschließen.

Theoretisch nicht zu erklären sind nun gerade die Antwortmuster der beiden dominanten Mischtypgruppen. Gewiß, Mi2 zeigt in der dritten Welle ein sehr

Tabelle 1: Antwortwahrscheinlichkeiten und Klassenanteile im Modell M4"
($L^2=78.70$; $df=166$)

Klasse	Mat.	PMat.	Mi ₁	Mi ₂	Mi ₃	Mi ₄
t₁:						
R+O,Part.	.02	.00	<u>.49</u>	.07	.25	.00
R+O,Infl	<u>.84</u>	.00	<u>.00</u>	.14	.16	.32
R+O,MFr	.12	.14	.08	<u>.65</u>	.00	.14
Part,Infl	.02	.03	.20	<u>.00</u>	<u>.42</u>	.00
Part,MFr	.00	<u>.83</u>	.20	.14	<u>.00</u>	.00
Infl,MFr	.00	<u>.00</u>	.03	.00	.17	<u>.54</u>
t₂:						
R+O,Part.	.05	.03	<u>.26</u>	.10	.00	.00
R+O,Infl	<u>.89</u>	.00	<u>.14</u>	.10	.16	.00
R+O,MFr	.03	.03	<u>.28</u>	<u>.76</u>	.10	.00
Part,Infl	.00	.00	.12	.00	<u>.74</u>	.00
Part,MFr	.00	<u>.86</u>	.17	.00	<u>.00</u>	.00
Infl,MFr	.03	.08	.03	.04	.00	<u>1.00</u>
t₃:						
R+O,Part	.10	.00	<u>.43</u>	.00	.00	.00
R+O,Infl	<u>.82</u>	.00	<u>.41</u>	.00	.04	.00
R+O,MFr	.00	.00	.00	<u>1.00</u>	.00	.00
Part,Infl	.00	.00	.00	.00	<u>.88</u>	.22
Part,MFr	.00	<u>.96</u>	.04	.00	<u>.08</u>	.15
Infl,MFr	.08	<u>.04</u>	.12	.00	.00	<u>.64</u>
Kl. anteile	.250	.214	.201	.191	.083	.061

einfaches Antwortmuster, doch wie soll man die Wahrscheinlichkeiten in den ersten beiden Befragungen deuten? Warum sollte bei dieser Gruppe der 'Sokrateseffekt' erst in der dritten Welle auftreten? Warum nimmt insbesondere die Neigung, mit Ruhe und Ordnung ein anderes Ziel als Meinungsfreiheit zu kombinieren, von der ersten zur zweiten Welle nicht deutlicher ab? Warum tendiert diese Gruppe andererseits nur in der ersten Welle schwach zur Kombination Part(izipation) und M(einungs)Fr(eiheit), nicht aber in den Wellen zwei und drei?

Noch erratischer jedoch ist das Antwortverhalten des ersten Mischtyps. Er läßt gewissermaßen keine Antwortkombination aus. Wenn er in der ersten Welle kaum eine Präferenz für die Kombinationen 2 (R+0, Infl) und 6 (Infl, MFr) hat, so entwickelt er in der dritten Welle gerade für diese Kombinationen eine besondere Vorliebe. In der zweiten Befragung weicht das Muster



von einem rein zufälligen Antwortverhalten (im Sinne des Auswürfelns) kaum noch ab, wobei die sonst dominante Kombination (R+O, Part) nicht einmal mit gleicher Wahrscheinlichkeit gewählt wird wie die Kombination (R+O, MFr). Vor allem hier hat man den Verdacht, daß die Schätzwerte weniger ein systematisches Antwortverhalten als vielmehr Eigenheiten des Datensatzes widerspiegeln. Vor die Wahl gestellt, ob man lieber ein etwas schlechter angepaßtes, dafür aber einfaches und interpretierbares Modell annehmen sollte oder aber ein besser angepaßtes mit uninterpretierbaren Koeffizienten, das aller Voraussicht doch nicht replizierbar ist, ziehe ich ganz klar die erste Alternative vor (4).

Sicher könnte man versuchen, durch zusätzliche Restriktionen jene Defekte zu korrigieren. Solche Maßnahmen würden nicht nur das eingangs erwähnte Problem verschärfen, sie könnten auch nicht über die schmale Datenbasis hinwegtäuschen, auf der die Schätzung der Modellparameter beruht. Ein Modell, bei dem aus 152 Beobachtungen, die auf 216 Zellen verteilt worden sind, ca. 50 Parameter geschätzt werden, ist wohl kaum mit inferenzstatistischen Argumenten zu verteidigen. So scheint es nicht ratsam, allen bislang schon präsentierten Modellen noch ein weiteres hinzuzufügen, das dann doch kein höheres Maß an empirischer Evidenz beanspruchen könnte als die Modelle M2-M4. Die Strategie der Modellanpassung wie vor allem auch das kleine N gestatten keinen harten statistischen Test. So muß es m.E. künftigen Analysen mit besseren Daten (und zusätzlichen Variablen) vorbehalten bleiben, eine Entscheidung zugunsten der einen oder anderen Deutung zu treffen.

Wenn wir uns auch nicht in allen Punkten einig sind, so habe ich die Kritik von Rolf LANGEHEINE doch als anregend und konstruktiv empfunden. Danken möchte ich auch dem Herausgeber der ZA-Information, der diese in meinen Augen sachliche und faire Diskussion ermöglicht hat.

Anmerkungen

- 1) Im einleitenden Abschnitt differenziert Rolf LANGEHEINE zwischen Reliabilität und Stabilität, wobei er als zeitstabil solche Personen bezeichnet, die in allen drei Wellen reliabel antworten. Ich halte den Ausdruck "zeitstabil" für nicht sehr glücklich, weil man im allgemeinen Modell zwischen beobachteter Stabilität (im Antwortverhalten) und Stabilität der latenten Eigenschaft unterscheiden muß. Zeitstabilität der latenten Klassen haben wir in allen unseren Modellen vorausgesetzt. Was LANGEHEINE als Anteil der Zeitinstabilen bezeichnen würde, habe ich daher als Anteil von Personen charakterisiert, die in mindestens einer Welle unzuverlässig antworten. Sieht man einmal davon ab, daß ich nur die Größenordnung und nicht die exakten Werte berichtet habe, so decken sich natürlich die von LANGEHEINE und mir berechneten Prozentzahlen.
- 2) LANGEHEINE und ich hatten zunächst den Verdacht, das Modell M4' sei unteridentifiziert. Wir haben, um dies zu prüfen, für die vier Klassenanteile, die Antwortwahrscheinlichkeiten der drei Werttypen und das N willkürlich numerische Werte vorgegeben, daraus die beobachteten Antworthäufigkeiten errechnet und diese als (fiktive) Daten eingelesen. Daß MLLSA trotz



Variation der Startwerte für M4' aus diesen Häufigkeiten immer wieder die unseren Berechnungen zugrundeliegenden Parameterwerte fand, spricht für die Identifikation von M4'. Bewiesen ist sie dadurch natürlich nicht.

- 3) Wenn das hier berichtete L^2 etwas geringer ist als das von LANGEHEINE, so weist das auf eine typische Schwierigkeit von Schätzungen dieser Art hin: Die Programme bleiben verhältnismäßig oft in lokalen Minima hängen, so daß man gut daran tut, zahlreiche Läufe mit unterschiedlichen Startwerten zu rechnen. Kleine L^2 -Differenzen können von großen Änderungen der Parameterwerte begleitet sein. Auch in unseren Lösungen differieren die Schätzwerte einiger Antwortwahrscheinlichkeiten dem Betrag nach um mehr als 0.1.
- 4) Es soll nicht unerwähnt bleiben, daß selbst M3' in Teilen überangepaßt ist, weil sich auch in diesem Modell für die Antwortwahrscheinlichkeiten von Mi., in der ersten Welle einige unplausible Werte ergeben. Man müßte auch hier noch Restriktionen einführen, wodurch sich die Modellanpassung geringfügig verschlechtern würde.

Prof. Dr. Wolfgang Jagodzinski, Universität Gießen, FB 03, Institut für Soziologie,
Karl-Klöckner-Str. 21, Haus E, 6300 Gießen

Archivexperten trafen sich im Dänischen Archiv

Die sozialwissenschaftlichen Datenarchive aus fast allen westeuropäischen Ländern haben sich vor Jahren zu einer Vereinigung mit der Bezeichnung CESSDA (Committee of European Social Science Data Archives) zusammengeschlossen. Hauptaufgabe dieser Vereinigung ist es, den internationalen Erfahrungsaustausch zwischen den Archiven zu fördern, die Archivbestände untereinander auszutauschen und nationalen Nutzern die internationale Suche bzw. den internationalen Datentransfer zu erleichtern.

Im Rahmen des interinstitutionellen Erfahrungsaustausches hat sich schon seit Jahren eine intensive Zusammenarbeit einzelner Archive entwickelt. Das hat dazu geführt, daß die Arbeitsvorgänge in den Datenarchiven harmonisiert wurden und ein ständiger Austausch von Erfahrungen über Archivierungstechniken sowie Datenverarbeitungssoftware stattfand. Einem solchen Erfahrungsaustausch diente auch die jüngste Konferenz, die vom Dänischen Datenarchiv (DDA) in Odense ausgerichtet wurde. Dieses erste "CESSDA Expert Seminar", das auf Initiative der Archivleiter durchgeführt wurde und der Förderung des Erfahrungsaustauschs dienen sollte, hatte die sogenannten Studienbeschreibungen zum Hauptthema. "Study descriptions" sind nach dem Verständnis der Archive methodische und inhaltliche Beschreibungen von empirischen Untersuchungen, die in Form von Datensätzen in die Archive eingebracht werden. Solche Studienbeschreibungen sind den Benutzern des Zentralarchivs als Kurzform beispielsweise aus dem Datenbestandskatalog bekannt, wo sie, in numerischer Reihenfolge sortiert, einen Gesamtüberblick über die Bestände des Hauses geben.

Bei der Betrachtung von Datenbestandskatalogen anderer Archive wird dem Leser auffallen, daß es eine hohe Vergleichbarkeit der Darstellungen gibt. Dies ist kein Zufall. Seit den 60er Jahren wurde ein umfangreiches Erfas-