Friedrich-Alexander-Universität
Erlangen-Nürnberg
Wirtschafts-und Sozialwissenschaftliche Fakultät

Diskussionspapier
62 / 2004

Skalentypen und Statistik

– Ein Kommentar zu Velleman & Wilkinson (1993) –

Ingo Klein

Lehrstuhl für Statistik und Ökonometrie
Lehrstuhl für Statistik und empirische Wirtschaftsforschung
Lange Gasse 20 · D-90403 Nürnberg

SKALENTYPEN UND STATISTIK
– Ein Kommentar zu Velleman & Wilkinson (1993) –

Ingo Klein

Lehrstuhl für Statistik und Ökonometrie
Universität Erlangen-Nürnberg
ingo.klein@wiso.uni-erlangen.de

Abstract

Velleman & Wilkinson (1993) give a lot of examples to show that the hierarchy of scalers introduced by Stevens is misleading. We will show that these examples gain are misleading if we accept that most real data sets stem come derived measurement. For derived measurement the absolute scale is very important. Count data, ranks, percent data, standardized variables all have an absolute scale.

Keywords: Scale type, permissible transformations, data transformation

1 Ausgangspunkt

Seit fast fünfzig Jahren wird in den Sozialwissenschaften intensiv diskutiert, ob die Wahl statistischer Methoden von dem Skalenniveau der zu analysierenden Variablen abhängt. Zumbo & Zimmermann (1993) sprechen von einer ,,levels of measurement controversy" oder ,,measurement–statistics debate". Ausgangspunkt dieser Kontroverse sind die Arbeiten des Psychologen S.S. Stevens aus den Jahren 1946 und 1951. In der ersten Arbeit gibt er eine Hierarchie von Skalentypen von der Nominal- über die Ordinal- und der Intervallskala zur Verhältnisskala an, wobei sich die Skalentypen durch unterschiedlich ,,große" Mengen zulässiger Skalentransformationen unterscheiden. In der zweiten Arbeit postuliert er, dass nur statistische Methoden ,,erlaubt" sind, die Invarianzeigenschaften bezüglich der zulässigen Skalentransformationen der zu analysierenden Variablen besitzen sollen. So ist das arithmetische Mittel bei Vorliegen ordinalskalierter Daten ,,nicht erlaubt", während der Median die notwendige Invarianzeigenschaft besitzt.

Da in den Wirtschafts- und Sozialwissenschaften häufig ordinalskalierte Variablen vorliegen, die ordnungserhaltende (d.h. streng monoton zunehmende) Abbildungen als zulässige Transformationen besitzen, führte Stevens' Definition ,,erlaubter" statistischer Methoden zu einer Blüte der sog. nichtparametrischen Statistik, die statt von den originalen Messwerten von Rangzahlen der Beobachtungen ausgeht. Dagegen sind die klassischen Verfahren der parametrischen Statistik, wie z.B. der t-Test auf Mittelwertdifferenzen bei unbekannten aber gleichen Varianzen und der F-Test auf Varianzquotienten, nur dann ,,erlaubt", wenn zumindest intervallskalierte Daten vorliegen.

Eine weitere Konsequenz der unterschiedlichen Skalenniveaus ist die Frage, welche Funktionen in der Lage sind, ohne Rückgriff auf Dimensionskonstanten, Variablen verschiedener Skalenniveaus ineinander so abzubilden, dass die sich ergebende Funktion ,,erlaubt" im Sinne von Stevens ist. Diese Fragestellung ist für Verhältnisskalen von Luce (1959, 1964) gel"st worden, wobei er als einzige stetige Lösung der sich ergebenden Funktionalgleichung eine Potenzfunktion erhält. Die Konsequenz dieses Ergebnis ist beispielsweise, dass sich die Urteile verschiedener Experten nur anhand des geometrischen Mittels zusammenfassen lassen, sofern die Expertenurteile verhältnisskaliert sind. Für Intervallskalen existiert keine einzige nicht-konstante stetige Möglichkeit der Zusammenfassung.

Dass eine derart starke Einschränkung der Vielfalt statistischer Methoden nicht ohne Widerspruch bleiben konnte, liegt auf der Hand. Ein entscheidender Kritikpunkt besteht darin, dass man realen Daten nicht ,,ansehen" kann, welches Skalenniveau sie besitzen. In Stevens' Theorie muss das Skalenniveau nämlich a priori bekannt sein. Durch die Repräsentationstheorie des Messens, wie sie z.B. in Pfanzagl (1971), Krantz et al. (1971), Suppes et al. (1989) und Luce et al. (1990) dargestellt wird, gelingt es, Skalenniveaus als Ergebnis eines Messvorganges durch strukturerhaltende Abbildung herzuleiten, wie sie aus der Theorie des ordinalen bzw. kardinalen Nutzens bekannt sind. In der Theorie der Skalentypen von Alper (1985, 1987) und Narens (1981) ist weiterhin gezeigt worden, dass unter gewissen Annahmen, nur die von Stevens vorgeschlagenen Skalenniveaus auftreten können. Eine Begründung der von Stevens erhobenen Forderung der Invarianz statistischer Methoden bezüglich zulässiger Skalentransformationen steht allerdings noch aus, obwohl Narens (1988) diese Invarianzforderung in die šberlegungen des Erlanger Programms von F. Klein aus dem Jahre 1872 für eine Grundlegung der Geometrie eingebettet hat. Erste Ansätze einer Theorie der ,,erlaubten" Statistiken gehen auf Klein (1984) und (1994) zurück.

Diese fehlende messtheoretische Begründung und eine automatische Anwendung des Invarianzkriteriums von Stevens in der Computer-gestützten Auswahl statistischer Methoden haben Velleman & Wilkinson (1993) veranlasst, die genannte Kontroverse aufzugreifen und ihre Arbeit programmatisch ,,Nominal, Ordinal, Interval, and Ratio Typologies Are Misleading" zu betiteln.

Im folgenden wird versucht, die von Velleman & Wilkinson vorgebrachten Kritikpunkte zu systematisieren und weitestgehend zu entkräften. Es wird dem allgemein und speziell auch in statistischen Expertensystemen praktizierten Vorgehen der Methodenwahl das Wort geredet, ohne die Probleme wissenschaftstheoretisch kaum fundierbarer, das Skalenniveau betreffender a priori Annahmen in Abrede zu stellen.

2 Kritik von Velleman & Wilkinson

Es lassen sich wissenschaftstheoretisch fundierte und pragmatisch orientierte Kritikpunkte unterscheiden. Da jedoch zur Zeit die wissenschaftstheoretischen Grundlagen der Theorie des Messens intensiv diskutiert werden und eine Betrachtung dieser Diskussion den Rahmen sprengen würde, sollen im folgenden nur die mehr pragmatischen Kritikpunkte angesprochen werden. Der Generaltenor der pragmatischen Kritik lautet: "It doesn't work." Velleman & Wilkinson geben dazu eine Vielzahl von Beispielen an, die ihre kritischen Thesen belegen sollen. Im folgenden werden diese Beispiele unter dem in dem Artikel von Velleman & Wilkinson verwendeten šberschriften aufgegriffen und diskutiert.

Alternative Skaleneinteilungen

Mosteller & Tukey (1977) geben die in Tabelle 1 aufgeführten Datentypen an, auf die sich Velleman & Wilkinson berufen, wobei die letzte Spalte, die von uns zugefügt wurde, die zugehörigen Skalentypen identifiziert. Diese Datentypen scheinen zunächst keine Restriktionen für statistische Methoden zu implizieren. Obwohl dies nicht von Mosteller & Tukey beabsichtigt ist, kann jedoch das jeweilige Skalenniveau sämtlicher Datentypen bestimmt werden, wie die Angaben der letzten Spalte zeigen. Dabei spielt die absolute Skala eine entscheidende Rolle. Sie ist dadurch gekennzeichnet, dass lediglich die identische Abbildung als zulässige Transformation auftritt. Grob gesprochen gibt es somit nur eine Möglichkeit der Messung absolutskalierter Größen. Zählergebnisse liefern das wohl prominenteste Beispiel für Messwerte einer absoluten Skala. Die absolute Skala tritt vornehmlich auf, wenn eine sog. abgeleitete Messung vorliegt. Dabei wird aus mehreren fundamental gemessenen Größen, durch eine zumeist definitorische Vorschrift, eine weitere Größe abgeleitet. Ein oft zitiertes Beispiel ist die physikalische Dichte, die sich als Quotient aus Masse und Volumen errechnet. Durch abgeleitete Messung ergeben sich insbesondere die beiden Datentypen ”Anteile” und ”Ränge”.

Namen	Ausprägungen	Skalentyp
Grade	geordnete Labels	Ordinalskala
Ränge		absolute Skala
Anteile	Prozentzahlen	absolute Skala
Zählergebnisse	nichtnegative Integerzahlen	absolute Skala
Beträge	nichtnegative reelle Zahlen	Intervallskala
”Balances”	unbeschränkte reelle Zahlen	Intervallskala

Tabelle 1: Datentypen nach Mosteller & Tukey (1977) und ihre Skalentypen

Diese Datentypen scheinen zunächst keine Restriktionen für statistische Methoden zu implizieren. Obwohl dies nicht von Mosteller & Tukey beabsichtigt ist, kann jedoch das jeweilige Skalenniveau sämtlicher Datentypen bestimmt werden, wie die Angaben der letzten Spalte zeigen. Dabei spielt die absolute Skala eine entscheidende Rolle. Sie ist dadurch gekennzeichnet, dass lediglich die identische Abbildung als zulässige Transformation auftritt. Grob gesprochen gibt es somit nur eine Möglichkeit der Messung absolutskalierter Größen. Zählergebnisse liefern das wohl prominenteste Beispiel für Messwerte einer absoluten Skala. Die absolute Skala tritt vornehmlich auf, wenn eine sog. abgeleitete Messung vorliegt. Dabei wird aus mehreren fundamental gemessenen Größen, durch eine zumeist definitorische Vorschrift, eine weitere Größe abgeleitet. Ein oft zitiertes Beispiel ist die physikalische Dichte, die sich als Quotient aus Masse und Volumen errechnet. Durch abgeleitete Messung ergeben sich insbesondere die beiden Datentypen ”Anteile” und ”Ränge”. Wenn z.B. m eine Ordinalskala ist und A eine Objektmenge, dann ist der Rang eines Objektes a definiert als die Anzahl der Objekte a_i aus A, deren Messwerte $m(a_i)$ kleiner oder gleich $m(a)$ sind. Diese Definition legt den Begriff ”Rang” eindeutig fest. Bezüglich

Angaben von Anteilen (z.B. in Prozent oder in Promille) stellen Velleman & Wilkinson fest, dass sich diese den gängigen zulässigen Skalentransformationen entziehen, da ihre Werte auf ein bestimmtes Zahlenintervall beschränkt sind. Dies ist aber kein Problem, da die Identität die einzige zulässige Transformation ist, sofern festgelegt ist, worauf sich die Anteile summieren sollen (z.B. 1, 100 oder 1000). Offensichtlich ist man in der Lage, diesen praxisrelevanten Datentypen ohne Schwierigkeiten ihr Skalenniveau anzusehen.

2.0.1 Stevens' Skalentypen sind unzureichend, um realen Datenskalen zu beschreiben

Mit einem weiteren Beispiel wird von Velleman & Wilkinson die Behauptung aufgestellt, dass häufig in der Praxis auftretenden Datentypen nicht notwendigerweise das Skalenniveau anzusehen ist.

Es werden vier Krankheitssymptome betrachtet, wobei jeweils mit 1 bzw. 0 festgehalten wird, ob ein System vorliegt oder nicht. Daraus ergibt sich die in der folgenden Tabelle dargestellte partielle binäre Ordnung. Die letzte Spalte gibt an, wie viele Symptome jeweils festzustellen sind.

			1111	4
		1110	0111	3
	1100	0110	0011	2
1000	0100	0010	0001	1

Velleman & Wilkinson behaupten, dass die horizontale Dimension eine Nominalskala und die vertikale Dimension (d.h. die letzte Spalte) eine Verhältnisskala angeben. Dieses Beispiel passt aber sehr wohl in die Terminologie von Stevens, wenn man wiederum absolute Skalen im Falle abgeleiteter Messung betrachtet. In dem genannten Beispiel handelt es sich nämlich um die Weiterverarbeitung eines multivariaten kategorialen Datensatzes (Nominalskala) durch eine abgeleitete Absolutskala. Mittels abgeleiteter Messung wird lediglich die Anzahl der Symptome festgestellt, d.h. gezählt. Zählprozesse führen aber, wie oben angegeben, stets zu absolutskalierten Merkmalen.

Stevens' Skalentypen sind keine fixen Eigenschaften der Daten

Die wohl schwerwiegendste Kritik von Velleman & Wilkinson basiert auf der Überzeugung, dass das Skalenniveau von Daten nicht a priori feststellbar ist, sondern von der jeweiligen Fragestellung abhängt. Velleman & Wilkinson belegen dies durch zwei Beispiele:

Bei einem Empfang der American Statistical Association werden am Eingang fortlaufend nummerierte Tickets ausgegeben. Aus der Nummer, die ein Teilnehmer erhalten hat, können in Abhängigkeit von der Fragestellung die unterschiedlichsten Informationen abgelesen werden:

1. Die Ticketnummern sind zunächst andere Namen der Teilnehmer und damit Messwerte einer nominalskalierten Variablen. Jeder Teilnehmer ist aufgrund seiner Ticketnummer eindeutig identifizierbar.
2. Im Rahmen einer Verlosung wird verkündet, dass die Ticketnummer 126 gewonnen hat. Der Teilnehmer mit der Ticketnnummer 56 schließt, dass er zu früh gekommen ist, um gewinnen zu können. 56 bzw. 126 lassen sich somit als Ausprägungen eines ordinalskalierten Merkmals auffassen.
3. Aus der Gleichmäßigkeit und der Geschwindigkeit der Ankunft der Teilnehmer kann der Teilnehmer mit der Ticketnummer 56 die Zeit errechnen, die er zu früh gekommen ist, um den Preis zu gewinnen. Velleman & Wilkinson schließen, dass in dieser Situation die Ticketnummer " an interval-scale-value" ist.
4. Schließlich kann die Ticketnummer 126 mit der Anzahl der im Raum befindlichen Personen verglichen und z.B. festgestellt werden, dass Personen bereits gegangen sein müssen.

Die Art der Vergabe der Ticketnummern erlaubt, diese als absolutskalierte Rangzahlen zu interpretieren. Die Ticketnummer eines bestimmten Teilnehmers gibt an, wie viele Teilnehmer vor ihm gekommen sind. Absolutskalierte Werte lassen sich auf beliebige Weise weiterverarbeiten. Die absolute Skala besitzt das höchste Skalenniveau, so dass sie insbesondere auch eine Verhältnis-, Intervall- oder Ordinalskala darstellt. Es ist deshalb nicht verwunderlich, wenn Abstandsvergleiche und Rangfolgen sinnvoll interpretiert werden können. Trotzdem ist im vierten Fall die Ticketnummer kein "interval-scale-value". Eine mit ihrer Hilfe errechnete Zeitangabe (z.B. 126 Sekunden, wenn die Teilnehmer im Sekundentakt ankommen) ist intervallskaliert. Diese Zeitangabe ist das Ergebnis einer abgeleiteten Messung. Nicht die unterschiedlichen Fragestellungen sind das entscheidende, sondern die verschiedenen Informationsgrade, die bei der Messung berücksichtigt werden können. Diese reichen von der Kenntnis der Art und Weise, in welcher Reihenfolge die Ticketnummern ausgegeben werden, bis zur Kenntnis der Geschwindigkeit, mit der die Gäste ankommen.

Die Testdatensätze der meisten Statistikpakete verfügen über ein Automobilbeispiel, in dem als Merkmal die Zylinderzahl erfasst wird. Velleman & Wilkinson betonen, dass die Zylinderzahl die Automobile zwar klassifiziert, aber auch eine Ordnung des Benzinverbrauchs entsprechend der Zylinderzahl möglich ist. Wiederum ist nicht die Fragestellung für den Skalentyp entscheidend. Die Anzahl der Zylinder basiert auf einem Zählprozess, der eine absolute Skala definiert. Grob gesprochen kann mit den Zylinderangaben jede Rechenoperation durchgeführt werden, und damit insbesondere auch solche, die zur Beantwortung interessanter Fragestellungen gehören.

Datentransformationen

Bekanntlich setzen viele statistische Verfahren voraus, dass die zu analysierenden Daten eine Reihe von Annahmen erfüllen. So wird für einen t-Test für den Vergleich zweier Mittelwerte die Varianzhomogenität und die Normalverteilung benötigt. Andere Verfahren erfordern die Symmetrie der Verteilung der Grundgesamtheit oder die Linearität des Zusammenhangs von zwei oder mehr Variablen. Diese Annahmen sind jedoch häufig nicht erfüllt, so dass nach Regeln gesucht wird, wie die Daten zu transformieren sind, damit die Annahmen wenigstens nahezu gelten. So haben Box & Cox (1962) eine Klasse von Transformationen vorgeschlagen, die streng monoton zunehmend und im Regelfall nicht-linear sind. Wurzel- und Logarithmustransformationen gehören insbesondere zur Klasse dieser Box-Cox-Transformationen. Velleman & Wilkinson betonen, dass diese Transformationen offenbar zulässige Transformationen einer Ordinalskala sind, obwohl die zu transformierenden Daten zumindest Messwerte einer Intervallskala darstellen, da ansonsten Konzepte, wie Linearität, Symmetrie, Homogenität und Normalverteiltheit keinen Sinn machen.

Eine Vermischung von Datentransformationen und zulässigen Skalentransformationen ist u.E. aber höchst fragwürdig. Datentransformationen sind ein bewusst angewendetes Hilfsmittel zur Sicherstellung notwendiger Voraussetzungen statistischer Methoden, während zulässige Skalentransformationen Ausdruck eines bestimmten bei der Messung substanzwissenschaftlicher Begriffe verfügbaren Informationsgehalts darstellen. Nicht-lineare Datentransformationen (insbesondere Logarithmen) können diesen Informationsgehalt transparenter machen, wie die von Hoaglin (1988) aufgelisteten Beispiele zeigen.

Das eigentliche Problem der Datentransformationen liegt nicht darin, ob sie auch zulässige Skalentransformationen sind, sondern in der Fragestellung, ob sich die mittels transformierter Daten gewonnenen Ergebnisse auf die Originaldaten übertragen lassen. Dies ist aber im Regelfall nicht möglich.

Kontinuum von (robusten) Statistiken

Plakativ führt die von Stevens eingeführte Definition ,,erlaubter" statistischer Methoden zu der Feststellung, dass der Median zur Ordinalskala und das arithmetische Mittel zur Intervallskala ,,gehören".

Velleman & Wilkinson gehen nun auf sog. L-Statistiken ein, die eine Linearkombination von Ordnungsstatistiken, d.h. der Größe nach geordneten Daten, sind. Insbesondere der getrimmte Mittelwert ist eine L-Statistik, wobei ein bestimmter festgelegter Prozentsatz der kleinsten und der größten Daten bei der arithmetischen Mittelung nicht berücksichtigt werden. Offensichtlich sind sowohl der Median als auch das gewöhnliche arithmetische Mittel Spezialfälle des getrimmten Mittelwertes. Velleman & Wilkinson (1994, S. 70) schließen nun aus dieser Tatsache:

> ,,This estimator is thus on a continuum between Stevens's ordinal and interval categories. ... In some sense, the trimmed mean seems to classifiy the data into a central body "interval" values and outlying tails of "ordinal" values."

Dieser Auffassung ist entschieden entgegenzutreten. Getrimmte Mittel werden angewendet, wenn Ausreißer oder extreme Beobachtungen vorliegen. Um diese zu bestimmen, wird ein Abstandsbegriff benötigt, der zumindest intervallskalierte Daten voraussetzt. Getrimmte Mittel sind dagegen vorteilhaft, wenn die Daten aus einer Grundgesamtheit stammen, deren Verteilung im Zentrum einer Normalverteilung und in den Rändern einer anderen Verteilung entspricht, die Ausreißer mit einer größeren Wahrscheinlichkeit zulässt als die Normalverteilung. Dann sind getrimmte Mittel aus Verteilungs-, aber nicht aus Skalenüberlegungen "geeignete" Statistiken.

Weiterhin ist es problematisch, von einem "Kontinuum" von Skalen zwischen Ordinal- und Intervallskala auszugehen. Unter gewissen Mächtigkeitsannahmen haben Narens (1981) bzw. Alper (1985, 1987) gezeigt, dass es keine echten Zwischentypen geben kann.

Schließlich hat Klein (1986) nachgewiesen, dass Statistiken, die über Robustheitseigenschaften bezüglich Verteilungsannahmen verfügen, auch solche bezüglich Skalenannahmen besitzen, so dass z.B. die Anwendung eines arithmetischen Mittel auch dann noch „erlaubt" ist, wenn die Annahme intervallskalierter Daten leicht verletzt ist.

Der Skalentyp ist keine präzise Kategorie

Velleman & Wilkinson zitieren Tukey (1961), der ausgeführt hat, dass fehlerbehaftete, eigentlich intervallskalierte Daten, wobei der Fehler von der Größe des Datums abhängt, keiner Intervallskala mehr folgen. Insbesondere sind dann Hypothesen über die exakte Gleichheit von Mittelwerten oder Varianzen fragwürdig. In dieser Situation sind zumeist Rangverfahren problemlos anwendbar, wobei die allerdings fehlerbehaftete Information über die Datenabstände nicht berücksichtigt wird.

Dazu ist anzumerken, dass zumindest für nicht allzu kleine Stichproben Rangverfahren ähnlich gute statistische Eigenschaften wie die klassischen parametrischen Verfahren besitzen, die ihre Optimalität zumeist aus der Normalverteilungsannahme ableiten. Weiterhin wird auch in der Theorie des Messens intensiv an probabilistischen Ansätzen gearbeitet, wie die Monographie von Falmagne (1985) zeigt, so dass von dort aus Implikationen für die Statistik zu erwarten sind.

3 Schlussfolgerung

Velleman & Wilkinson haben zwar mit einer Vielzahl von Beispielen versucht, die von Stevens eingeführte Hierarchie von Skalen und ihre Implikationen für die Statistik in Frage zu stellen. Die von ihnen angegebenen pragmatischen Kritikpunkte scheinen allerdings wenig stichhaltig zu sein, wenn berücksichtigt wird, dass viele reale Datensätze aus einer abgeleiteten Messung stammen und häufig zu einer absoluten Skala gehören.

Annahmen über den Skalentyp der Daten sind Annahmen über deren standardisierten Informationsgehalt. Das Kriterium von Stevens über ,,erlaubte" statistische Methoden stellt lediglich sicher, dass für eine sinnvolle Interpretation der Ergebnisse dieser Methoden nicht sehr viel mehr Information in den Daten hätte enthalten sein müssen, als tatsächlich vorliegt. Es ist sozusagen ein ,,Kriterium der Widerspruchsfreiheit", das die Kompatibilität von Dateninformation und Ergebnisinterpretation sichert und damit ein kleiner Teil eines sicher nicht vollständig zu explizierbaren komplexen Bedeutsamkeitsbegriffs ist. In diesem Sinne ist die Zuordnung von Methoden zu Skalentypen und damit standardisierter Information auch in statistischen Expertensystemen zu befürworten.

Literatur

[1] Alper, T.M. (1985). A note on real measurement structures of scale type $(m, m+1)$. *Journal of Mathematical Psychology* **29**, 73-81.

[2] Alper, T.M. (1987). A classification of all order-preserving homeomorphism groups of the reals that satisfy finite uniqueness. *Journal of Mathematical Psychology* **31**, 135-154.

[3] Box, G.E.P. & Cox, D.R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistical Society, Series B* **26**, 211-252.

[4] Falmagne, J.C. (1985). *Elements of psychophysical theory*. Springer: New York.

[5] Hoaglin, D.C. (1988). Transformation in everyday experience. *Chance* **1**, 40-45.

[6] Klein, F. (1921). *Gesammelte Abhandlungen I.* Berlin.

[7] Klein, I. (1984). *Das Problem der Auswahl geeigneter Maßzahlen in der deskriptiven Statistik.* Physica-Verlag: Würzburg.

[8] Klein, I. (1986). Qualitative Robustheit statistischer Maßzahlen. *Allgemeines Statistisches Archiv* **48**, 158-169.

[9] Klein, I. (1994). *Mögliche Skalentypen, invariante Relationen und wissenschaftliche Gesetze.* Vandenhoeck & Ruprecht: Göttingen.

[10] Krantz, D.H., Luce, R.D., Suppes, P. & Tversky, A. (1971). *Foundations of measurement. Vol 1: Additive and polynomial representation.* Academic Press: New York, London.

[11] Luce, R.D. (1959). On the possible psychophysical laws. *Psychological Review* **66**, 81-95.

[12] Luce, R.D. (1964). A generalization of a theorem of dimensional analysis. *Journal of Mathematical Psychology* **1**, 278-284.

[13] Luce, R.D., Krantz, D.H., Suppes, P. & Tversky, A. (1990). *Foundations of measurement. Vol 3: Representation, axiomatization, and invariance.* Academic Press: New York, London.

[14] Mosteller, F. & Tukey, J.W. (1977). Data analysis and regression. Addison-Wesley: Boston.

[15] Narens, L. (1981). On the scales of measurement. *Journal of Mathematical Psychology* **24**, 249-275.

[16] Narens, L. (1988). Meaningfulness and the Erlanger programm of Felix Klein. *Mathématiques Informatique et Science Humaines* **101**, 61-72.

[17] Pfanzagl, J. (1971). *Theory of Measurement.* Physica-Verlag: Würzburg, Wien.

[18] Stevens, S.S. (1946). On a theory of scales and measurement. *Science.* **103**, 677-680.

[19] Stevens, S.S. (1951). Mathematics, measurement and psychophysics. In S.S. Stevens (ed.), *Handbook of experimental psychology*. Wiley: New York, 1-49.

[20] Suppes, P., Krantz, D.H., Luce, R.D. & Tversky, A. (1989). *Foundations of Measurement. Vol 2: Geometrical, Threshold, and Probabilistic Representations.* Academic Press: New York.

[21] Tukey, J.W. (1961). Data analysis and behavioral science or learning to bear the quantitative man's burden by shunning badmandments. In: Jones, L.V. (1986). *The Collected Works of John W. Tukey, Vol. III.* Wadsworth: Belmont, 391-484.

[22] Velleman, P.F. & Wilkinson, L. (1993). Nominal, ordinal, interval, and ratio typologies are misleading. *American Statistician* **47**, 65-72.

[23] Zumbo, B.D. & Zimmermann, D.W. (1993). Is the selection of statistical methods governed by level of measurement. *Canadian Psychology* **43**, 390-400.