Influences of Measurement Theory on Statistical Analysis & Stevens' Scales of Measurement

Mehtap ERGÜVEN*

Abstract

In this study, progress and improvement of measurement theory is presented and mainly S.S. Stevens' opinions and implementations on types of scales are discussed. Fundamentals of Measurement Theory, its historical evolution, basic goals of the measurement theory are described and summarized. Influences of measurement theory on statistical analysis and relation between Stevens' scale types and statistical methods are introduced. Types of scales are discussed in detail. Importance of preparing relevant data to evaluate real sociological facts is emphasized.

Keywords: Measurement theory, Stevens' types of scales, nominal scale, ordinal scale, interval scale, and ratio scale

Introduction

Finding real aspects of data, which are most suitable and important for us and using those aspects to construct a model of a concept, is a very significant issue in measurement. Measurement theory is a branch of applied mathematics, which together with statistics allows us to validly define measurements and metrics and implement statistical analysis on our data.

Measurement is an essential concept in science. Conclusions of empirical studies are based on values measured on research objects. The purpose of measurement is to differentiate abilities across people (Suen, Principles of Test Theories, 1990). During life, we are tested many times. Different kinds of questionnaires, rating scales, examinations, and some other measurement methods are used to evaluate and clarify ability, knowledge, or level of test takers. It is therefore crucial to assess the objectivity, reliability, and quality of measurements.

Studies on measurements start in 19th century, taking roots in works of French and German psychiatrists that verified the influence of mental diseases in motor, sensorial, and behavioral-cognitive skills, and those of English researchers in the field of genetics, which highlighted the importance of measuring individual differences with the use of well-defined methodologies. (E.P. Araujo,, D. F. Andrade, S. Bortolotti, 2009)

The 'classical' school of measurement was developed in physics and other sciences by the end of the nineteenth century. In the classical view, measurement discovered a numerical relationship between a standard object and the one measured. The property was seen as inherent in the object. This viewpoint is deeply ingrained in our language and society (Chrisman, 1998).

Measurement, the assignment of numbers to objects to represent their gradual properties, can be 'derived' or 'fundamental'. In derived measurement, we obtain the desired value of a magnitude for an object from other values we already have and which we related with the unknown value in a specific way. Derived measurement is by far the most common kind of measurement in scientific practice, but it is clear that, measurement cannot always be derived. Fundamental measurement is essential for it is 'where everything begins'. In fundamental, or direct, measurement we obtain the desired values with no previous measurements at all directly from qualitative empirical data (Diez).

Objectivity of measurement

Measurement theory enables us to analyze and manipulate data validly. Usage of statistics and probability helps to clarify quantitatively possible variances, errors. Main target in measurement is obtaining meaningful and objective results. When we compare properties of objects not all of our quantitative statements are objective. In the comparison of some properties of diamond and chalk, we can have a look to the following statements: The quotient of their masses equal to 100, the quotient of their temperatures is 2, and the quotient of hardness is 0.1. Only the first statement expresses something that depends on objects' properties, but other statements depend on these properties and on the conventions adopted in the construction of the measurement scale. For this reason, we can say that the first statement is objective, while the others are not. If the temperature is measured in degrees Celsius, the statement is true, but it is false with respect to degrees Fahrenheit. Although the statement that chalk's temperature is twice diamond's is not objective in this sense, the statement that the difference between the chalk's temperature at noon and at midnight is five times the difference between diamond's temperature at noon and at midnight is objective. This raises a question of "why some quantitative statements are objective and others are not". Measurement theory answers this question, by investigating the conditions that make measurement possible and by studying the extent to which we can use the measures obtained to make objective statements about objects (Diez).

* Lecturer, Faculty of Computer Technologies and Engineering, International Black Sea University, Tbilisi, Georgia.
E-mail: merguven@ibsu.edu.ge

Historical Development of Measurement Theory

According to Boumans (2005) Maxwell indicates that: "All the mathematical sciences are founded on relations between physical laws and laws of numbers, so that the aim of exact science is to reduce the problems of nature to the determination of quantities by operations with numbers". (Boumans, 2005)

The theory of measurement is an interdisciplinary subject that grew out of the attempt to put the foundations of measurement on firm mathematical foundation. Measurement Theory is the result of two different complementary research traditions.

The first begins with Helmholtz (1887). Helmholtz took up Maxwell's view and continued to think in this direction; and continuous with Hölder and Campbell, and focuses on comparative combinatorial systems and real morphisms. Helmholtz was the first to ask the main question on how fundamental measurement is possible. He was also the first one to answer it by providing a set of conditions that the system must satisfy. He did not demonstrate that his conditions were sufficient for numerical representation. This was achieved by Hölder (1901). He studied the necessity/sufficiency of a set of conditions for the numerical representation of a qualitative comparative-combinatorial system.

Hölder gives seven conditions, or axioms, that the domain D of objects, the qualitative relation greater than or equal to $\succsim$ and the qualitative operation **o** must satisfy and demonstrates that these conditions are jointly sufficient for there to be an isomorphism from < D, $\succsim$,**o** > onto < Re+, ≥, +>; that is for there to be a 1-1 mapping f:D → Re+ from the domain of objects into the positive real numbers so that

i) a $\succsim$ b iff f(a) ≥f(b)

ii) aob $\sim$ c iff f(a)+f(b)= f(c).

In this sense numbers represent magnitudes, since numbers assigned to objects are such that qualitative $\succsim$-facts among objects are 'replicated' by quantitative ≥-facts about the assigned numbers, and the same goes for qualitative o-facts and +-facts. This result is known as Hölder's Theorem.

After Hölder, Huntington (1902) presented similar results. Wiener (1921) was also heading in the same direction.

The second tradition originates in the work of Stevens and his collaborators on scale types, transformations, and invariance (Diez).

The effects of fundamental measurement theory for statistical analyses continue to be discussed. On one side is the view that the scale on which a set of measurements lies determines the type of statistical treatments that are suitable for application to the measurements. The opposing view is that there is no relation holding between the measurement scale and statistical procedures; essentially anything goes, relative to the measurement stipulations. (James Townsend , Gregory Ashby, 1984).

Representational Theory

First aspect of fundamental measurement is the representation theorem. The representational theory provides for the assignment of numbers to the empirical objects in such a way that interesting qualitative empirical relations among the objects are reflected in the numbers themselves as well as in important properties of the number system. Often in the case of infinite set of objects, one may simply prove that the assignment exists or state a method by which the numbers can be assigned (James Townsend , Gregory Ashby, 1984).

In measurement theory scales (assignments) are identified with homomorphisms. Formally, an admissible transformation of a scale is a transformation of numbers assigned so that one gets a homomorphism (Roberts, 2009).

In the formal representational theory this is expressed as: Take a well-defined, non-empty, class of extra-mathematical entities, X. Let there exist on that class a set of empirical relations R={R1 , ..., Rn }. Let us further consider a set of numbers N (in general a subset of the set of real numbers Re) and let there be defined on that set a set of numerical relations P ={P1, ...,Pn} . Let there exist a mapping M with domain X and a range in N,

M:X → N which is a homomorphism of the empirical relationship system $\langle$X,R$\rangle$ and the numerical relational system $\langle$N,P$\rangle$ (Finkelstein 1975, 105).This is illustrated below in figure 1 where $x_i \in$ ***X and ni*** $\in$ ***N. M*** is so called 'scale of measurement'

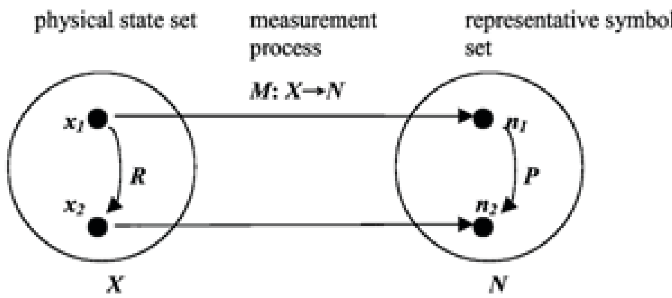

Figure1: *General Measurement Process(Boumans 2005)*

Note: Group homomorphism from group ***(G,*)*** to group ***(H, •)*** is a function ***h: G → H*** such that for all u and v in G it holds that ***h (u*v) =h (u).h (v)***

Measurement theory is supposed to analyze the concept of a scale of measurement. It distinguishes various types of scale and describes their uses, and formulates the conditions required for the existence of scales of various types.

However, the representational theory of measurement has turned too much into a pure mathematical discipline, leaving out the question of how the mathematical structures gain their empirical significance in actual measurement. Heidelberger (1994a, 1994b) discussed this problem of empirical significance. (Boumans, 2005)

In the development process of representational theory, Heidelberger emphasizes that, most followers of the representational theory of today have adopted an 'operationalist' interpretation.

The operational theory avoids the assumption of an underlying reality, requiring only that measurement consists of precisely specified operations; scientific theories concern only relationships among measurements. The classical theory, like the representational theory, assumes an objective reality, but, unlike the representational theory, holds that only quantitative attributes are measurable, and measurement involves the discovery of the magnitudes of these attributes. In the classical theory, like the operational theory, meaningfulness comes from empirical support for scientific theories describing the interrelationships of various measurements. (Sarle, 1997)

Operationalist interpretation is best illustrated by Stevens' dictum (Stevens 1959, 19) :

"Measurement is the assignment of numerals to objects or events according to rule 'any rule'. Of course, the fact that numerals can be assigned under different rules leads to different kinds of scales and different kinds of measurements, not all of equal power and usefulness. Nevertheless, provided a consistent rule is followed, some form of measurement is achieved" (Boumans, 2005).

Stevens' Scales of Measurement

Stevens adopted the representationalist philosophy in a 'nominalist' form (Michell, 1993), defining measurement as the 'assignment of numbers to objects according to a rule' (Chrisman, 1998).

According to Townsend and Ashby, Stevens was responsible for some critically important ideas regarding the use and misuse of measurement scales, the theory has progressed far beyond his work. Indeed several of his most valuable concepts gain their true significance only in the context of later developments.

Stevens (1946) was defending that measurement exists in a variety of forms and that scale of measurement fall into certain definite classes. These classes are determined both by the empirical operations invoked in the process of "measuring" and by the formal (mathematical) properties of scales. Furthermore, the statistical manipulations that can legitimately be applied to empirical data depend upon the type of scale against which the data are ordered. The type of scale achieved depends upon the character of the basic empirical operations performed. These operations are limited ordinarily by the nature of the thing being scaled and by our choice of procedures, but once selected; the operations determine that there will eventuate one or another of the scales (Stevens, 1946) which are shown in table1.

Table 1 reproduces Stevens' original table exactly so that his presentation is not clouded by the reinterpretations developed over the past fifty years. (Chrisman, 1998).

***Table1:** Stevens' scales of measurement (1946)*

Scale	Basic Empirical Operations	Mathematical Group Structure	Permissible statistics (invariantive)
NOMINAL	Determination of equality	*Permutation group* $x' = f(x)$ $f(x)$ means any one-to-one substitution	Number of cases Mode
ORDINAL	Determination of greater or less	*Isotonic group* $x' = f(x)$ $f(x)$ means any monotonic increasing function	Median Percentiles
INTERVAL	Determination of equality of intervals or differences	*General linear group* $x' = ax + b$	Mean Standard deviation Rank-order correlation Product-moment correlation
RATIO	Determination of equality of ratios	*Similarity group* $x' = ax$	Coefficient of variation

Mathematical Determination of Scales of Measurement

In the theory of measurement, we think of starting with a set A of objects that we want to measure. We shall think of a scale of measurement as a function f that assigns a real number ***f(a)*** to each element a of A. More generally we can think of ***f(a)*** as belonging to another set B.

The representational theory of measurement gives conditions under *which a function is acceptable scale* of measurement. Following ideas of Stevens (1946, 1951, 1959) expressing an admissible transformation as a function that sends, transfers one acceptable scale into another, for example Centigrade into Fahrenheit and kilograms into pounds.

Theoretically an admissible transformation of a scale can be thought of as a function $\emptyset$ that takes ***f(a)*** into $(\emptyset o f)(a)$. Stevens classified scales into four types according to the associated class of admissible transformations. These are nominal, ordinal, interval, and ratio scales.

If the admissible transformations are the (strictly) monotone increasing transformations, this scale is an ***ordinal scale***. Comparing sizes is meaningful in ordinal scale: $f(a)>f(b)$

If the admissible transformations are the form:

$\emptyset(x)=\alpha x+\beta, \alpha>0$, It is called ***interval scale***.

Temperature scale is an interval scale since the transformation from Centigrade into Fahrenheit involves the following form:

$$\emptyset(x)=(9/5)x+32.$$

If the numbers assigned to two pairs of objects are equally different, then the pairs of objects must be equally different in the real world. For interval scales, it is meaningful to compare assigned numbers as shown below:

If $a-b>c-d$ then, $f(a)-f(b)>f(c)-f(d)$. 'Or vice versa'.

We call a scale a ratio scale, if the admissible transformations are the form:

$$\emptyset(x)=\alpha x, \alpha>0.$$

Such transformations change the units. Transformation from kilograms into pounds involves the admissible transformation $\emptyset(x)=2,2x$ hence, mass is an example of a ratio scale. If person A is 80 kg and B is 40 kg, it is obvious that person A is twice heavier than person B. Comparison of ratios $\frac{f(a)}{f(b)} > \frac{f(c)}{f(d)}$ is meaningful for ratio scale (Roberts, 2009).

Brief explanation of the types of scales is given in the next section.

Types of Scales

Nominal Scales

Nominal scales are composed of sets of categories in which objects are classified. A nominal scale is simply some placing of data into categories, without any order or structure. The nominal scale does not express any values or relationships between variables. Variables measured on a nominal scale are categorical or qualitative variables.

This scale has no numeric significance. The only possible arithmetic operation on this scale is counting. According to the

Stevens, only non-parametric statistics such as computation of chi-square and finding mode can be used here. Yes/no, male/female, married/single, etc. are examples of dichotomous nominal data and nationality, consisting of multiple values, such as 'Georgian', 'Turkish', 'Australian' etc. is example of non-dichotomous data.

Even at this level, regression analysis is possible, using *dummy variables;* for example, gender can be treated as a dummy variable equaling 0 for subjects of male gender and 1 for subjects of female gender. They can be used either as an independent variable in an ordinary least squares regression, or as dependent variables in the probit or logistic regression. "A probit model is a type of regression where the dependent variable can only take two values, for example married or not married." (Wikipedia, 2012)

Table 2: *Example of nominal scale usage in questionnaire.*

	yes	no
Does your lecturer explain his/her grading system at the beginning of the semester?		
Does your lecturer use technological tools?		

Ordinal Scales

The ordinal scale arises from the operation of the rank ordering (Stevens, 1946). The main characteristic of this type of scale is that, the categories have a logical order or ordered relationship to each other. Ordinal variables do not tell anything about the absolute magnitude of the difference between 1st and 2nd or 4th and 5th. For a scale to be at the ordinal level of measurement, the categories comprising the scale must be mutually exclusive and ordered (Knapp, 1990)

Economic situation, joining to the social activities, level of success, grades for academic performance (A, B, C ...) are some variables, which are useful for statistical implementations in the ordinal scale. Examples of the type of statistical operations appropriate to this scale are finding median, mode, rank order correlation, variance (non-parametric). An ordinal level of measurement example is given below:

Rank the students according to their English Exam results from best to worst.

-----Giorgi A.
-----Natia S.
-----Aytaç K.

Interval Scales

On this type of measurement scales, one unit of scale shows the same magnitude on the trait or characteristic being measured across the whole range of the scale. This means that we can interpret differences in the distance along the scale. All quantitative attributes are measurable with help of interval scale. Beside ***equality of units***, zero does not represent the absolute lowest value. Rather, it is point on the scale with numbers both above and below it.

It is clear that, a student who scores 80% is probably a better student than someone who scores 50%. The difference between the two scores is 30%. In an interval scale, the data can be ranked and for which the difference between the two values can be calculated and interpreted. The zero point on an interval scale is *arbitrary* and is *not a true zero*; therefore, it is *not* possible to make decision about *how many times higher one score is than another.* For instance in temperature, a temperature of 30 degrees Fahrenheit is not twice as warm as one of 15 degrees Fahrenheit and they are not in the ratio 2:1! Like temperature, IQ might also lie on interval scale. Interval scale data would use parametric statistical techniques: Mean and standard deviation, Correlation – r, Regression, Analysis of variance, Factor analysis plus a whole range of advanced multivariate and modeling techniques.

Stevens clearly demonstrated the difference between ordinal and interval scales. An ordinal variable arises from a scale for which all order- reserving (monotonic) transformations are admissible; that is, they leave the scale form invariant. For an interval variable, the only admissible transformations are those of the linear-form $y = bx + a$ (Knapp, 1990).

The most common examples of interval scales are scores obtained using objective tests such as multiple-choice tests of achievement. It is widely assumed that each multiple-choice test item measures a single point's worth of the trait being measured and that all points are equal to all other points. However, such tests do not measure at the ratio level because the zero on such tests is arbitrary not absolute. To see this, consider someone who gets a zero on a multiple-choice final examination. Does the zero mean that the student has absolutely no knowledge of or skills in the subject area? Probably a score of zero only indicates that they know nothing on that test, 'not that they have zero knowledge of the content domain'. (M.L., 1997)

Ratio Scales

Ratio scales permit the researcher to compare both *differences* in scores and the *relative magnitude of scores*. This has the properties of an interval scale together with a fixed origin or zero point. Actually, in the physical sciences and engineering measurement is done on ratio scales. Mass, length, duration, energy, and electric charge are variables of physical measures that can be used in ratio scales. Phrases such as "four times" and "twice" are meaningful at the ratio level. When we talk about interval scale, we mentioned that it is not possible to find the ratio of units, since there is not absolute zero point. I can explain this situation with the following example.

Let us consider three students A, B, C and their mathematics examination results as 40, 50, and 80 respectively. Only if the beginning point is zero we can say that student C is twice as successful as A. However, if the base point is starting from 20 it is obvious that student C is three times more successful than student A.

Table3: *General characteristics of scales' types.*

Scale	***Characteristic***	***Logical Operations Allowed***
Nominal	Naming	=, ≠
Ordinal	Ordering	=, ≠, >, <
Interval	Equal interval, no absolute zero	=, ≠, >, <, +, -
Ratio	Equal interval, absolute zero	=, ≠, >, <, +, -, ×, ÷

The miss point of the interval scale is that zero point is chosen arbitrary. (Belgeler, 2013). Ratio scale is like the interval-scale variable, however it has a non-arbitrary zero value. Beside all the statistics permitted for interval scales, ratio scale allows finding geometric mean, harmonic mean, coefficient of variation, logarithms too.

Conclusion

Measurement theory encourages people to think about the meaning of their data and important assumptions behind the analysis. The purpose of this article is giving brief information about historical development of measurement theory and its different implementations. Stevens showed that strong assumptions are required for reliable statistics. Here I try to clarify the main differences between scale types and I explained feasible analyses with respect to scale types.

After Stevens, different suggestions and discussions started. His opponents could justifiably protest that in *science all facts are permissible*. Not with standing this, however, it is somewhat worrisome when the conclusions derived from measurements depend on quite arbitrary aspects of the chosen measurement scale. Therefore, there may have been some point to Stevens' prescriptions. (Michell, 1986)

Beside the opinions of antagonists, using different scales in any survey will help to the respondent to prevent them from clicking the highest, lowest or middle rating all the time. Another benefit of using different kinds of scales in the survey is that each scale provides us a unique perspective on the data that we are analyzing (Taylor, 2012). Generally, in questionnaires, we use Stevens' scale types but the people are not aware of it. There is a hierarchy in the level of measurement and as I explained at lower levels assumptions are less restrictive. In the upper level, the current level contains all the properties of the previous one. Main principles of statistical observations are coincided with Stevens' scale types. As a result, Measurement theory encourages people to provide meaningful information about reality. If the target is to obtain the qualitative research results, using appropriate scale type will give us results that are most beneficial and reliable.

References

[1] Belgeler. (2013, April 12). Retrieved April 2013, from Belgeler: http://www.belgeler.com

[2] Boumans, M. (2005, December). Measurement Outside the Laboratory. Philosophy of Science, 850-863.

[3] Chrisman, N. R. (1998). Beyond Stevens: A revised approach to measurement for geographic information. 25(4), pp. 231-242.

[4] Diez, J. (n.d.). Hidstory of measurement theory. In J. Diez, Hidstory of measurement theory (Vol. 3). Barcelona, Spain: Encyclopedia of life support systems.

[5] E.P. Araujo,, D. F. Andrade, S. Bortolotti. (2009, Agust 24). Item Response Theory. 1000-1008. Brazil.

[6] James Townsend , Gregory Ashby. (1984). Measurement scales and statistics: The misconception misconceived. Psychological Bulletin, 96(2), 394-401.

[7] Knapp, T. R. (1990, March). Treating Ordinal Scales as Interval Scales: An Attempt To Resolve the Controversy. 39.

[8] M.L., P. (1997). Understanding research methods:An overview of the essentials. Los Angeles, USA: Pyrczak Publishing.

[9] Michell, J. (1986, November). Measurement Scales and Statistics A Clash of Paradigms. Psychological Bulletin, 100(3), pp. 398-407.

[10] Roberts, F. S. (2009). Meaningful and Meaningless Statements in Epidemiology and Public Health. NJ, Washington, USA.

[11] Sarle, W. S. (1997, September 14). Measurement theory: Frequently asked questions. NC, Washington, US.

[12] Stevens, S. S. (1946, June). On the Theory of Scales of Measurement. (H. University, Ed.) Science, 103, 677-680.

[13] Steyer, R. (1999). Steyer, R. 1999, Classical (Psychometric) Test Theory. Jena, Germany. Jena, Germany.

[14] Suen, H. K. (1990). Principles of test theories. New Jersy, USA: Lawrence Erlbaum Associates, Inc.

[15] Taylor, I. (2012, January). 4 Measurement Scales Every Researcher Should Remember.