Zustimmungsanteile und Mittelwerte von Liker-skalierten Items

Borg, Ingwer; Gabler, Siegfried

ZUSTIMMUNGSANTEILE UND MITTELWERTE VON LIKERT-SKALIERTEN ITEMS

INGWER BORG & SIEGFRIED GABLER

Bei der Darstellung von Umfrageergebnissen von Likert-skalierten Items verwendet man in der nicht-wissenschaftlichen Praxis als Ergebnisstatistiken oft den Prozentsatz der Personen, die den Items zugestimmt haben. Diese Statistiken sind einfacher und zuverlässiger zu verstehen als die in der Wissenschaft gebräuchlicheren Skalenmittelwerte. Andererseits scheint klar zu sein, dass die Zustimmungsprozente nur relativ grobe Beschreibungen der Antworten sind. Hier wird jedoch sowohl empirisch mit den Daten aus vier großen Umfragen als auch theoretisch mittels eines Modells, das die empirischen Antwortverteilungen auf der Likert-Skala aus einer stets normalen latenten Urteilsverteilung ableitet, gezeigt, dass der Informationsgehalt der Zustimmungsprozente im allgemeinen nicht geringer ist als der von Mittelwerten, wenn man eine konstante Varianz der latenten Verteilung unterstellt oder wenn man zusätzlich die Ablehnungsprozente berichtet. Die verbreitete Verwendung von Zustimmungsprozenten an Stelle von Mittelwerten ist also nicht nur aus kommunikativen Gründen gut begründet.

When reporting the results of surveys using Likert-scaled items, it is common-place in the non-scientific context to use percent agreement rather than scale means. Percent agreement is easier to understand and more reliable in communicating than scale mean values which are more popular in science. It seems clear, though, that percent agreement is but a coarse approximation of the information contained in the mean. We here show that this is not so, using empirical evidence based on four large surveys, and developing a model that relates the answer distributions of Likert items back to latent judgement distributions that are always normal. Both lines of arguments show that using percent agreement – and assuming either a constant item variance or reporting also percent disagreement – is well justified not only for reasons of better communication.

1. Wie werden Antworttendenzen für Items berichtet?

Relativ wenig Beachtung findet in der Umfrageforschung die Frage, wie man ihre Ergebnisse so darstellen kann, dass sie auch von fachfremden Personen leicht und korrekt verstanden werden. Der erfahrene Sozialforscher mag hierbei zunächst an die Vermittlung multivariater Strukturen oder die Bewertung der Umfrageergebnisse denken. Den Konsumenten hierbei zu leiten ist eine offensichtliche Herausforderung. In der Praxis beginnt das Problem aber bereits viel früher, nämlich bei dem Versuch, dem Konsumenten die zentralen statistischen Tendenzen einzelner Items optimal zu kommunizieren.

In der Umfrageforschung verwendet man vielfach Items, deren Frageteil als Feststellung formuliert wird, etwa wie folgt: „Ich bin mit meinem Job zufrieden." Seine Antwort kann der Befragte auf einer bipolaren Zustimmungs- bis Ablehnungsskala ausdrücken. Diese Skala hat meist fünf Kategorien, die mit „stimme voll zu", „stimme eher zu", „teils-teils", „stimme eher nicht zu" bzw. „stimme überhaupt nicht zu" etikettiert werden („Likert-Skala"). Als Ergebnis der Umfrage wird schließlich der Mittelwert der numerisch kodierten Antwortkategorien berichtet, also z.B. „Die Arbeitszufriedenheit der Befragten ist 3,8".

Ein solcher Wert ist jedoch für den Konsumenten nicht ohne weiteres informativ. Er muss dazu wissen, ob die Kodierung von 1 bis 5 oder von 5 bis 1 erfolgte. Beides ist möglich. Ersteres findet sich häufig in der angewandten Psychologie und wird dort als „Notenskala" bezeichnet. Letzteres ist eher in wissenschaftlichen Publikationen üblich, weil es dort Konvention ist, eine stärkere Ausprägung des gemessenen Attributs mit einem größeren Skalenwert zu verbinden. Der Rezipient der Umfrage muss also zumindest beachten, wie die Skala gepolt war. Rückfragen („Was war noch mal 1?") sind in der Praxis daher nicht selten. Und keine Rückfragen bedeutet noch nicht, dass die Werte richtig rezipiert wurden. Ein anderes Problem ist, dass die Skala wenig vertraut ist. Das führt zu der Frage: „Ist 3,8 gut oder schlecht?" oder zumindest zu „Ist 3,8 groß oder klein?"

Zudem wird nicht immer die klassische 5-stufige Skala verwendet. Man findet in der Praxis häufig auch 3-, 7- oder 9-stufige Antwortskalen. Manche Autoren verzichten auch auf die Mittelkategorie. Wieder andere Autoren bieten zusätzlich noch eine „Weiß-Nicht"-Kategorie an.

Ein Skalenmittelwert wird damit für den gelegentlichen Konsumenten von Umfragebefunden nicht leicht lesbar. An Stelle der Skalenmittelwerte werden daher vor allem in der Praxis von Organisationsbefragungen, wo ein breites und tiefes „Survey Feedback" der Befragungsergebnisse in die Organisation ein wesentliches Motiv für die Befragung ist (Nadler 1977; Borg 2001), andere Statistiken berichtet. Im einfachsten Fall wird die gesamte Verteilung für jedes Item berichtet. Meist wird diese Verteilung jedoch weiter vereinfacht: Berichtet werden entweder nur noch die Anteile der Personen, die sich dem Item gegenüber „zustimmend“, „gemischt“ oder „ablehnend“ geäußert haben (Macey 1996; Edwards et al., 1997) oder überhaupt nur noch der Prozentsatz der Personen, die sich zustimmend geäußert haben (Bergler/Piwinger 2000; Borg 2000a; Bruennecke/Canisius 1991; Rogelberg et al., im Druck). Das gilt insbesondere für kommerzielle Institute oder für Umfragekonsortien wie die Mayflower Group oder die ITSG (Kraut 1997; Church/ Waclawski 2001), die ihre Normwerte oder „Benchmarks“ in Form derartiger Zustimmungsprozente (*Ja%*) zur Verfügung stellen.

Die *Ja%*-Statistiken haben zweifellos kommunikative Vorteile. Nachfragen nach der Zahl der Antwortkategorien und der Polung der Skala entfallen hier automatisch. Zudem ist ein Prozentwert eine Maßzahl, die jedem vertraut ist. Andererseits erscheint es intuitiv klar, dass die *Ja%*-Statistik den zentralen Trend eines Items nur relativ grob beschreibt. Anders als beim Mittelwert bleiben in der *Ja%*-Statistik die Intensitäten der zustimmenden Antworten unberücksichtigt. Neutral-ambivalente und ablehnende Antworten werden sogar vollständig ignoriert. In der Praxis wird dies bisweilen kritisiert: Warum erst so differenziert fragen, wenn dann nur noch ein Teil der Antworten und dann auch noch zusammenfassend berichtet wird?

Kritik zur *Ja%*-Statistik kommt aber auch aus anderer Richtung. So bewerten Rogelberg et al. (im Druck) die *Ja%*-Statistik als „extrem problematisch“. Ihr Hauptkritikpunkt ist, dass bei ihrer Verwendung psychometrische Information verloren geht. Sie schlagen daher an Stelle dieser Statistiken die Verwendung von Mittelwerten vor, die so reskaliert sind, dass sie ebenfalls leicht verständlich werden. Eine dieser Transformationen bildet den Mittelwert ab auf eine „grade point“-Skala von 4 bis 1, eine Art amerikanischer Notenskala. Eine andere transformiert den Mittelwert auf eine „Testwertskala“, bei der 0=min und 100=max ist. Die Verwendung einer solchen Skala soll den Mittelwert leichter interpretierbar machen.

Die Transformation des Mittelwerts auf eine solche Skala führt aber zu neuen Problemen. Grundsätzlich gilt, dass die transformierten Statistiken von den Daten weiter entfernt sind als z.B. der Mittelwert oder *Ja%*. Nehmen wir an, wir hätten das für Mitarbeiterbefragungen typische Item „Ich plane ernsthaft, die Firma in den kommenden 12 Monaten zu verlassen“. Auf der Testwertskala hätte sich nun ein Wert von „30“ ergeben. Was können wir daraus lernen? Wie viele Personen planen nun, die Firma zu verlassen? Sind das viele oder wenige? Diese Fragen sind nicht unmittelbar zu beantworten. Dazu kommt, dass in der Praxis 100er-Skalen oft als Prozentskalen gelesen werden, was hier natürlich ganz falsch wäre. Zudem muss die Transformation selbst erklärt und verstanden werden – eine zusätzliche Anforderung für den Endnutzer der Daten.

Eine „grade point"-Skala ergibt sich in Deutschland, wenn man die übliche 5-stufige Likert-Skala als Notenskala kodiert, also mit 1=„stimme voll zu" bis 5=„lehne voll ab". Hier zeigt sich aber gleich ein weiteres Problem: Die deutsche Notenskala entspricht nicht der amerikanischen Notenskala. Das schränkt zumindest die Nützlichkeit von Notenskalen in internationalen Studien ein. Prozentwerte versteht dagegen weltweit jeder gleich. Ein anderes Problem ist die Konnotation der Notenskalen von „gut" bis „schlecht". Die *Ja%*-Statistik hat dieses Problem nicht, vorausgesetzt man liest sie als Prozentsatz der Zustimmung und nicht als „percent favorable", wie in der Praxis häufig üblich (Folkman 1998; Rogelberg et al., im Druck; Edwards et al., 1997).

Die weite Verbreitung der *Ja%*-Statistiken in der Praxis und ihre attraktiven Eigenschaften als Transporteur von Umfrageergebnissen lassen es angezeigt erscheinen, der Frage, wie diese Statistiken und Skalenmittelwerte empirisch und theoretisch zusammenhängen, genauer nachzugehen. Damit wollen wir die Diskussion, welche dieser Statistiken man verwenden sollte und was dabei ggf. zu beachten ist, auf eine solidere Grundlage stellen. Wir bauen dabei auf einer Arbeit von Borg (1989) auf, die enge Zusammenhänge zwischen Skalenmittelwerten und Zustimmungsanteilen zeigt. Wir verwenden hier jedoch weit größere Stichproben und entwickeln eine explizite theoretische Fundierung für die Beziehung dieser Statistiken.

2. Empirische Beziehung von *Ja%* und Mittelwert

Im folgenden gehen wir zunächst rein empirisch vor. Wir fragen direkt, welche Beziehung *Ja%*-Statistiken und Mittelwerte für Likert-skalierte Items aufweisen. Dazu betrachten wir vier große Organisationsumfragen, in denen alle Items nach dem Likert-Ansatz erhoben wurden.

Studie A ist eine weltweite Mitarbeiterbefragung aus dem Jahr 2000, durchgeführt als Online-Umfrage in einem HighTech-Unternehmen mit ca. 22.000 Mitarbeitern (siehe dazu auch Borg, 2000b). Der Rücklauf der Befragung betrug 89%. Der Fragebogen umfasste 94 Items zu den üblichen Themen einer Mitarbeiterbefragung (Arbeitsplatzbedingungen, Arbeit selbst, Bezahlung, Vorgesetzter usw.). Alle Items waren im Likert-Format formuliert. Die 5-stufige Antwortskala lautete in der deutschen Sprachversion: „stimme voll zu" (kodiert als 1), „stimme zu", „teils-teils", „stimme nicht zu" und „stimme überhaupt nicht zu" (kodiert als 5).

Studie B ist eine Mitarbeiterbefragung aus dem Jahr 1998, durchgeführt in einem Produktionsbetrieb der Automobilindustrie, der im Drei-Schicht-Betrieb arbeitet. Der Betrieb hatte ca. 27.000 Mitarbeiter, ca. 80% davon Arbeiter. Verwendet wurden hier 67 Items und eine 5-stufige Likert-Skala mit den Labels „trifft voll zu“ (=1), „trifft eher zu“, „teils-teils“, „trifft eher nicht zu“, „trifft überhaupt nicht zu“ (=5). Die Befragung wurde während der Arbeitszeit – bei angehaltener Produktion (siehe dazu Fotos in Borg, 2000a, S. 165f.) – als Vollbefragung mit Papier und Bleistift durchgeführt. Der Rücklauf betrug 95%.

Bei Studie C handelt es sich um eine Mitarbeiterbefragung aus dem Jahr 1998 in einem deutschen Logistikunternehmen mit ca. 220.000 Mitarbeitern. Der Fragebogen enthielt einen Kern von 54 Items, die allen Mitarbeitern gestellt wurden. Die Befragung wurde postalisch mit Versand des Fragebogens an die Privatadresse der Mitarbeiter durchgeführt. Der Rücklauf betrug 49%. Die Labels der Likert-Skala lauteten: „stimme voll zu“ (=1), „stimme zu“, „teils-teils“, „stimme nicht zu“, „stimme überhaupt nicht zu“ (=5).

Bei Studie D handelt es sich um eine Vollbefragung der Mitarbeiter eines Autowerks im Jahr 1998. Verwendet wurden hier 85 Items mit 7-stufiger Antwortskala, etikettiert mit „stimmt voll und ganz“ (=1), „stimmt“, „stimmt in etwa“, „teils-teils“, „stimmt eher nicht“, „stimmt nicht“, „stimmt überhaupt nicht“ (=7). Der Erhebungsmodus war die Wahllokalmethode (Borg 2000a). Die Erhebung erfolgte während der Arbeitszeit. Der Rücklauf lag bei 97%.

Die Abbildungen 1-4 zeigen die Beziehung der Zustimmungs- und Ablehnungswerte (*Ja%* bzw. *Nein%*) der jeweiligen Items zu den Skalenmittelwerten dieser Items. Wie man sieht sind die Beziehungen jeweils sehr eng. Die Trendlinien – hier bestimmt mittels der DWLS-Funktion („distance weighted least squares“ aus Systat V9) – sind weitgehend linear, vor allem im mittleren Bereich, so dass die Korrelation der *Ja%*- (*Nein%*-) und der Mittelwerte fast perfekt sind: Sie betragen -0,98 (0,91), -0,99 (0,96), -0,97 (0,87) bzw. -0,99 (0,92). Nichtlinearitäten erkennt man vor allem über den Extremwerten der Likert-Skala. Die Trends verlaufen insgesamt leicht S-förmig.

3. Theoretische Beziehung von *Ja%* und Mittelwert

Die Aufgabe, seine Meinung oder Einstellung zu einem Statement auf einer Likert-Skala auszudrücken, ist eine recht komplexe Anforderung. Dabei interessiert uns hier nur der letzte Schritt des Urteilsprozesses: Ein Urteil ist i.S. einer bestimmten Stärke, dem Item zuzustimmen oder es abzulehnen, gebildet und muss nun vom Befragten so „formatiert“ (Tourangeau et al., 2000) werden, dass es in die Likert-Skala des Fragebogens passt, also auf eine Skala abgebildet werden, die (a) in wenige Kategorien gerastert und (b) nach oben und unten beschränkt ist. Für das latente Kontinuum nehmen wir an, dass weder (a) noch (b) gilt: Vielmehr unterstellen wir, dass die Verteilung der Urteile der befragten Personen auf diesem Kontinuum normal ist.

Empirisch findet man natürlich, dass die Verteilungen der Antworten auf die Meinungs- und Einstellungsitems von Mitarbeiterbefragungen keineswegs immer normal sind. Das ist jedoch leicht zu erklären. Ein Item kann ja z.B. ganz extrem formuliert sein, so dass den Befragten oft nichts anderes übrig bleibt, als die Extremkategorien zu wählen. Was soll man etwa auf die Frage „Ich bin gelegentlich schlechter Laune" anderes antworten als „stimme voll zu" oder „stimme zu"? Ähnliches gilt für Items besonderer Art wie z.B. Werteitems, die i.d.R. stark schiefe Antwortverteilungen haben. Was soll man anderes antworten auf die Frage „Mir ist es wichtig, ein sicheres Einkommen zu haben" als „stimme voll zu" oder zumindest „stimme zu"? Die Antwortverteilungen sind nur dann annähernd normal, wenn das Item eine mittlere Schwierigkeit hat. Bei extremen Itemmittelwerten werden die Antwortverteilungen schiefer und am Ende sogar J-förmig. Trotzdem sind solche Verteilungen praktisch immer eingipflig (Gabler/Borg 1996).

Diese Formen der empirischen Verteilungen sind dann verträglich mit der Annahme latenter Normalverteilungen, wenn man unterstellt, dass diese auf das beschränkte Intervall der Likert-Skala gestaucht werden. Inhaltlich ausgedrückt heißt das, dass Personen, deren Einstellung extremer ist, als es die Likert-Skala auszudrücken erlaubt, ihre Einstellungswerte auf die angebotene Skala transformieren müssen. Dabei nehmen wir an, dass zwei Punkte inhaltlich festliegen: Der Schnittpunkt zwischen einem zustimmenden und einem „teils-teils"-Urteil (z_0); und ebenso der zwischen einem ablehnenden und einem „teils-teils"-Urteil (z_1). Wir nehmen an, dass diese Punkte auf einer 5-stufigen Likert-Skala bei 2,5 bzw. bei 3,5 liegen, also genau im Mittelpunkt zwischen den Kategorien 2= „stimme eher zu" und 3=„teils-teils" bzw. zwischen 3=„teils-teils" und 4=„stimme eher nicht zu". (Auf der 7-stufigen Likert-Skala nehmen wir entsprechend die Werte 3,5 und 4,5 als Fixpunkte an.) Die Stauchung sollte also so erfolgen, dass die wahren Skalenwerte in Antwortwerte abgebildet werden, die zu diesen Fixpunkten gleich geordnet sind. Diese Punkte sind deshalb Fixpunkte, weil hier ein qualitativer Übergang stattfindet z.B. von „gerade-noch"-Zustimmung zu einem „gemischten" Urteil. Die Befragten sollten sich, so nehmen wir an, darüber relativ sicher sein, ob sie zustimmen wollen oder nicht, während die Intensität einer eventuellen Zustimmung eher dem angebotenen Antwortformat angepasst wird.

Die wahre Urteilsverteilung ist laut Modell also eine (μ,σ)-normalverteilte Zufallsvariable X, die mittels der Transformationsfunktion $T(x)$ so in ein endliches Intervall (a,b) abgebildet wird, dass die *Ja%* (der Anteil p_0) und die *Nein%* (der Anteil p_1) in der Ausgangsverteilung mit den *Ja%* bzw. den *Nein%* in der transformierten Verteilung – und weiter dann auch auf der Likert-Skala – übereinstimmen. Sind μ und σ bekannt, lassen sie sich in die entsprechenden (theoretischen) Zustimmungs- bzw. Ablehnungsanteile umrechnen, und umgekehrt, mittels

$$p_0 = \Phi\left(\frac{z_0 - \mu}{\sigma}\right) \text{ bzw. } p_1 = 1 - \Phi\left(\frac{z_1 - \mu}{\sigma}\right)$$

und

$$\mu = \frac{z_1\Phi^{-1}(p_0) - z_0\Phi^{-1}(1-p_1)}{\Phi^{-1}(p_0) - \Phi^{-1}(1-p_1)} \;;\; \sigma = \frac{z_1 - z_0}{\Phi^{-1}(1-p_1) - \Phi^{-1}(p_0)} \;.$$

Die Transformation $T(x)$, die die latente Normalverteilung (auf der Skala von $-\infty$ bis $+\infty$) auf die beschränkte Antwortskala abbildet, ist eine wichtige Spezifikation des hier entwickelten Modells. Für $T(x)$ sollte gelten, dass sie die unendliche Skala so auf die Likert-Skala staucht, dass dabei die Punkte, an denen sich die Bedeutung der Antworten von „zustimmend" zu „gemischt" bzw. von „gemischt" zu „ablehnend" ändert, fest bleiben. Die anderen Punkte sollten so verschoben werden, dass sich ihre relative Lage zu diesen Fixpunkten nicht ändert und dass sich ihre Ordnung untereinander nicht ändert. Als eine metrische Spezifikation der Transformation $T(x)$, die diese Anforderungen erfüllt, wird in Abbildung 5 die logistische Verteilungsfunktion zugrunde gelegt. Die Reihenfolge der Bilder ist gegen den Uhrzeigersinn zu lesen, d.h. ausgehend von der Normalverteilung gelangt man über die Transformationsfunktion zur transformierten Dichte auf der „gestauchten Skala", die schließlich das Histogramm rechts oben liefert, also die (modellgenerierte) Häufigkeitsverteilung auf der Likert-Skala. Die Höhe des Balkens bei der Likert-Skala an der Stelle i (i=1,...,5) ergibt sich aus der Fläche der transformierten Dichte $f_1(z)$ über dem Intervall [i-0,5; i+0,5]. Der rechnerische Prozess ist formal im mathematischen Anhang dargestellt.

Wenn das vorgestellte Modell die Realität adäquat beschreibt, sind die in empirischen Untersuchungen ermittelten mittleren Skalenwerte nach dem zentralen Grenzwertsatz Schätzungen für den Erwartungswert einer Zufallsvariablen Z mit Dichte $f_1(z)$. Die mittleren Skalenwerte werden umso näher bei diesem Erwartungswert liegen, je größer der Stichprobenumfang gewählt wird. Dem Zusammenhang auf Datenebene zwischen *Ja%* bzw. *Nein%* und den mittleren Skalenwerten in den Abbildungen 1-4 entspricht auf Modellebene der Zusammenhang zwischen p_0 bzw. p_1 und dem Erwartungswert von Z. Dieser Erwartungswert lässt sich allgemein berechnen als

$$E(Z) = \int_a^b z f_1(z)dz = \frac{1}{\sigma}\int_{-\infty}^{\infty} T(x)\varphi(x \mid \mu, \sigma)dx$$

und ist daher gleich dem Erwartungswert von *T(X)*, wobei *X* eine (μ,σ)-normalverteilte Zufallsvariable ist, die mittels $T(x)$ in das endliche Intervall (a,b) abgebildet wird. Bei einer 5-stufigen Skala ist a=0,5 und b=5,5. Die analytische Berechnung dieses Erwartungswertes ist im allgemeinen relativ schwierig, die genaue Gestalt der Transformation auch nicht entscheidend. Im mathematischen Anhang ist eine geschlossene Formel für den Erwartungswert explizit angegeben, falls bei der Transformationsfunktion die Laplace-Verteilung Verwendung findet.

Abbildung 6 zeigt für die 94 Items aus Studie A gute Übereinstimmung zwischen den mittleren Skalenwerten und den Erwartungswerten, die sich aus dem Modell ergeben. Genauer wurde aus den *Ja%*- und *Nein%*-Angaben für die 94 Items jeweils der Wert für μ und von σ ermittelt, wie das oben im Abschnitt 3 beschrieben ist. Über die Transformation auf die gestauchte Skala lässt sich dann jeweils der Erwartungswert von *Z* – wie im mathematischen Anhang angegeben – berechnen. Die so erhaltenen 94 Erwartungswerte werden zusammen mit den mittleren Skalenwerten aus den Daten in die Grafik eintragen. Wie der Abbildung 6 zu entnehmen ist, wird das oben dargestellte Modell durch die Daten eindrucksvoll gestützt. Würde man die in den Abbildungen 1 bis 4 eingezeichneten mittleren Skalenwerte durch die Erwartungswerte aus der transformierten Verteilung ersetzen, würden sie fast deckungsgleich ausfallen.

Eine Alternative zum obigen Vorgehen ist die, nicht *Ja%* und *Nein%* zur Spezifikation des Modells zu verwenden, sondern *Ja%* und σ. Im einfachsten Fall setzt man σ=1 für alle Items. Theoretisch bedeutet das, dass dieses Modell unterstellt, dass die Urteilsverteilungen von jedem Item nicht nur normal, sondern auch gleich „breit“ sind. Dann ist die Beziehung des Zustimmungsanteils und des Lageparameters der Verteilung auf dem latenten Urteilskontinuum die aus jedem Statistiklehrbuch bekannte S-förmige Funktion, die insbesondere einen breiten linearen Mittelabschnitt hat. Wird diese Funktion dann über eine Transformation, wie sie in Abbildung 5 unten links zu sehen ist, auf die Likert-Skala gestaucht, bleibt die Linearität des Mittelteils erhalten. Damit ergibt sich eine sehr einfache Beziehung von *Ja%* und mittlerem Skalenwert, bzw. zwischen p_0 und dem Erwartungswert von Z aus der transformierten Verteilung, wie Abbildung 7 im Falle einer 5 -er Skala zeigt.

Die Spezifikation σ=1 ist zumindest für die Daten der Abbildungen 1-3 betrachteten Umfragen empirisch gut begründet, weil jedes Item in diesen Studien eine Varianz von ungefähr 1 hat. (Genauer gilt: Für Studie A ist die mittlere Standardabweichung $\bar{s} = 0{,}91$, mit einer Standardabweichung von $s(s)$=0,09; für B gilt $\bar{s} = 1{,}11$ mit $s(s)$=0,10; für C ist die mittlere Standardabweichung $\bar{s} = 0{,}99$ mit $s(s)$=0,10.) Für Studie D (Abbildung 4), die eine 7-stufige Antwortskala verwendet, ist die mittlere Standardabweichung der verschiedenen Items ebenfalls sehr ähnlich, aber etwas größer mit $\bar{s} = 1{,}33$ bei $s(s)$=0,24.

4. Diskussion und Ausblick

Als Fazit der obigen Untersuchungen kann man zunächst festhalten, dass sich rein empirisch in vier großen Umfragestudien zeigt, dass die *Ja%* (und ebenfalls die *Nein%*) in enger, fast linearer Beziehung zu den mittleren Skalenwerten stehen. Insofern kann man recht genau von den vermeintlich nur „groben" Prozentstatistiken auf die Mittelwerte schließen. Die Zustimmungs- und Ablehnungsprozente enthalten also im wesentlichen die gleiche Information wie die Mittelwerte. Nur bei extremen Prozentwerten wird die Beziehung nichtlinear, was aber nicht weiter verwunderlich ist, weil z.B. bei *Ja%*=100 kein abgestufter Rückschluss mehr möglich ist, mit welcher Intensität die Befragten dem Item zugestimmt haben.

Theoretisch erscheint ein Modell am besten begründbar, das davon ausgeht, dass die empirischen *Ja%* und *Nein%* den Zustimmungs- bzw. Ablehnungsanteilen auf einem latenten Urteilskontinuum entsprechen, auf dem die Verteilung der Antworten für jedes Item stets normal ist. Die Nichtnormalität der empirischen Verteilung entsteht dadurch, dass der Befragte seine Antwort auf die in wenige Kategorien gestufte Likertskala „stauchen" bzw. „rastern" muss.

Der praktische Nachteil dieses Modells ist der, dass es keine einfache Beziehung von *Ja%* und Mittelwert herstellt. Vielmehr besagt es, dass erst *Ja%* und *Nein%* <u>zusammen</u> die Information des Mittelwerts ergeben. Empirisch zeigt sich aber in den Abbildungen 1-4, dass *Ja%* und *Nein%* <u>jeweils</u> in einer engen, fast linearen Beziehung zum mittleren Skalenwert stehen. Diese Beziehung basiert – aus Sicht des Modells der latenten Normalverteilung – darauf, dass die Standardabweichung aller Items in den verschiedenen Studien jeweils sehr ähnlich ist. Diese Konstanz der Itemstreuung allerdings von vornherein als Modellspezifikation zu fordern, erscheint uns apriori weniger plausibel zu sein als die Annahme, dass *Ja%* und *Nein%* jeweils den wahren Zustimmungs- bzw. Ablehnungsanteilen entsprechen. Noch schwerer begründbar ist es, einen bestimmten Wert für diese Streuung zu postulieren.

Was also tun? Am sinnvollsten erscheint es uns, im jeweiligen empirischen Kontext ein Diagramm wie in den Abbildungen 1-4 zu erzeugen und damit zu prüfen, wie eng und welcher Art der Zusammenhang der *Ja%* und der Itemmittelwerte rein empirisch ist. Die obigen Untersuchungen zeigen, dass man hierbei durchaus optimistisch sein kann in dem Sinn, dass sich die *Ja%*-Statistiken dabei als statistisch solide Alternativen zu den Skalenmittelwerten erweisen werden. Wegen ihrer kommunikativen Vorteile sind sie dann die Statistik der Wahl. Bei diesem Vorgehen sieht man zudem, wo die Beziehung von *Ja%* und Mittelwert nichtlinear zu werden beginnt, bei welchen *Ja%*-Werten es also angezeigt ist, den Mittelwert zumindest als Hilfsgröße mitzuführen.

Schließlich noch eine eher technische Anmerkung. In den Abbildungen 1-4 sieht man, dass ein *Ja%*-Wert von 50 recht genau dem Mittelpunkt zwischen der Antwortkategorie „teils-teils" und der Antwortkategorie „stimme eher zu" entspricht. Dieser Mittelpunkt liegt bei 2,5 auf der 5-stufigen Skala (Abbildungen 1-3) und bei 3,5 auf der 7-stufigen Skala (Abbildung 4). Unterstellt man also eine kontinuierliche Intensität der Zustimmung oder Ablehnung des Items, dann zeigt sich hier die von der Likert-Skala geforderte diskrete Rasterung des Urteils so, wie oben in den theoretischen Überlegungen unterstellt: Der Übergang des Urteils von „teils-teils" zu „stimme eher zu" liegt fast genau in der Mitte des Intervalls zwischen den Skalenwerten dieser beiden Urteile, d.h., bei genauerer Betrachtung, etwas rechts davon.

Theoretisch lässt sich dieser Sachverhalt noch begründen. Für $p_0 = 0{,}5$ ist $\mu = z_0$ unabhängig von p_1. Da die Transformationsfunktion in der Regel bis $(z_0 + z_1)/2$ konvex, danach konkav verläuft, kann man sich grafisch leicht klarmachen, dass in diesem Fall der Erwartungswert von $T(X)$ mit wachsenden

$$\sigma = \frac{z_1 - z_0}{\Phi^{-1}(1 - p_1)}$$

d.h. mit wachsendem p_1 zunimmt, allerdings relativ schwach. Also ist insbesondere im Falle einer 5-er Likert-Skala mit $z_0 = 2{,}5$ und $z_1 = 3{,}5$ der Minimalwert von $E\ T(X)$ für σ=0 gegeben, was dann zur unteren Schranke $E\ T(X) = z_0$ führt. Bei großem Stichprobenumfang ist demnach in diesem Spezialfall als mittlerer Skalenwert ein Wert etwas größer als z_0 =2,5 zu erwarten. Analoges gilt auf der 7-stufigen Skala.

Was bleibt in der Zukunft zu untersuchen? Bei den vorliegenden realen Datensätzen konnte man von großen Stichproben ausgehen. In weiteren Untersuchungen muss gezeigt werden, wie robust sich das vorgestellte Transformationsmodell bei kleinen Stichproben verhält bzw. wie stark der Stichprobenfehler sich dann auswirkt. Die Wahl der Transformationsfunktion hat ebenfalls Auswirkung auf die Ergebnisse und bedarf einer genaueren Analyse. Schließlich sollte noch verfolgt werden, genau welche Beziehung die empirische Standardabweichung der Itemskalenwerte zum σ der latenten Urteilsverteilung hat, weil wir das Konstantsetzen von σ in unserem alternativen Modell vor allem aus der Beobachtung relativ ähnlicher Standardabweichungen der Skalenwerte zu begründen versucht haben. Erste Ergebnisse weisen auf einen nichtlinearen Zusammenhang hin. Zwischen den Standardabweichungen der transformierten Werte und den empirischen Standardabweichungen scheint es allerdings wieder näherungsweise einen linearen Zusammenhang zu geben.

5. Mathematischer Anhang

Es sei F_θ(x) eine auf $\Re$ definierte Verteilungsfunktion und T: $\Re \rightarrow (a,b)$ eine monoton steigende Funktion, die die reellen Zahlen in das Intervall (a,b) abbildet. Definiert man auf (a,b) eine Funktion F_1: $(a,b) \rightarrow \Re$ durch

$$F_1(z) = F_\theta\,(T^{-1}(z))$$

so ist F_1 eine Verteilungsfunktion.

Wir spezifizieren obige Aussage. Es sei $F_\theta(x)=\Phi(x|\mu,\sigma)$ die Verteilungsfunktion einer (μ,σ)-normalverteilten Zufallsvariablen und

$$T(x) = a + (b-a)F_0\left(\frac{F_0^{-1}(\frac{z_1-a}{b-a})(x-z_0) - F_0^{-1}(\frac{z_0-a}{b-a})(x-z_1)}{z_1 - z_0}\right)$$

mit einer auf $(-\infty,\infty)$ positiven streng monotonen Verteilungsfunktion $F_0(x)$. Dann ist für $a < z < b$

$$T^{-1}(z) = \frac{(z_1-z_0)F_0^{-1}\left(\frac{z-a}{b-a}\right) + z_0 F_0^{-1}\left(\frac{z_1-a}{b-a}\right) - z_1 F_0^{-1}\left(\frac{z_0-a}{b-a}\right)}{F_0^{-1}\left(\frac{z_1-a}{b-a}\right) - F_0^{-1}\left(\frac{z_0-a}{b-a}\right)}$$

und daher

$$F_1(z) = \Phi\left(T^{-1}(z)\,|\,\mu,\sigma\right) = \Phi\left(\frac{(z_1-z_0)F_0^{-1}\left(\frac{z-a}{b-a}\right) + z_0 F_0^{-1}\left(\frac{z_1-a}{b-a}\right) - z_1 F_0^{-1}\left(\frac{z_0-a}{b-a}\right)}{F_0^{-1}\left(\frac{z_1-a}{b-a}\right) - F_0^{-1}\left(\frac{z_0-a}{b-a}\right)}\,|\,\mu,\sigma\right)$$

Die transformierte Dichte lautet

$$f_1(z) = \frac{1}{\sigma}\varphi\left(T^{-1}(z)\,|\,\mu,\sigma\right)\cdot T^{-1'}(z)$$

$$= \frac{1}{\sigma}\varphi\left(\frac{(z_1-z_0)F_0^{-1}\left(\frac{z-a}{b-a}\right)+z_0F_0^{-1}\left(\frac{z_1-a}{b-a}\right)-z_1F_0^{-1}\left(\frac{z_0-a}{b-a}\right)}{F_0^{-1}\left(\frac{z_1-a}{b-a}\right)-F_0^{-1}\left(\frac{z_0-a}{b-a}\right)}\,|\,\mu,\sigma\right)\cdot\frac{(z_1-z_0)f_0^{-1}\left(\frac{z-a}{b-a}\right)}{F_0^{-1}\left(\frac{z_1-a}{b-a}\right)-F_0^{-1}\left(\frac{z_0-a}{b-a}\right)}\frac{1}{(b-a)}$$

Angenommen $F_0(x)$ ist die Laplace'sche Verteilungsfunktion mit Parametern λ und β, d.h. die Dichte $f_0(x)$ lässt sich schreiben als

$$f_0(x) = \frac{1}{2\beta}e^{-\frac{|x-\lambda|}{\beta}} \qquad \text{für} \quad -\infty < x < \infty \,.$$

Damit ist $T(x)$ (unabhängig von λ und β) gleich

$$T(x) = a + (b-a)\begin{cases} \frac{1}{2}e^{c(x-\frac{z_0+z_1}{2})} & \text{für} \quad x \leq \frac{z_0+z_1}{2} \\ 1-\frac{1}{2}e^{-c(x-\frac{z_0+z_1}{2})} & \text{für} \quad x > \frac{z_0+z_1}{2} \end{cases}$$

wobei

$$c = \frac{-2\ln\left(2\frac{b-z_1}{b-a}\right)}{z_1-z_0}$$

ist. Mit einigem Rechenaufwand erhält man mit

$$d = \frac{z_0+z_1}{2} - \mu$$

$$
\begin{aligned}
E(Z) = E(T(X)) &= \frac{1}{\sigma} \int_{-\infty}^{\infty} T(x)\varphi(x \mid \mu, \sigma)dx \\
&= a + \frac{b-a}{2} e^{-cd+0,5c^2\sigma^2} \Phi\left(\frac{d - c\sigma^2}{\sigma} \right) + (b-a)\left(1 - \Phi\left(\frac{d}{\sigma}\right)\right) \\
&- \frac{b-a}{2} e^{cd+0,5c^2\sigma^2} \left(1 - \Phi\left(\frac{d + c\sigma^2}{\sigma}\right)\right).
\end{aligned}
$$

Literatur

Bergler, R./Piwinger, M., 2000: Die Mitarbeiterbefragung als Instrument der Entwicklung von Strategien der Unternehmensentwicklung bei Vorwerck. S. 73-102 in: M. Domsch (Hrsg.), Handbuch Mitarbeiterbefragung. Heidelberg: Springer.

Borg, I., 1989: Zur Präsentation von Umfrageergebnissen. Zeitschrift für Arbeits- und Organisationspsychologie, 33, 90-95.

Borg, I., 2000a: Führungsinstrument Mitarbeiterbefragung: Theorien, Tools und Praxiserfahrungen. Göttingen: Verlag für angewandte Psychologie.

Borg, I., 2000b: Früh- versus Spätantworter. ZUMA-Nachrichten, 47, 7-19.

Borg, I., 2001: Mitarbeiterbefragungen. S. 373-396 in: H. Schuler (Hrsg.), Lehrbuch der Personalpsychologie. Göttingen: Hogrefe.

Bruennecke, K./Canisius, E., 1991: Open Line – Eine Mitarbeiterbefragung der Hewlett-Packard GmbH. S. 95-107 in: M. Domsch/A. Schneble (Hrsg.), Mitarbeiterbefragungen. Heidelberg: Physica Verlag.

Church, A.H./Waclawski, J., 2001: Organizational surveys. San Francisco: Jossey-Bass.

Edwards, J.E./Thomas, M.D./Rosenfeld, P./Booth-Kewley, S., 1997: How to conduct organizational surveys. Newbury Park, CA: Sage.

Folkman, J., 1998: Employee surveys that make a difference. Provo, UT: Executive Excellence Publishing.

Gabler, S./Borg, I., 1996: Unimodalität und Unimodalitätstests. ZUMA-Nachrichten, 38, 33-44.

Kraut, A.I., (Hrsg.), 1996: Organizational surveys: Tools for assessment and change. San Francisco, CA: Jossey-Bass.

Macey, W.H., 1996: Dealing with the data: collection, processing, and analysis. S. 204-232 in: A.I. Kraut (Hrsg.), Organizational surveys: Tools for assessment and change. San Francisco, CA: Jossey-Bass.

Nadler, D.A., 1977: Feedback and organization development: Using data-based methods. Reading, MA: Addison-Wesley.

Rogelberg, S.G./Church, A.H./Waclawski, J./Stanton, J.M., (im Druck): Organizational survey research: overview, the internet/intranet and present pratices of concern. In S.G. Rogelberg (Hrsg.), Handbook of methods in organizational and industrial psychology. London: Blackwell.

Tourangeau, R./Rips, L.J./Rasinski, K.A., 2000: The psychology of survey response. Cambridge, UK: Cambridge University Press.

Abbildung 1: Beziehung von *Ja%* (Punkte) und *Nein%* (Quadrate) für 94 Items aus der Mitarbeiterbefragung A.

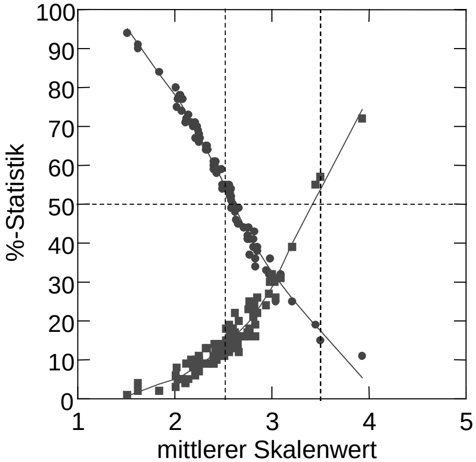

Abbildung 2: Beziehung von *Ja%* (Punkte) und *Nein%* (Quadrate) für 67 Items aus der Mitarbeiterbefragung B.

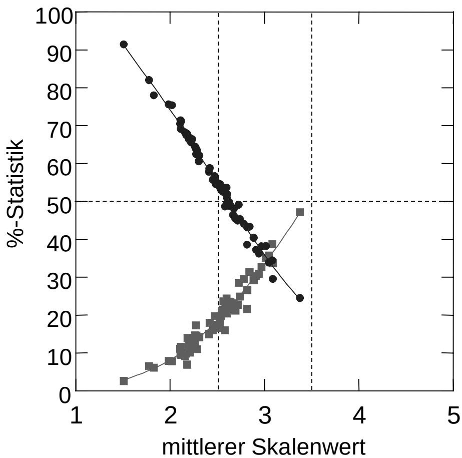

Abbildung 3: Beziehung von *Ja%* (Punkte) und *Nein%* (Quadrate) für 54 Items aus der Mitarbeiterbefragung C.

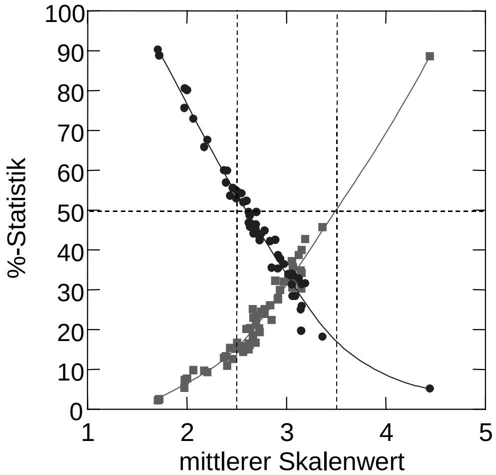

Abbildung 4: Beziehung von *Ja%* (Punkte) und *Nein%* (Quadrate) für 85 Items aus der Mitarbeiterbefragung D.

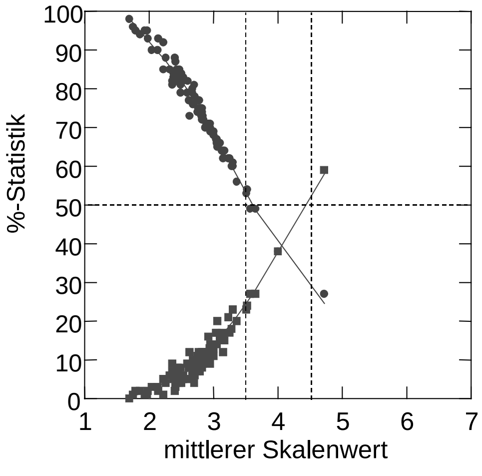

Abbildung 5: Illustration der Abbildung der latenten Antwortverteilung (oben links) auf die Likertskala (oben rechts) über die logistische Verteilungsfunktion (unten links) und eine 5-stufige Rasterung (letzter Schritt: Mitte rechts) unter Erhalt der Zustimmungs-, Gemischt- und Ablehnungsanteile.

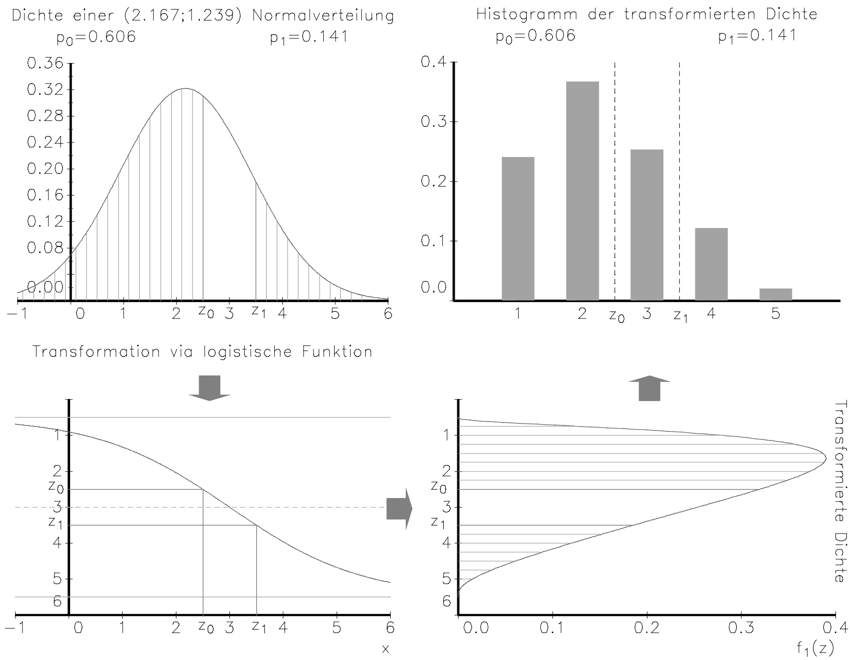

Abbildung 6: Mittlere (empirische) Skalenwerte relativ zu den aus den *Ja%*- und *Nein%*-Anteilen per Modell berechneten erwarteten Urteilswerten auf der gestauchten Skala (Studie A); die Regressionsgerade lautet $\hat{y} = 0{,}04 + 1{,}01 \cdot x$.

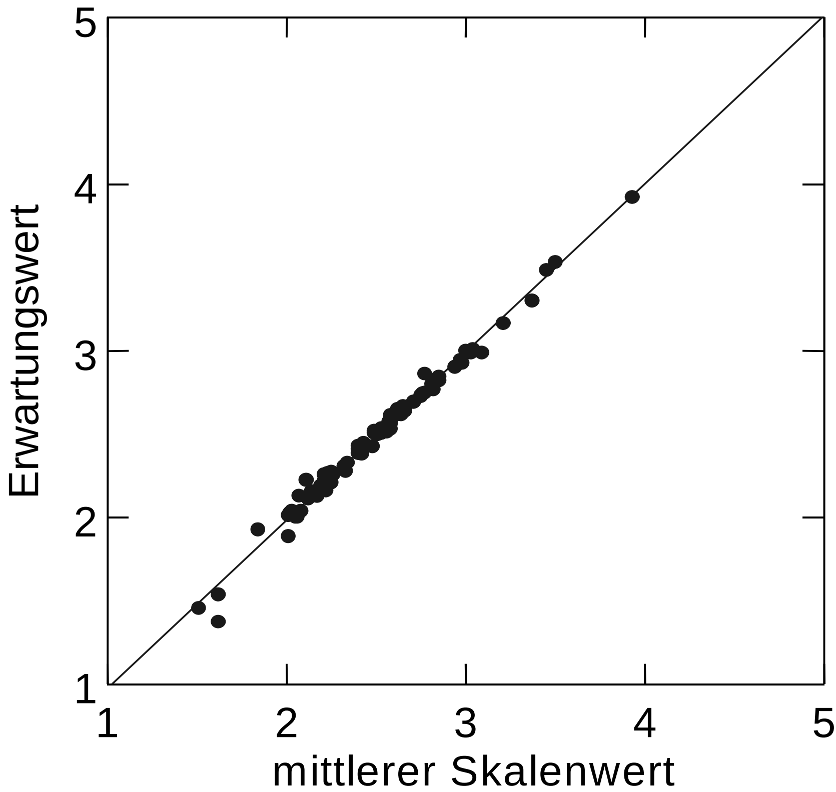

Abbildung 7: Beziehung von p_0 und dem Erwartungswert von Z in der gestauchten Verteilung bei σ=1 (5-er Skala).