

P R O B L E M
und
E N T S C H E I D U N G

Arbeiten zur Organisationspsychologie
aus der Abteilung für Angewandte Psychologie
des Psychologischen Instituts der Universität München
und der Fachgruppe Psychologie der Universität Augsburg

Heft 14

München - Augsburg 1975

Zum Problem der Messung von Einstellungsänderungen durch
Meßwiederholung

Gisela Albers, Hermann Brandstätter und Ulf Peltzer

Aus Praktikabilitätsgründen werden in der sozialpsychologischen Forschung häufig Meßwertdifferenzen aus Vorher-Nachher-Erhebungen als Meßwerte der Einstellungsänderung verwendet. Die beiden schwerwiegendsten Argumente gegen diese Vorgehensweise werden vorgestellt, indem die möglichen Auswirkungen aufgezeigt werden.

Getrennt für den Fall der klassischen Vorher-Nachher-Messung und den Fall der fortlaufenden Meßwerterhebungen werden Auswertungsvorschläge diskutiert, die diese Fehlerquellen vermeiden oder zumindest reduzieren. Dabei wurden komplizierte Kontrollgruppen-Designs außer Acht gelassen, da sie in der sozialpsychologischen Forschung entweder nur unter großem ökonomischen Mehraufwand zu realisieren oder in Einzelfällen gar nicht sinnvoll sind, da für bestimmte Fragestellungen echte Kontrollgruppenbedingungen nicht hergestellt werden können.

1. Vorbemerkungen

Im Zusammenhang mit Untersuchungen zur Wirkung sozialer Verstärkung in Gruppendiskussionen (Arbeiten der Augsburger Forschungsgruppe, siehe u.a. Problem und Entscheidung: Heft 8, 1972; Heft 12, 1974; Heft 14, 1975) stellte sich als ein zentrales Problem die Messung von Einstellungsänderungen. In diesen Untersuchungen wurde die Messung von Einstellungen aus Praktikabilitäts-Gründen mittels mehrstufiger Rating-Skalen vorgenommen. In einigen Untersuchungen wurde die relevante Einstellung jeweils vor und nach der Diskussion erhoben, in anderen zusätzlich fortlaufend während der Diskussion. In beiden Fällen wurden Differenzen zwischen einzelnen Meßwerten als Grundlage für die Variable Einstellungsänderung berechnet. Die Problematik dieses Vorgehens und alternative Ansätze sollen im folgenden diskutiert werden.

2. Probleme der Differenzenbildung

Einstellungen sind häufig instabil, auch wenn keine Beeinflussung erkennbar ist. Selbst unter der Annahme, daß alle Einflüsse auf die jeweilige Einstellung kontrollierbar wären, ergäbe sich eine Zufalls-Streuung aufgrund des nicht absolut reliablen Meßinstruments. Besonders bei der Verwendung von Rating-Skalen dürfte der Meßfehler im Vergleich zu aufwendigeren Verfahren relativ groß sein.

Die Zufalls-Streuung aufgrund unkontrollierbarer Merkmalschwankungen und der nicht perfekten Reliabilität des Meßinstrumentes hat unerwünschte Auswirkungen.

2.1. Es tritt der Regressionseffekt (Galton, 1885) auf. Das heißt, bei der wiederholten Messung ist die Wahrscheinlichkeit für Vpn, die bei der ersten Messung im Extrembereich liegen, bei der zweiten Messung im Mittelbereich zu liegen, auch dann relativ groß, wenn keine systematische Einwirkung stattgefunden hat. Dieser Effekt ist desto stärker, je größer die Zufalls-Streuung der Einzelmessungen ist. Ein möglicher Treatment-Effekt läßt sich häufig nicht vom Regressionseffekt trennen.

2.2. Da bei wiederholter Messung einer Variable vor und nach dem Treatment eine hohe positive Korrelation zwischen den beiden Messungen zu erwarten ist, wird die Reliabilität der Meßwertdifferenzen gegenüber der Reliabilität der Einzelmessung stark reduziert. Aus den Axiomen der klassischen Testtheorie läßt sich folgende Beziehung zwischen der Reliabilität der Differenzwerte und der Reliabilität der Ausgangswerte ableiten:

$$R(x_{t_2} - x_{t_1}) = 1 - \frac{1 - \bar{R}_x}{1 - r_{x_{t_1} x_{t_2}}}$$

$R(x_{t_2} - x_{t_1})$: Reliabilität der Differenzwerte
 \bar{R}_x : mittlere Ausgangsreliabilität
 $r_{x_{t_1} x_{t_2}}$: Interkorrelation der Messungen

So ergibt sich z.B. für eine Ausgangsreliabilität von $\bar{R}_x = .95$ und einer Interkorrelation von $r_{x_{t_1} x_{t_2}} = .90$ eine Reliabilität der Differenzen von

$$R(x_{t_2} - x_{t_1}) = .50!$$

3. Alternativen

Die problematischen Auswirkungen bei der einfachen Differenzenbildung von Meßwerten lassen es ratsam erscheinen, sich um alternative Auswertungsansätze zu bemühen. Im folgenden werden Vorschläge für zwei Datenarten diskutiert:

1. Anfangs-End-Veränderung (t_1, t_2)
2. Fortlaufende Meßwert-Erhebung (t_1, t_2, \dots, t_n)

3.1. Für Untersuchungen mit Anfangs-End-Messung bieten sich vier Möglichkeiten an:

1. Vergleich der Endwerte
2. Berechnung der "wahren" Differenzen
3. Kovarianzanalyse (Winer, 1971)
4. Regression der Endwerte auf die Anfangswerte

3.1.1. Bei Random-Zuordnung der Vpn auf die einzelnen Zellen des Versuchsplans müßten die Verteilungen der Anfangswerte bei hinreichend großem Stichprobenumfang in allen Zellen gleich sein hinsichtlich Mittelwert, Varianz, Schiefe und Exzeß. Für diesen Fall würde die Analyse der Endwerte ausreichen. Allerdings werden die Verteilungen in der Praxis zumindest leicht unterschiedlich sein, da die Stichproben in komplexeren, sozialpsychologischen Untersuchungen nicht sonderlich groß sein können.

3.1.2. Relativ häufig wird in Untersuchungen, z.B. Längsschnittuntersuchungen zur Intelligenz, mit den sogenannten wahren Differenzen oder wahren Werten gerechnet, die über die Reliabilität der Meßwerte geschätzt werden können. Da es bei den Skalierungen der Einstellung mittels Rating-Skalen ohne zusätzliche aufwendige Reliabilitäts-Untersuchungen kein erfolgversprechendes Verfahren zur Bestimmung oder Abschätzung der Zuverlässigkeit der Daten gibt, scheint dieser Ansatz nicht praktikabel.

3.1.3. Bei nicht vollständiger Random-Zuordnung der Vpn zu den einzelnen Zellen des Versuchsplans wird in der Literatur häufig die Kovarianzanalyse empfohlen. Nach Cronbach u. Furby (1970) sind die Voraussetzungen der Kovarianzanalyse:

- a) randomisierte Stichprobenaufteilung (!)
- b) Korrelation der Kovariaten mit den abhängigen Variablen soll größer/gleich .40 betragen
- c) Linearität der Regressionen unter allen Bedingungskombinationen
- d) Homogenität der Regressionen unter allen Bedingungskombinationen.

Cronbach u. Furby halten die Anwendung der normalen Kovarianzanalyse bei dem hier behandelten Problem für unzulässig. Sie schlagen vor, eine Regression der Anfangs- auf die Endwerte (also zeitlich gesehen rückwärts) innerhalb der Treatment-Gruppe durchzuführen und dann mit den geschätzten Anfangswerten als Kovariate und den empirischen Endwerten als abhängiger Variablen eine Kovarianzanalyse zu rechnen. Dabei soll die "Rückwärts"-Regression zur Reduzierung der treatment-unabhängigen Varianz dienen.

Leider geht der Artikel von Cronbach und Furby nur kurz auf die Voraussetzungen und Anwendungsmöglichkeiten ihres Vorschlags ein, und der von ihnen in diesem Zusammenhang zitierte Artikel von Lord (1960) beschäftigt sich mit diesem Verfahren überhaupt nicht. Uns erscheint das genannte Vorgehen als eine unerlaubte Methode der Fehlerreduktion. Angesichts dieser Problemlage dürfte es weiterhin gerechtfertigt sein, die Kovarianzanalyse für die Auswertung von Meßwiederholungsuntersuchungen zu verwenden und zwar sowohl für den Fall der vollständig wie auch der nicht vollständig randomisierten Stichprobenaufteilung. Dies natürlich nur, wenn die Voraussetzungen der Kovarianzanalyse (3.1.3. b, c, d) erfüllt sind, die vor jeder Analyse geprüft werden müssen.

3.1.4. Die einfachste, erfolgversprechende Berechnung von Veränderungsmaßen ist die für jede Bedingungskombination getrennt berechnete Regression der Endwerte auf die Anfangswerte. Sie ermöglicht es gleichzeitig, die Voraussetzungen der Kovarianzanalyse zu überprüfen.

3.1.4.1. Nach der Berechnung der Regressionen folgt die Prüfung auf:

- a) Linearität jeder Regressionskurve
- b) Homogenität aller Regressionsgraden.

3.1.4.2. Falls sich weder in 3.1.4.1. a) noch in 3.1.4.1. b) signifikante Abweichungen zeigen, kann eine Kovarianzanalyse mit der Anfangsmessung als Kovariante und der Endmessung als abhängiger Variable gerechnet werden (siehe 3.1.3.).

3.1.4.3. Falls Linearität, nicht aber die Homogenität der Steigerungen der Regressionsgeraden gegeben ist, fällt die Kovarianzanalyse als zulässiges Verfahren aus, jedoch ist die Interpretation der Unterschiede der Regressionsgeraden relativ einfach.

3.1.4.4. Falls es Abweichungen von der Linearität gibt, kann nur eine Interpretation der Kurvenverläufe für jede einzelne Bedingungskombination erfolgen.

3.2. Für Einstellungsdaten, die fortlaufend über eine bestimmte Zeitdistanz erhoben werden, bieten sich zwei Möglichkeiten an:

1. Kovarianzanalyse mit Meßwiederholung
2. Multivariate Varianzanalyse mit individuellen Kurvenparametern als abhängige Variable.

3.2.1. Grundsätzlich ist es, ähnlich wie bei der Anfangs-End-Messung, möglich, für die fortlaufende Skalierung die Regressionskurven für jede Bedingungskombination zu berechnen, in diesem Fall allerdings zusätzlich für jeden Durchgang getrennt, und bei Erfüllung der Voraussetzungen anschließend eine Kovarianzanalyse mit Meßwiederholung durchzuführen. Dagegen spricht, daß aufgrund der überaus hohen Interkorrelationen zwischen den einzelnen Durchgängen, die mit wachsendem Zeitabstand abnehmen, die Voraussetzung der Homogenität der Varianzen und Kovarianzen auf jeden Fall verletzt sein werden. Ein entsprechendes Verfahren, wie es Greenhouse u. Geisser (1959) für den Fall einer Varianzanalyse mit Meßwiederholung und inhomogenen Kovarianzen vorgeschlagen haben, gibt es offenbar nicht. Eine Übertragung dieses Verfahrens, das auf eine konservative Korrektur der Freiheitsgrade hinausläuft, auf die Kovarianzanalyse ist nicht ohne weiteres zulässig.

3.2.2. Rahlfs u. Bedall (1971), die einen Vorschlag von Grizzle (1970) aufgreifen, schlagen für die Analyse zeitabhängiger Daten die Multivariate Varianzanalyse (MANOVA) bei vorhergehender Datenreduktion vor. Die Datenreihe (Verlaufskurve) jeder Vp wird mittels einer Regressionsanalyse auf einige wenige Parameter reduziert, mit denen dann eine MANOVA gerechnet wird. In den meisten Fällen wird vermutlich das Polynom $y = a + bt - ct^2$, eventuell sogar der lineare Fall $y = a + bt$ ausreichen. Untersuchungsspezifische Kurvenverläufe wie z.B. die Oszillation der Meinung in einer Diskussion, in der pro und contra-Argumente abwechseln, lassen sich durch die Erweiterung des Polynoms durch einen sin- bzw. cos-Ausdruck beschreiben. In einem solchen Fall wären dann die Parameter für die Gleichung $y = a + bt + ct^2 + d \cos(\pi t)$ zu bestimmen, so daß die MANOVA mit 4 abhängigen Variablen gerechnet würde. Der Parameter d wäre ein Ausdruck für die Amplitudenhöhe der Meinungsschwankung, der Parameter b für die Richtung und Stärke einer Meinungsänderung. Der Parameter a ist nicht eindeutig zu interpretieren; man könnte ihn mit Einschränkungen als "wahren" Anfangswert auffassen. Der Parameter c gäbe Aufschluß darüber, inwieweit eine Beeinflussung der Meinung, sofern sie vorhanden ist, konstant über die Zeit anhält. Vorteile dieses Verfahrens sind:

- a) die Aggregation von Individuen-Meßwerten wird nicht auf der ersten Stufe der Auswertung vorgenommen
- b) viele jeweils hochkorrelierende Meßwerte werden zu wenigen, reliableren Parameterschätzungen zusammengefaßt, wobei die MANOVA Korrelationen zwischen den Parametern nicht ausschließt.

Nachteil dieses Verfahrens ist, daß der Regressionseffekt (siehe 2.1.) nur dann ausgeschaltet werden kann, wenn der "Trend zur Mitte" sich nicht gleichmäßig mit der Zeit vollzieht.

Literatur

- Cronbach, L.J. u. Furby, L.: How we should measure "change" or should we? Psych. Bull., 1970, 74, 68-80
- Galton, F.: Regression towards mediocrity in hereditary stature. Journal of the Anthropological Institute, 1885, 15, 246-263
- Greenhouse, S.W. u. Geisser, S.: On methods in the analysis of profile data. Psychometrika, 1959, 24, 95-112
- Grizzle, J.E.: An example of the analysis of a series of response curves and an application of multivariate multiple comparisons. In: Bose et al. (Eds.) Essays in probability and statistics The University of North Carolina, Press Chapel Hill, 1970, 311-326
- Lord, F.M.: Large sample covariance analysis when the control variable is fallible. Am Stat. Ass. J., 1960, 65, 307-321
- Rahlf's, V.W. u. Bedall, F.K.: Biometrie und klinische Pharmakologie. Int. J. clin. Pharmacol., 1971, 96-109
- Winer, B.J.: Statistical Principles in experimental design. McGraw Hill, New York 1971²

Der Einfluß stellvertretender sozialer Verstärkung auf den Entscheidungsverlauf der Beobachter von Gruppendiskussionen

Lutz von Rosenstiel, Gisela Stocker-Kreichgauer
unter Mitarbeit von Gisela Albers¹⁾

Vorhergesagt wurde, daß

- die überdauernde Einstellungsänderung und
- die aktuelle Einstellungsänderung

von Beobachtern kontroverser Diskussionen vom Ausmaß der stellvertretenden sozialen Verstärkung abhängig ist.

Vorhergesagt wurde weiterhin, daß

- die Einstellungsänderungen der Beobachter spezifisch auf die verstärkten Argumente hin erfolgen und
- die Veränderungsbereitschaft (Variabilität) umso größer ist, je freundlicher und je intensiver die Reaktionen des bei der Diskussion anwesenden Publikums sind.

Zusätzlich wurde vorhergesagt, daß die überdauernden, aktuellen und spezifischen Einstellungsänderungseffekte sowie die Veränderungsbereitschaft bei einer Fernsehdarbietung der Diskussion stärker sind als beim Hörfunk und hier wiederum stärker sind als beim Sitzungsprotokoll.

Geprüft wurden unsere Hypothesen an 257 nach dem Zufall aus dem Augsburger Telefonbuch ausgewählten Vpn., von denen 128 Personen verwertbare Daten lieferten. Die statistischen Prüfungen erfolgten z. T. mit parametrischen, z. T. mit nonparametrischen Verfahren.

Es zeigte sich, daß es im erwarteten Sinne zu überdauernden Einstellungsänderungen kam, die aktuelle Wirkung jedoch nur bei reinen Belohnungsbedingungen zu beobachten war, während bei Bestrafungsbedingungen Bumerangeffekte auftraten. Die Einstellungsänderungen erfolgten bevorzugt auf die verstärkten Argumente im Sinne der Verstärkung. Die Veränderungsbereitschaft war erwartungsgemäß umso größer, je angenehmer die sozialen Verstärkungen erschienen und je intensiver sie ausfielen.

Die Medieneffekte fielen deutlich schwächer aus. Die Ergebnisse sprechen jedoch - meist allerdings nur tendenziell - dafür, daß der überdauernde Einfluß erwartungsgemäß in der Fernsehbedingung am stärksten ist, die Veränderungsbereitschaft aber gerade bei diesem Medium am geringsten erscheint. Interpretierbare Unterschiede zwischen den Wirkungen des Hörfunks und des Sitzungsprotokolls konnten nicht aufgezeigt werden.

1) Wir danken L. Dargatz, U. Ruchti u. B. Wißner für ihre Mitarbeit bei den empirischen Untersuchungen und bei der Datenauswertung