SCIENCE

Vol. 103, No. 2684 Friday, June 7, 1946

On the Theory of Scales of Measurement

S. S. Stevens

Director, Psycho-Acoustic Laboratory, Harvard University

FOR SEVEN YEARS A COMMITTEE of the British Association for the Advancement of Science debated the problem of measurement. Appointed in 1932 to represent Section A (Mathematical and Physical Sciences) and Section J (Psychology), the committee was instructed to consider and report upon the possibility of "quantitative estimates of sensory events"—meaning simply: Is it possible to measure human sensation? Deliberation led only to disagreement, mainly about what is meant by the term measurement. An interim report in 1938 found one member complaining that his colleagues "came out by that same door as they went in," and in order to have another try at agreement, the committee begged to be continued for another year.

For its final report (1940) the committee chose a common bone for its contentions, directing its arguments at a concrete example of a sensory scale. This was the Sone scale of loudness (S. S. Stevens and H. Davis. *Hearing.* New York: Wiley, 1938), which purports to measure the subjective magnitude of an auditory sensation against a scale having the formal properties of other basic scales, such as those used to measure length and weight. Again the 19 members of the committee came out by the routes they entered, and their views ranged widely between two extremes. One member submitted "that any law purporting to express a quantitative relation between sensation intensity and stimulus intensity is not merely false but is in fact meaningless unless and until a meaning can be given to the concept of addition as applied to sensation" (Final Report, p. 245).

It is plain from this and from other statements by the committee that the real issue is the meaning of measurement. This, to be sure, is a semantic issue, but one susceptible of orderly discussion. Perhaps agreement can better be achieved if we recognize that measurement exists in a variety of forms and that scales 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 the scales. Furthermore—and this is of great concern to several of the sciences—the statistical manipulations that can legitimately be applied to empirical data depend upon the type of scale against which the data are ordered.

A Classification of Scales of Measurement

Paraphrasing N. R. Campbell (Final Report, p. 340), we may say that measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules. The fact that numerals can be assigned under different rules leads to different kinds of scales and different kinds of measurement. The problem then becomes that of making explicit (a) the various rules for the assignment of numerals, (b) the mathematical properties (or group structure) of the resulting scales, and (c) the statistical operations applicable to measurements made with each type of scale.

Scales are possible in the first place only because there is a certain isomorphism between what we can do with the aspects of objects and the properties of the numeral series. In dealing with the aspects of objects we invoke empirical operations for determining equality (classifying), for rank-ordering, and for determining when differences and when ratios between the aspects of objects are equal. The conventional series of numerals yields to analogous operations: We can identify the members of a numeral series and classify them. We know their order as given by convention. We can determine equal differences, as $8-6=4-2$, and equal ratios, as $8/4=6/3$. The isomorphism between these properties of the numeral series and certain empirical operations which we perform with objects permits the use of the series as a *model* to represent aspects of the empirical world.

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 listed in Table 1.[1]

The decision to discard the scale names commonly encountered in writings on measurement is based on the ambiguity of such terms as "intensive" and "extensive." Both ordinal and interval scales have at times been called intensive, and both interval and ratio scales have sometimes been labeled extensive.

TABLE 1

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 Contingency correlation
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

It will be noted that the column listing the basic operations needed to create each type of scale is cumulative: to an operation listed opposite a particular scale must be added all those operations preceding it. Thus, an interval scale can be erected only provided we have an operation for determining equality of intervals, for determining greater or less, and for determining equality (not greater and not less). To these operations must be added a method for ascertaining equality of ratios if a ratio scale is to be achieved.

In the column which records the group structure of each scale are listed the mathematical transformations which leave the scale-form invariant. Thus, any numeral, x, on a scale can be replaced by another numeral, x', where x' is the function of x listed in this column. Each mathematical group in the column is contained in the group immediately above it.

The last column presents examples of the type of statistical operations appropriate to each scale. This column is cumulative in that *all* statistics listed are admissible for data scaled against a ratio scale. The criterion for the appropriateness of a statistic is *invariance* under the transformations in Column 3. Thus, the case that stands at the median (mid-point) of a distribution maintains its position under all transformations which preserve order (isotonic group), but an item located at the mean remains at the mean only under transformations as restricted as those of the linear group. The ratio expressed by the coefficient of variation remains invariant only under the similarity transformation (multiplication by a constant). (The rank-order correlation coefficient is usually deemed appropriate to an ordinal scale, but actually this statistic assumes equal intervals between successive ranks and therefore calls for an interval scale.)

Let us now consider each scale in turn.

NOMINAL SCALE

The *nominal scale* represents the most unrestricted assignment of numerals. The numerals are used only as labels or type numbers, and words or letters would serve as well. Two types of nominal assignments are sometimes distinguished, as illustrated (a) by the 'numbering' of football players for the identification of the individuals, and (b) by the 'numbering' of types or classes, where each member of a class is assigned the same numeral. Actually, the first is a special case of the second, for when we label our football players we are dealing with unit classes of one member each. Since the purpose is just as well served when any two designating numerals are interchanged, this scale form remains invariant under the general substitution or permutation group (sometimes called the symmetric group of transformations). The only statistic relevant to nominal scales of Type A is the number of cases, e.g. the number of players assigned numerals. But once classes containing several individuals have

[1] A classification essentially equivalent to that contained in this table was presented before the International Congress for the Unity of Science, September 1941. The writer is indebted to the late Prof. G. D. Birkhoff for a stimulating discussion which led to the completion of the table in essentially its present form.

been formed (Type B), we can determine the most numerous class (the mode), and under certain conditions we can test, by the contingency methods, hypotheses regarding the distribution of cases among the classes.

The nominal scale is a primitive form, and quite naturally there are many who will urge that it is absurd to attribute to this process of assigning numerals the dignity implied by the term measurement. Certainly there can be no quarrel with this objection, for the naming of things is an arbitrary business. However we christen it, the use of numerals as names for classes is an example of the "assignment of numerals according to rule." The rule is: Do not assign the same numeral to different classes or different numerals to the same class. Beyond that, anything goes with the nominal scale.

Ordinal Scale

The *ordinal scale* arises from the operation of rank-ordering. Since any 'order-preserving' transformation will leave the scale form invariant, this scale has the structure of what may be called the isotonic or order-preserving group. A classic example of an ordinal scale is the scale of hardness of minerals. Other instances are found among scales of intelligence, personality traits, grade or quality of leather, etc.

As a matter of fact, most of the scales used widely and effectively by psychologists are ordinal scales. In the strictest propriety the ordinary statistics involving means and standard deviations ought not to be used with these scales, for these statistics imply a knowledge of something more than the relative rank-order of data. On the other hand, for this 'illegal' statisticizing there can be invoked a kind of pragmatic sanction: In numerous instances it leads to fruitful results. While the outlawing of this procedure would probably serve no good purpose, it is proper to point out that means and standard deviations computed on an ordinal scale are in error to the extent that the successive intervals on the scale are unequal in size. When only the rank-order of data is known, we should proceed cautiously with our statistics, and especially with the conclusions we draw from them.

Even in applying those statistics that are normally appropriate for ordinal scales, we sometimes find rigor compromised. Thus, although it is indicated in Table 1 that percentile measures may be applied to rank-ordered data, it should be pointed out that the customary procedure of assigning a value to a percentile by interpolating linearly within a class interval is, in all strictness, wholly out of bounds. Likewise, it is not strictly proper to determine the mid-point of a class interval by linear interpolation, because the linearity of an ordinal scale is precisely the property which is open to question.

Interval Scale

With the *interval scale* we come to a form that is "quantitative" in the ordinary sense of the word. Almost all the usual statistical measures are applicable here, unless they are the kinds that imply a knowledge of a 'true' zero point. The zero point on an interval scale is a matter of convention or convenience, as is shown by the fact that the scale form remains invariant when a constant is added.

This point is illustrated by our two scales of temperature, Centigrade and Fahrenheit. Equal intervals of temperature are scaled off by noting equal volumes of expansion; an arbitrary zero is agreed upon for each scale; and a numerical value on one of the scales is transformed into a value on the other by means of an equation of the form $x' = ax + b$. Our scales of time offer a similar example. Dates on one calendar are transformed to those on another by way of this same equation. On these scales, of course, it is meaningless to say that one value is twice or some other proportion greater than another.

Periods of time, however, can be measured on ratio scales and one period may be correctly defined as double another. The same is probably true of temperature measured on the so-called Absolute Scale.

Most psychological measurement aspires to create interval scales, and it sometimes succeeds. The problem usually is to devise operations for equalizing the units of the scales—a problem not always easy of solution but one for which there are several possible modes of attack. Only occasionally is there concern for the location of a 'true' zero point, because the human attributes measured by psychologists usually exist in a positive degree that is large compared with the range of its variation. In this respect these attributes are analogous to temperature as it is encountered in everyday life. Intelligence, for example, is usefully assessed on ordinal scales which try to approximate interval scales, and it is not necessary to define what zero intelligence would mean.

Ratio Scale

Ratio scales are those most commonly encountered in physics and are possible only when there exist operations for determining all four relations: equality, rank-order, equality of intervals, and equality of ratios. Once such a scale is erected, its numerical values can be transformed (as from inches to feet) only by multiplying each value by a constant. An absolute zero is always implied, even though the zero value on some scales (e.g. Absolute Temperature) may never be produced. All types of statistical measures are applicable to ratio scales, and only with these scales may we properly indulge in logarithmic transformations such as are involved in the use of decibels.

Foremost among the ratio scales is the scale of number itself—cardinal number—the scale we use when we count such things as eggs, pennies, and apples. This scale of the numerosity of aggregates is so basic and so common that it is ordinarily not even mentioned in discussions of measurement.

It is conventional in physics to distinguish between two types of ratio scales: *fundamental* and *derived.* Fundamental scales are represented by length, weight, and electrical resistance, whereas derived scales are represented by density, force, and elasticity.

These latter are *derived* magnitudes in the sense that they are mathematical functions of certain fundamental magnitudes. They are actually more numerous in physics than are the fundamental magnitudes, which are commonly held to be basic because they satisfy the criterion of *additivity.* Weights, lengths, and resistances can be added in the physical sense, but this important empirical fact is generally accorded more prominence in the theory of measurement than it deserves. The so-called fundamental scales are important instances of ratio scales, but they are only instances. As a matter of fact, it can be demonstrated that the fundamental scales could be set up even if the physical operation of addition were ruled out as impossible of performance. Given three balances, for example, each having the proper construction, a set of standard weights could be manufactured without it ever being necessary to place two weights in the same scale pan at the same time. The procedure is too long to describe in these pages, but its feasibility is mentioned here simply to suggest that physical addition, even though it is sometimes possible, is not necessarily the basis of all measurement. Too much measuring goes on where resort can never be had to the process of laying things end-to-end or of piling them up in a heap.

Ratio scales of psychological magnitudes are rare but not entirely unknown. The Sone scale discussed by the British committee is an example founded on a deliberate attempt to have human observers judge the loudness ratios of pairs of tones. The judgment of equal intervals had long been established as a legitimate method, and with the work on sensory ratios, started independently in several laboratories, the final step was taken to assign numerals to sensations of loudness in such a way that relations among the sensations are reflected by the ordinary arithmetical relations in the numeral series. As in all measurement, there are limits imposed by error and variability, but within these limits the Sone scale ought properly to be classed as a ratio scale.

To the British committee, then, we may venture to suggest by way of conclusion that the most liberal and useful definition of measurement is, as one of its members advised, "the assignment of numerals to things so as to represent facts and conventions about them." The problem as to what is and is not measurement then reduces to the simple question: What are the rules, if any, under which numerals are assigned? If we can point to a consistent set of rules, we are obviously concerned with measurement of some sort, and we can then proceed to the more interesting question as to the kind of measurement it is. In most cases a formulation of the rules of assignment discloses directly the kind of measurement and hence the kind of scale involved. If there remains any ambiguity, we may seek the final and definitive answer in the mathematical group-structure of the scale form: In what ways can we transform its values and still have it serve all the functions previously fulfilled? We know that the values of all scales can be multiplied by a constant, which changes the size of the unit. If, in addition, a constant can be added (or a new zero point chosen), it is proof positive that we are not concerned with a ratio scale. Then, if the purpose of the scale is still served when its values are squared or cubed, it is not even an interval scale. And finally, if any two values may be interchanged at will, the ordinal scale is ruled out and the nominal scale is the sole remaining possibility.

This proposed solution to the semantic problem is not meant to imply that all scales belonging to the same mathematical group are equally precise or accurate or useful or "fundamental." Measurement is never better than the empirical operations by which it is carried out, and operations range from bad to good. Any particular scale, sensory or physical, may be objected to on the grounds of bias, low precision, restricted generality, and other factors, but the objector should remember that these are relative and practical matters and that no scale used by mortals is perfectly free of their taint.