

Sensitivity—A Criterion for the Comparison of Methods of Test

J. Mandel and R. D. Stiehler

In the evaluation of many methods of test, the two usual criteria—precision and accuracy—are insufficient. Accuracy is only applicable where comparisons with a standard can be made. Precision, when interpreted as degree of reproducibility, is not necessarily a measure of merit, because a method may be highly reproducible merely because it is too crude to detect small variations.

To obtain a quantitative measure of merit of test methods, a new concept—sensitivity—is introduced. If M is a measure of some property Q , and σ_M its standard deviation, the sensitivity of M , denoted ψ_M , is defined by the relation $\psi_M = (dM/dQ)/\sigma_M$. It follows from this definition that the sensitivity of a test method may or may not be constant for all values of the property Q . A statistical test of significance is derived for the ratio of sensitivities of alternative methods of test. Unlike the standard deviation and the coefficient of variation, sensitivity is a measure of merit that is invariant with respect to any functional transformation of the measurement, and is therefore independent of the scale in which the measurement is expressed.

1. Introduction

In the physical sciences, there frequently is a choice between several methods for the determination of a particular characteristic. In such cases means are necessary to compare the relative merits of the various methods. The customary procedure for evaluating a test method, particularly in analytical chemistry, is to determine accuracy by comparing the values found on known samples with the theoretical values, and to express precision by the reproducibility of the experimental values as measured by the standard deviation. Alternative methods can then be compared on the basis of both precision and accuracy. In the evaluation of many methods of test, particularly those for polymeric materials, these criteria are insufficient. This paper presents a single criterion by which the relative merit of methods of test can be evaluated. The main advantage of the new criterion—referred to as sensitivity—is that it takes into account, not only the reproducibility of the testing procedure, but also its ability to detect small variations in the characteristic to be measured.

The need for such a criterion has been felt by various workers. Newton [1]¹ discusses the fallacy of comparing alternative test methods on the sole basis of their respective standard deviations of error. According to Throdahl [2], Mooney considers a *coefficient of discrimination*, defined as the ratio of the difference between the average values obtained from two sets of samples to the standard deviation within samples. Dillon [3] compares two plasmometers on the basis of their *selectivities*, the concept of selectivity being defined by him as the “percentage difference between two observations on different mixtures divided by the average maximum per-

centage error.” Roth and Stiehler [4], in comparing the precisions of strain and stress measurements, convert the standard deviation of strain into stress units and then consider the ratio of this converted standard deviation to that of stress; alternatively, they consider the ratio of the variance “between batches” to that “within batches” as a criterion for the sensitivity of either method. The latter criterion is also applied by Buist and Davies [5] and by Newton, Scott, and Whorlow [6], who refer to it as the *discriminating power*. Reichel [7] introduces the concept of “*technische Güte*” to characterize the merit of methods of chemical analysis.

In this paper, a general mathematical definition is proposed for the sensitivity concept, which is an intrinsic measure of merit, of particular value for the comparison of two or more alternative test methods.

2. Sensitivity in the Case of Proportionality

In most analytical methods in chemistry the desired material is not determined directly but is calculated from measurements of a proportional quantity of some related material. For example, in the determination of zinc, the amount of this metal is calculated from the quantity of zinc oxide, zinc sulfate, or other zinc compound actually measured. In comparing the relative merits of the use of these alternative compounds, a pertinent consideration, besides the magnitude of experimental error, is the ratio of the equivalent weight of the zinc compound to that of zinc. It is recognized that a larger ratio is preferable, provided that the experimental error is not increased in the same proportion. A correct evaluation of alternative methods, involving zinc compounds of different equivalent weight, can be obtained from the following considerations:

¹ Figures in brackets indicate the literature references at the end of this paper.

The percentage of zinc in the unknown is given by the equation

$$Zn = \frac{100P}{W} \times \frac{[Zn]}{[Zn \text{ compound}]}, \quad (1)$$

where P is the weight of the Zn compound measured; W is the weight of the sample; $[Zn]$ is the equivalent weight of zinc; and $[Zn \text{ compound}]$ is the equivalent weight of the zinc compound measured.

Let Q equal the percentage of zinc, R the ratio of the equivalent weights of zinc and the zinc compound measured, and M the weight of zinc compound per gram of sample. Then

$$Q = 100MR. \quad (2)$$

From this relation it follows [8] that the standard deviation for the determination of zinc is given by the equation

$$\sigma_Q = 100R\sigma_M. \quad (3)$$

Equation (3) shows that the precision of the zinc determination is improved when (1) the quantity $100R$ is small, and (2) the error of measurement of the zinc compound (σ_M) is small.

If the weight of zinc compound per gram of sample is plotted against the percentage of zinc, a straight line is obtained, as shown in figure 1. The line passes through the origin and has a slope equal to the reciprocal of $100R$. Let the slope be designated as K . Equation (3) can now be written

$$\sigma_Q = \frac{\sigma_M}{K}. \quad (4)$$

Thus, high precision in the determination of Q (i. e., a small value for σ_Q) reduces to the requirement that the quantity K/σ_M be large. The absolute value of the quantity K/σ_M is defined as the sensitivity of the measurement of M for the determination of Q and is denoted by ψ . Thus

$$\text{Sensitivity} = \psi = \frac{|K|}{\sigma_M}. \quad (5)$$

It is obvious that the merit of the method is dependent on more than the reproducibility of measurement of M . It also depends on the rate of change in M with a change in Q or the ability to discriminate between small changes in Q .

3. Sensitivity in the General Case

In many methods, particularly when dealing with polymeric materials, the measured quantity M and the desired quantity Q are not linearly related. An example is the measurement of refractive index to determine the percentage of bound styrene in GR-S synthetic rubber. Additional difficulties arise when it becomes impossible to define a single criterion Q for the characterization of the properties in which one is interested. In these cases it is necessary to consider a measurable quantity M that is in some sense related to these properties. An example of this type is given by vulcanization tests on rubbers, where stress-strain measurements are used as an index or measure of the degree of vulcanization. Whether or not a quantity Q can be defined, and whatever the relation may be between a characteristic Q and the measured quantity M , the criterion defined as sensitivity can effectively be used for evaluating and comparing methods of test.

Figure 2 illustrates a case in which Q is susceptible of exact definition and the relation between M and Q is curvilinear. If it is desired to differentiate between the two close values, Q_1 and Q_2 , by means of the corresponding measurements M_1 and M_2 , it is again apparent that the success of the operation will depend on two circumstances: (1) the magnitude of the difference $M_2 - M_1$, for a given difference $Q_2 - Q_1$; i. e., the magnitude of the slope $(M_2 - M_1)/(Q_2 - Q_1)$; and (2) the precision of measurement; i. e., the smallness of the standard deviation. Indeed, if σ_M is too large, the regions of uncertainty of M_1 and M_2 may overlap, and the discrimination fail. As before, these two desiderata can be combined in a single criterion, the sensitivity, defined according

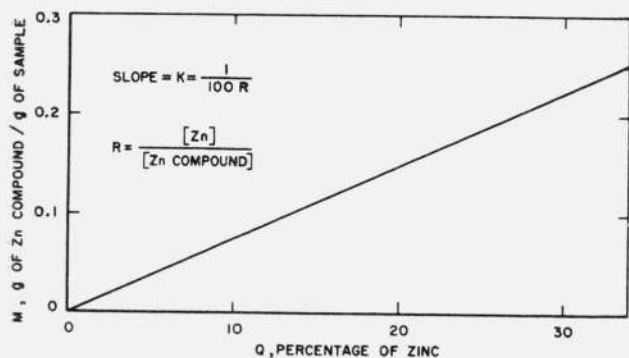


FIGURE 1. Sensitivity for proportional relationship.

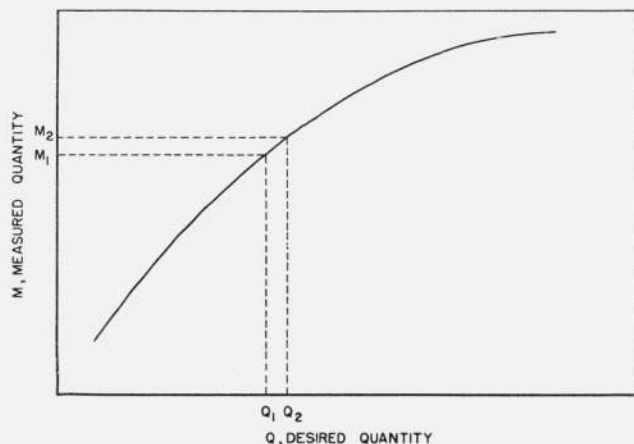


FIGURE 2. Sensitivity for curvilinear relationship.

to eq (5) as the absolute value of the ratio of the slope $K = (M_2 - M_1)/(Q_2 - Q_1)$ to the standard deviation of M , σ_M . The larger the sensitivity, the more useful will be the test method M for the characterization of Q . It should be noted, however, that in the general case, K is no longer constant but varies with the value of Q . Thus, even in cases in which the experimental error (measured by σ_M) remains constant, the sensitivity may vary with the value of Q . Only when the error is proportional to K is the sensitivity constant.

If the properties under consideration cannot be expressed by means of a single criterion Q , it is not possible to determine the absolute sensitivity of a method of test. It is possible, however, to determine the relative sensitivities of two or more methods used to characterize these properties. This important application of the sensitivity concept can best be shown by first considering a case in which a single criterion Q exists, and two alternative measuring methods M and N , both related to Q , are to be compared. For example, density and refractive-index methods for determining the bound styrene in GR-S may be compared without knowing the actual percentage of bound styrene. Let ψ_M and ψ_N be the sensitivities corresponding to the two methods. From eq (5) it follows that the ratio of the sensitivities is given by

$$\frac{\psi_M}{\psi_N} = \frac{|K_M/K_N|}{\sigma_M/\sigma_N} = \frac{|K'|}{\sigma_M/\sigma_N} \quad (6)$$

The meaning of K' is found as follows:

$$K' = \frac{K_M}{K_N} = \frac{\Delta M/\Delta Q}{\Delta N/\Delta Q} = \frac{\Delta M}{\Delta N} \quad (7)$$

Thus K' is the slope of a curve of M plotted as a function of N . From eq (5) it follows that the dimension of sensitivity is that of $1/Q$, since σ_M has the dimension of M , and K is of dimension M/Q . On the other hand, the ratio of the sensitivities of alternative test methods given in eq (6) is dimensionless. This fact, as well as eq (7), shows that the comparison of two methods, by means of the ratio of their sensitivities, does not necessitate a knowledge of their relation to the theoretical Q . All that is required is a knowledge of their mutual relationship.

In the case of bound styrene, the relation between density and refractive index can be established from a series of samples of different bound styrene contents without a knowledge of bound styrene in any sample. Of course, the bound styrene content could be determined by some absolute method, and the absolute sensitivities of the refractive index and density methods for measuring this property could be established.

In the case of stress-strain measurements, on the other hand, the characteristic—degree of vulcanization—cannot be represented by a single quantity Q and consequently no absolute sensitivities for either method can be calculated. Nevertheless, relation (6), with K' given by (7), can be applied, since it

does not involve the quantity Q , and the sensitivity ratio can be used to compare the measurement of tensile stress [9] and the measurement of strain [4]. The relationship between these two methods of measurement for a GR-S synthetic rubber compound, according to Roth and Stiehler [4], is given by the equation:

$$SE^n = C \quad (8)$$

where S represents tensile stress, E represents strain, and n and C are constants for any particular type of vulcanizates.

If the logarithmic derivative is taken, it follows that

$$\frac{dS}{S} = -n \frac{dE}{E} \quad (9)$$

As n is of the order of 1.5, it might be expected that measurements of tensile stress would detect variations in the vulcanizates better than measurements of strain. However, Roth and Stiehler [4] show that the error of measurement of strain is much smaller than that of the usual measurement of tensile stress; hence, the sensitivity of strain measurements is greater.

From eq (9) it follows that the slope of the strain versus tensile-stress curve is

$$\frac{dE}{dS} = -\frac{E}{nS}$$

and consequently,

$$\frac{\psi_E}{\psi_S} = \frac{E/nS}{\sigma_E/\sigma_S} \quad (10)$$

This expression is found to exceed unity, as shown in table 1, which lists data pertinent for the calculation of the sensitivity ratio, for tensile-stress and strain values obtained in three different plants and for two cures [10]. It should be noted that the ratio of the two sensitivities varies with the degree or time of cure, since the factor E/nS decreases as vulcanization progresses. The advantages of the strain test are therefore greatest for tests on vulcanizates that are undercured. The data also show that the greater sensitivity of the strain test is due to its better reproducibility.

TABLE 1. Comparison of tensile stress and strain measurements of GR-S synthetic rubber

Cure at 292° F	Plant	K' ($E/1.6 S$) ^a	Standard deviation		Ratio of sensitivities (strain/ stress)
			Strain at 400 psi	Stress at 300% elongation	
25	A	.610	1.6	9.5	3.6
	B	.542	3.1	22.5	3.9
	C	.362	2.1	15.4	2.6
100	A	.0706	0.83	14.8	1.3
	B	.0703	1.84	35.8	1.4
	C	.0641	1.17	37.1	2.0

^a The value 1.6 taken for n is an upper limit for GR-S synthetic rubber. For values of n smaller than 1.6, the ratios in the last column will be larger.

It should be noted that the application of the sensitivity criterion in comparing two test methods implies that a definite functional relationship exists between the properties measured by the two methods. This restriction is not introduced by the sensitivity concept, but rather a limitation inherent in any valid comparison. If a characteristic Q can be adequately measured by two different methods M and N , both methods must be functions of Q and therefore functionally related to each other. In many cases, M and N , in addition to depending on Q , will also depend on other factors not common to both. A comparison of M and N for the determination of Q is then only valid under conditions in which the results yielded by M and N are solely governed by variations in Q , i. e., all noncommon factors must be held constant for all samples involved in the comparison. Failure to satisfy this condition will result in data of M and N that may well show significant correlation, but not necessarily a definite functional relationship either with each other or with the characteristic Q .

It is also important to note that the functional relationship assumed to exist between the methods M and N need not be known for the application of the sensitivity criterion.

4. Test of Significance for the Sensitivity Ratio

It has been shown that a measure of the relative merit of a test method M with respect to an alternative method N is given by the sensitivity ratio:

$$\frac{\psi_M}{\psi_N} = |K'| \frac{\sigma_N}{\sigma_M},$$

where K' is the slope of the curve of M versus N in the region of the curve at which the comparison is made. If this ratio exceeds unity, M is superior to N . Since, in general, both K' and the quantities σ_M and σ_N will be determined experimentally, the ratio ψ_M/ψ_N can only be approximated, and its estimate will be subject to random fluctuations.

In practice it is fortunately quite often the case that the two tests are carried out on the same sample or in such a manner that their relationship is known with much higher precision than either of the two measurements. Thus, a comparison of the relative merits of measuring the rate of tread wear of tires by weight loss or by depth loss can be made by measuring both losses on the same tire. While either of these experimental quantities depends on highly variable climatic and road conditions, the relation between the two is practically free from these effects because both are obtained under the same identical conditions.

In such cases, the fluctuations in the sensitivity ratio can be considered to be due entirely to the uncertainty in the ratio s_N/s_M where s is a sample estimate for the corresponding σ .

To determine whether the ratio $|K'| \sigma_N/\sigma_M$ exceeds unity, a statistical test is made of the hypothesis

$|K'| \sigma_N/\sigma_M = 1$, against the alternative hypothesis $|K'| \sigma_N/\sigma_M > 1$.

The quantity $F = (s_N^2/\sigma_N^2)/(s_M^2/\sigma_M^2)$ is known to be distributed in accordance with the F -statistic [11]. Consequently,

$$\frac{\sigma_N}{\sigma_M} = \frac{s_N}{s_M} \frac{1}{\sqrt{F}}$$

and

$$|K'| \frac{\sigma_N}{\sigma_M} = |K'| \frac{s_N}{s_M} \frac{1}{\sqrt{F}} \quad (11)$$

If F_0 is the tabulated value of the F -statistic at the desired level of significance, the quantity $|K'| (s_N/s_M) 1/\sqrt{F_0}$ represents a lower confidence limit for the sensitivity ratio $|K'| \sigma_N/\sigma_M$. If this lower limit exceeds unity, it may be concluded, at the confidence level chosen, that M is more sensitive than N .

In the example shown in table 1, the numbers of degrees of freedom used in the estimation of the standard deviations ranged from 38 to 48. Examining the data of plant A and the 100-minute cure, for which there were 48 degrees of freedom for each standard deviation, F_0 , at the 5 percent level of significance, equals 1.61; and consequently, the lower confidence limit of the sensitivity ratio equals

$$1.3 \frac{1}{\sqrt{F_0}} = 1.3 \frac{1}{\sqrt{1.61}} = 1.0.$$

From this value it can be concluded that strain, even in the least favorable of the cases examined, is at least as sensitive as stress, and most likely more sensitive.

If the experimental error in the estimate of the slope K' is not negligible, the above test of significance is not valid. In such cases, the correct statistical procedure for testing the significance of the sensitivity ratio depends on the type of relationship between the two test methods (linear, quadratic, logarithmic, etc.) as well as on the design of the experiment used to establish the relationship. No attempt is made in this paper to deal with the statistical theory for these more complex situations.

5. Effect of Scale of Measurement

There exist many cases in which measurements of physical or chemical properties can be expressed in more than one scale. For example, in measuring the light-absorption characteristics of materials, the results can be expressed either in optical density or in percentage transmittance. Another example is the measurement of refractive indices: In many instruments, a scale is provided that allows the direct reading of the refractive index rather than the angles of refraction and of incidence. In these cases the different scales of measurement correspond to functionally related quantities, but the functions

relating them are not linear. An important advantage of the sensitivity concept is its nondependence on the scale of measurement. The standard deviation, being expressed in the same units as the measurement, has a value that depends on the unit and scale in which the measurement is expressed. The coefficient of variation, which is defined as the ratio of the standard deviation to the mean value, is nondimensional, because both these quantities are expressed in the same units. However, except for scales that are proportional to each other, the coefficient of variation is dependent on the scale in which the measurement is expressed.

Consider, for example, the logarithmic transformation of a measurement y :

$$z = \ln y.$$

The standard deviation of z is then approximated [8] by the expression

$$\sigma_z = \frac{d \ln y}{dy} \sigma_y = \frac{\sigma_y}{y}.$$

It is evident, from this formula, that the coefficient of variation of z , σ_z/z , is in general different from that of y , σ_y/y . It can be shown that the only transformation that leaves the coefficient of variation rigorously unaltered is a proportional transformation: $z = ky$, i. e., a simple change of units. (To the extent that the approximate expression $\sigma_z = |dz/dy| \sigma_y$ is applicable—[for details see 12, secs. 27.7 and 28.4]—the coefficient of variation is also unaltered under the transformation $z = k/y$.)

On the other hand, the sensitivity of the transformed variable z , for *any* transformation

$$z = f(y) \quad (12)$$

is identical to that of the original variable y , to the

extent that the following calculation of the ratio of the two sensitivities is applicable:

$$\frac{\psi_z}{\psi_y} = \frac{\left| \frac{dz}{dy} \right|}{\sigma_z / \sigma_y} = \frac{\left| \frac{dz}{dy} \right|}{\left| \frac{dz}{dy} \right| \sigma_y / \sigma_y} = 1. \quad (13)$$

It is evident from eq (13) that sensitivity is not affected by any transformation of the measurement, and is therefore independent of the scale in which the measurement is expressed.

6. References

- [1] R. G. Newton, Proc. Second Rubber Technol. Conf., p. 233, Institution of the Rubber Industry, London (W. Heffer & Sons, Ltd., Cambridge, England, 1948).
- [2] M. C. Throdahl, J. Colloid Sci. **2**, 187 (1947); Rubber Chem. and Technol. **21**, 164 (1948).
- [3] J. H. Dillon, Physics **7**, 73 (1936); Rubber Chem. and Technol. **9**, 496 (1936).
- [4] F. L. Roth and R. D. Stiehler, J. Research NBS **41**, 87 (1948) RP1906; India Rubber World **118**, 367 (1948); Rubber Chem. and Technol. **22**, 201 (1949).
- [5] J. M. Buist and O. L. Davies, Trans. Inst. Rubber Ind. **22**, 68 (1946); Rubber Chem. and Technol. **20**, 288 (1947).
- [6] R. G. Newton, J. R. Scott, and R. W. Whorlow, Proc. Intern. Rheol. Congr. **II**, 204; **III**, 61 (1948).
- [7] E. Reichel, Z. anal. Chem. **109**, 385 (1937).
- [8] W. E. Deming, Statistical adjustment of data, ch. III (John Wiley & Sons, Inc., New York, N. Y., 1943).
- [9] ASTM Book of Standards, pt. 6, D-412-51T, p. 134 (American Society for Testing Materials, Philadelphia, Pa., 1952).
- [10] Private communication from the Office of Synthetic Rubber, Reconstruction Finance Corp.
- [11] R. L. Anderson and T. A. Baneroff, Statistical theory in research (McGraw-Hill Book Co., New York, N. Y., 1952).
- [12] H. Cramer, Mathematical methods of statistics (Princeton University Press, Princeton, N. J., 1946).

WASHINGTON, February 8, 1954.