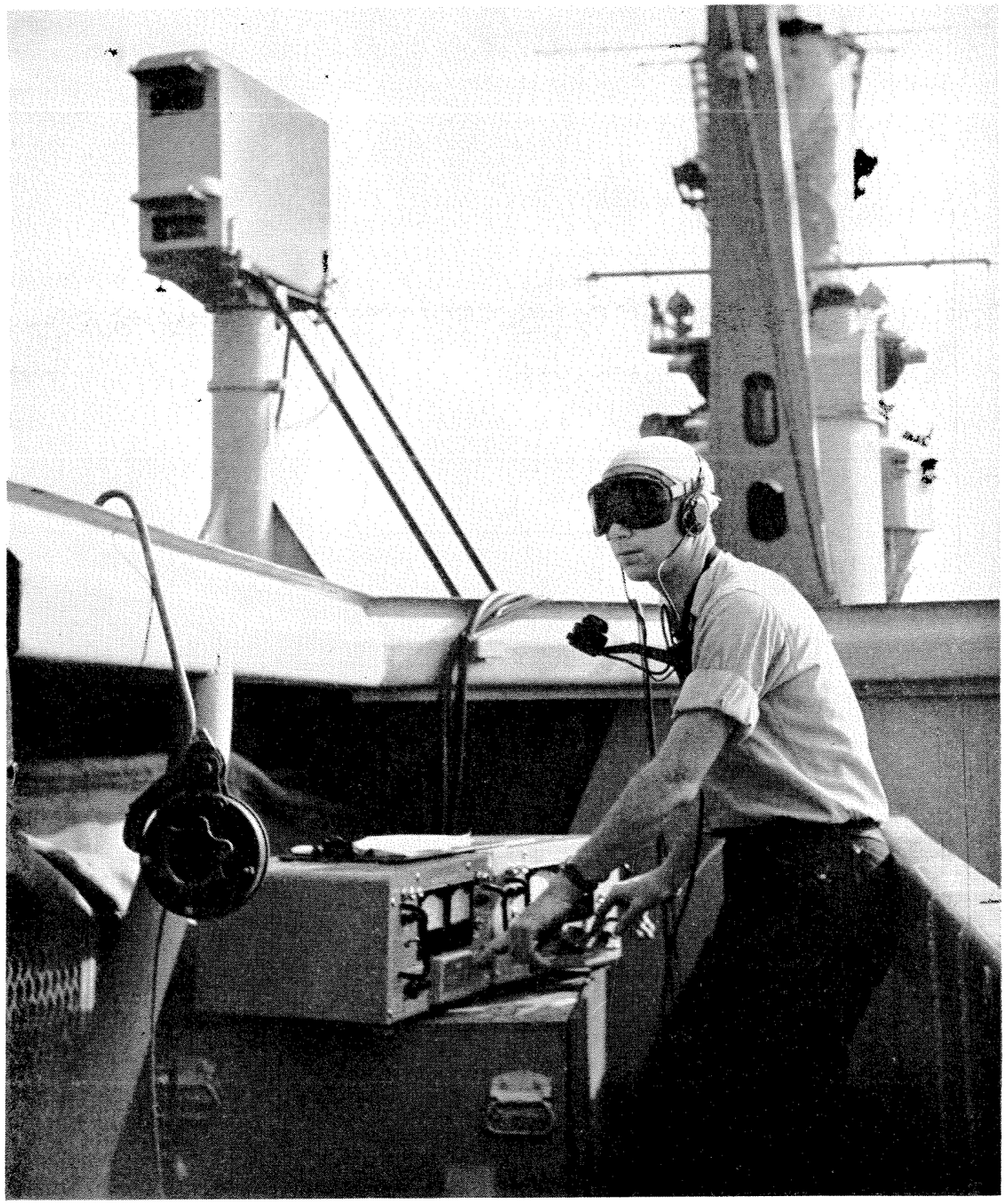

"Trodi"—Touchdown-Rate-of-Descent Indicator

(An electrical-optical brain developed by North American Aviation gives instant data on how fast an airplane "hits the deck" or lands. To seek rate of descent data, Trodi sends out two parallel beams of light, thin vertically and wide horizontally. They are one foot apart. A mirror system on the incoming airplane cuts the top beam, reflecting the light back to a photoelectric cell, which starts an electrical charge into a condenser. The descending airplane then cuts the second and lower beam, reflects it, and stops the charge going into the condenser. The electrical charge stored during the interval between beams is quickly translated by Trodi from voltage to rate of descent in feet per second.)

Describing UNCERTAINTIES *in* SINGLE-SAMPLE EXPERIMENTS

By S. J. KLINE[1] and F. A. McCLINTOCK[2]

Introduction

EVERYONE who uses the results of experiments must sooner or later ask, "How reliable are these results? To how many figures can they be depended upon?" In most cases the experimenter tries to eliminate all errors from his results, but we all know from experience that this happy goal is unattainable. Therefore the honest experimenter must provide the reader with some measure of the reliability of the results.

From the point of view of reliability estimates, experiments fall into two overlapping categories—single-sample and multiple-sample experiments. Ideally, we would like to repeat all measurements enough times using enough observers and enough diverse instruments so that the reliability of the results could be assured by the use of statistics. Experiments in which uncertainties are evaluated by such repetition will be called multiple-sample experiments. The estimation of reliability in multiple-sample experiments has been the subject of many publications. In particular, the American Society for Testing Materials has published a manual (1)[3] covering the presentation of results in controlled multiple-sample experiments. This manual has been available for twenty years and can serve as a standard for the presentation of the type of data covered.

Unfortunately, in most engineering experiments it is not practical to estimate all of the uncertainties of observation by repetition. If for no other reasons, the time required and the costs of operation and personnel are too great to permit repetition of all aspects of large-scale experiments. Experiments of the type in which uncertainties are not found by repetition will be called single-sample experiments.

There is almost nothing in print on methods for the description and analysis of uncertainties in single-sample experiments. The authors are not only unaware of a standard on the subject, but are unaware of any treatment reinforced by data covering an appreciable variety of experiments. Perhaps as a result of this many engineering colleges hardly mention the subject in their undergraduate curricula. The engineering literature in turn reflects this lack of instruction. Even a few of the society test codes which are scrupulous in other matters appear to be incorrect in the calculation of uncertainties. For this reason the authors hope that this presentation will start discussion of a possible standard, that it will stimulate research to provide badly needed data, and that it will encourage educators to review the adequacy of the treatment in their undergraduate curricula.

The scope and, consequently, the importance of single-sample experiments is much greater than at first might be imagined. This is due to four factors all of which tend to lessen the effect of repetition. Consequently, many experiments that appear to be multiple-sample are actually in part single-sample experiments.

The first of these factors was demonstrated by Pearson (2) who showed that observation of scales by a single observer, in general, did not give consistent results even though all extraneous variables appeared to have been removed. In particular, he showed that even a sample of 20 or 30 readings might have a mean value significantly different from the true mean as established by 500 or more samples. In the same paper Pearson also demonstrated that observations of scales by different observers are not necessarily independent due to some unexplained causes as well as to number bias, the tendency to read consistently high or low at certain points. Tuemmler (3) also has noted an apparent difference between the results of various laboratories using equipment of the same design to perform the same tests. The last, and perhaps the most significant factor, since it usually gives rise to the largest errors, is that instruments of different designs, in general, will not give the same results. Hence, if a single instrument is used for a set of observations, some error which is inherent in the instrument will be sampled only once, no matter how many times each reading is repeated.

In single-sample experiments it is inevitable that the statements of reliability will be based in part on estimates. This must be true since by definition statistics cannot be applied to all of the errors. Very often these estimates will be no better than ±50 per cent of the uncertainty; but ±50 per cent may be entirely satisfactory, particularly if the uncertainty is of the order of a few per cent or less of the original data.

A complete method for treatment of uncertainties in a given experiment must provide the answer to three questions: What is a rational way for estimating and describing the uncertainties in the variables? What is a proper method for calculating the propagation of these uncertainties into the results? What must be presented in a report to give a reasonably complete but concise picture of the reliability of the experiment?

Definition of Terms

Before proceeding, let us define certain terms more carefully. By "uncertainty" we mean a possible value the error might have.[4] For a single observation, the error, which is the difference between the true and observed values, is a certain fixed number. But the uncertainty, or what one thinks the error might be, may vary considerably depending upon the particular circumstances of the observation. "Variable" will mean a basic quantity observed directly in the laboratory as opposed to the "result" which is obtained by making corrections to or calculations with the recorded values of the variables. The recorded values of the variables will be called "data." In a few cases, of course, the results will be the same as the data. "Propagation of uncertainty" will mean the way in which uncertainties in the variables affect the uncertainty in the results. The

[1] Assistant Professor, Mechanical Engineering Department, Stanford University, Stanford, Calif.; formerly, Instructor, Mechanical Engineering Department, Massachusetts Institute of Technology, Cambridge, Mass. Jun. ASME.
[2] Assistant Professor, Mechanical Engineering Department, Massachusetts Institute of Technology. Jun. ASME.
[3] Numbers in parentheses refer to the Bibliography at the end of the paper.

[4] This important distinction between error and uncertainty is believed due to Airy (4).

terms standard deviation, mean, and frequency-distribution function will be used in the accepted statistical meanings as given for example by Hoel (5).

UNCERTAINTY DISTRIBUTION

In order to arrive at a rational method for describing the uncertainties in the variables, it is necessary to discuss the sources and nature of the uncertainties. There are many types of errors which can contribute to the uncertainty in each variable.

One useful classification of these errors is as follows: Accidental errors, fixed errors, and mistakes (6, 7). Accidental errors are those varying errors which cause repeated readings to differ without apparent reason. Accidental errors arise from instrument friction and time lag, personal errors, and many other causes. Fixed errors are those which cause repeated readings to be in error by the same amount without apparent reason. (If a reason were known, presumably a suitable correction would be made and the error eliminated.) Fixed errors arise from such causes as a burr on the lip of a Pitot tube or a lever arm of erroneous length. Mistakes are completely erroneous readings of scales, watches, and so on. Each of these types of errors will be considered in turn.

Accidental errors can be studied by taking repeated observations of the value of a variable. Such a sequence of readings may fall into various patterns, some of which are shown in

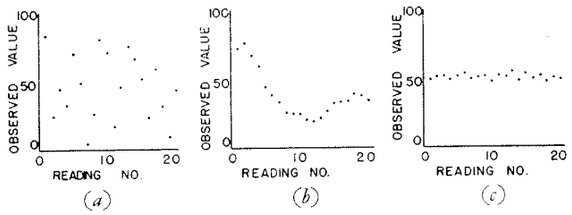

FIG. 1 EXPERIMENTAL OBSERVATIONS

Fig. 1. The sequences shown in Fig. 1(*a*, *b*) are uncontrolled or inhomogeneous; there is no telling what the trends are or how far they will go. The sequence of Fig. 1(*c*) is more predictable; even though there is a variation from one reading to another, the readings all tend to fall in a given region. Such a sequence is called homogeneous. The experiment from which it was obtained is said to be controlled. If a very large number of data are obtained, a frequency-distribution function can be constructed to describe them. The defining characteristic of the frequency-distribution function, $f(v)$, is that the fraction of values lying between v and $v + dv$ is $f(v)dv$. The distribution function corresponding to Fig. 1(*c*) is shown in Fig. 2. Accidental errors usually have a frequency distribution similar to that of Fig. 2, in that small errors are more likely than large ones and there is no definite upper limit to the possible size of an error.

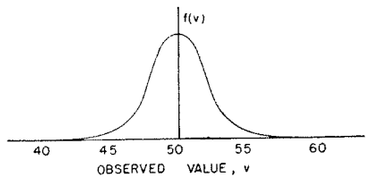

FIG. 2 FREQUENCY DISTRIBUTION

The frequency distribution for accidental errors has been assumed to be normal, or Gaussian, by many authors. While this may be true in many cases, it is not true in all. Cases of distribution functions for accidental errors which are non-normal in form (2, 8) et al., are now too well documented to assume that all such distributions will be normal.

In the case of a small number of observations, it is no longer possible to describe the distribution function exactly. Yet it is still possible to make some precise operational statements about the characteristics of the distribution function, using the statistical concept of a confidence limit.

No measure of the scatter of the errors can be obtained from a single-sample experiment. Consequently, the experimenter must rely on his past experience and judgment. The best he can do is make a statement of what he thinks would happen if the experiment were repeated an indefinitely large number of times. It would seem natural to make such a statement in the language of probability and statistics. But statistics are properly based on calculations with measured numbers, while statements regarding single-sample experiments must be based on what one thinks these numbers will be. Hence a statistical description would be misleading. Therefore, the term "uncertainty distribution" will be used instead of frequency distribution.

The uncertainty distribution is the distribution of errors which the experimenter believes would be found in a given variable if the variable were sampled a great many times. The uncertainty distribution caused by accidental errors could be measured by repeated readings of the variable in question. The uncertainty distribution then would be exactly the same as the frequency distribution. Thus the existing frequency distributions for accidental errors can be used to determine what the shape of uncertainty distributions for accidental errors should be.

The few available frequency distributions indicate that the uncertainty distribution caused by accidental errors should have a shape similar to the frequency distribution shown in Fig. 2. Therefore its general characteristics are also the same; namely, small errors are more likely than large errors, plus and minus errors are about equally likely, and no finite maximum error exists.

A close examination of fixed errors shows that the classification of errors as accidental and fixed is really a relative matter. As an example, consider the use of a thermocouple and precision potentiometer to measure furnace temperature. Errors may result from variations in temperature within the furnace both in time and space, from the effect of the thermocouple on the local temperature, from incorrect calibration of the wire or potentiometer, from deterioration of the wire or drift of the potentiometer-battery calibration during use, and from personal errors in balancing and reading the potentiometer. An examination of these errors shows that which errors are called fixed and which are called accidental depends largely on the scope of the experiment. If one observer, using one thermocouple and one potentiometer, were to make readings of temperature at one point, then almost all of the errors must be considered fixed even if the reading were taken many times. However, if several observers, using several different types of temperature-measuring devices, measured the temperature repeatedly at a given point, almost all of the errors could be called accidental. If the same type of apparatus must be used by all observers, a fixed error may remain.

For example, if only one thermocouple installation could be used in the foregoing experiment an error due to the distortion of the temperature field by the thermocouple will not be sampled. Such errors can be estimated by theoretical means, and the uncertainty in these calculations can be thought of as having an uncertainty distribution. Thus the fixed errors have an uncertainty distribution which can be visualized in terms of calculations and the use of more instruments and observers. Consequently, the uncertainty distribution developed in connection with the uncertainties resulting from accidental errors

can be applied directly to the uncertainties which result from fixed errors. The shape of the uncertainty distribution due to fixed errors is also believed to be similar to Fig. 2.

Mistakes are those errors which result from completely erroneous readings of watches, scales, and so on. These errors, in general, will be discarded by a careful observer if they are very large. Consequently, small errors are more likely than large ones, positive and negative errors are about equally likely, and no finite maximum error can be stated. Although the errors tend to occur in discrete steps, they do have an uncertainty distribution which can be visualized in terms of the use of many observers and scale intervals.

The entire error in a given reading due to all of the causes mentioned has no distribution function. It is just a certain finite number. On the other hand, the entire uncertainty or lack of knowledge about the value of a reading can be described completely in terms of an uncertainty distribution since each of its components can be described in this way. However, there is insufficient knowledge of uncertainties to warrant the use of a complete uncertainty distribution for each variable. In addition, the distribution would be too cumbersome for routine use and would require difficult if not impossible mathematics for the calculation of the uncertainties in the results. Some shorthand notation is needed which is consistent both with the concept of the uncertainty distribution outlined and with the state of the experimenter's knowledge of the uncertainties, and yet which is simple enough for routine use.

UNCERTAINTY INTERVAL FOR A VARIABLE

A satisfactory notation for the uncertainty of a variable must include a statement of the best estimate of the true value as well as a statement about the magnitude of the error in the estimate. The best estimate of the true value is usually described by giving the mean of the readings.

A simple but adequate description of the error in the estimate is more difficult to frame. In the case of frequency distributions, the statistician often uses the standard deviation. But use of a standard deviation to describe uncertainties has two distinct disadvantages. (*a*) For nearly normal distributions, it describes an interval such that the odds are approximately 2 to 1 that the error in a particular reading will lie inside the interval. However, the experimenter usually wants his odds to be at least 10 or 20 to 1, rather than 2 to 1. (*b*) It would be misleading to use the term standard deviation which connotes a root-mean-square value calculated from actual measured numbers because in single-sample experiments the numbers must be estimated.

Another measure of scatter sometimes employed is the range. This measure has been employed in some of the existing literature under the name of "maximum error." In the present nomenclature the maximum error would be called the "maximum uncertainty." This concept may have meaning to a manufacturer who must achieve complete interchangeability of parts. In critical manufacturing cases 100 per cent inspection is often used to force a maximum error.[5] But in the case of experimental uncertainty 100 per cent inspection cannot be applied. If just one of the great number of causes for error in a given variable has an uncertainty distribution with very long tails, then the final uncertainty distribution also must have long tails. Since every sampled distribution known to the authors has long tails, it is most unlikely that any uncertainty distribution for the entire error in a given variable can be described properly by a maximum uncertainty. One might argue that there is a maximum value of the uncertainty which the error will never exceed; but even granting this, a rigorous interpretation of the "never" almost always leads to values of the uncertainty which make the experiment unacceptable.

Another method of notation for describing a distribution which is better suited for description of uncertainty distributions is to specify an interval based on certain odds.

For example, the distribution of Fig. 2 indicates that the odds are roughly 20 to 1 that any given reading will lie within ± 4 of the mean of the distribution. Conversely, if only the value of a single reading is known, the position of the mean can be described by "20 to 1 the mean of the distribution lies within ± 4 of the reading." The odds the experimenter would be willing to wager on his estimate of where the true value lies would depend on how large the interval was made.

In the case cited, for example, he would be willing to wager only 1 to 2 that the true value lies within ± 1 of a given reading, but he would be willing to bet 20 to 1 that it lies within ± 4, or 100 to 1 that it lies within ± 10. This method of description is flexible in that the experimenter can set his odds to conform with the standards of reliability required by any given experiment. A complete description of the uncertainty distribution could be given by the statement of the odds associated with all possible intervals. But a statement of just one interval is all that is justified by our limited knowledge and it does provide a reasonable index of the reliability.

Considering the various factors just discussed, the authors believe that a good concise way to describe the uncertainty in each variable is to specify the mean of the readings and an uncertainty interval based on specified odds. Representing the mean by m (arithmetic mean of observed values), the uncertainty interval by w, and the odds by b, this becomes

$$m \pm w, (b \text{ to } 1) \quad [1]$$

As an example one might give

$$\text{Pressure} = 50.2 \pm 0.5 \text{ psia } (20 \text{ to } 1)$$

This states that the best value for the pressure is believed to be 50.2 psia and the odds are 20 to 1 that the true value lies within ± 0.5 psia of this best estimate. The uncertainty interval, which is denoted by w, is not a variable but a fixed value selected so that the experimenter would be willing to wager b to 1 that the error is less than w.

Determination of the actual value of the uncertainty interval based on given odds is one of the jobs of the experimenter. As already noted, at least some of these intervals will have to be based on estimates rather than experiments, and the estimates often may be no better than ± 50 per cent. Despite this, the experimenter owes it to himself and to his readers to go ahead and do the best he can; no one else is in an equally good position to make the required estimates which are essential to rational design and to interpretation of the results.

Such estimates are, of course, not pure guesses. Factors such as instrument backlash, sensitivity, and fluctuation, as well as the accuracy of the basic theory of operation of the instrument, sometimes can be accounted for. Calibration of the instrument against some type of standard is sometimes available, and experience based on prior experiments or auxiliary experiments can be used. This part of the subject is covered already in the standard textbooks on instrumentation and is too detailed in nature for adequate treatment here. Readers desiring further information should see (6, 8, 10) and other texts, manufacturers' catalogs, and the literature of their specialty.

Equation [1] together with the foregoing discussion gives a method by which the experimenter can describe the uncertainties in each of the basic variables in what the authors believe to be a sufficiently accurate and simple manner for routine use. It is then necessary to determine how these uncertainties propagate into the results.

[5] Even in these cases we know from experience that assembly sometimes fails. An enlightening discussion of this subject including cost effects is given by Pike and Silverberg (9).

UNCERTAINTY INTERVAL IN THE RESULT

Let the result R be a function of n independent variables, $v_1, v_2, \ldots v_n$

$$R = R(v_1, v_2 \ldots\ldots v_n) \qquad [2]$$

For small variations in the variables, this relation can be expressed in linear form as

$$\delta R = \frac{\partial R}{\partial v_1}\delta v_1 + \frac{\partial R}{\partial v_2}\delta v_2 + \ldots\ldots + \frac{\partial R}{\partial v_n}\delta v_n \qquad [3]$$

The uncertainties in the variables v_i are represented completely by an uncertainty distribution but can be adequately described by uncertainty intervals w_i based on certain odds. Therefore we must examine how to find the uncertainty interval for the result w_R based on essentially the same odds as the intervals for each of the variables. Certain theorems of statistics concerning the way in which frequency distributions combine will be helpful in finding a reasonable value for w_R.

Theorem 1. If R is a linear function of n independent variables, and if the maximum deviation of the ith variable from its mean is $(\pm\delta v_i)_{max}$ then the maximum deviation of R from its mean value is given by

$$\delta R_{max} = \left|\frac{\partial R}{\partial v_1}\delta v_{1_{max}}\right| + \left|\frac{\partial R}{\partial v_2}\delta v_{2_{max}}\right| + \ldots\ldots + \left|\frac{\partial R}{\partial v_n}\delta v_{n_{max}}\right| \qquad [4]$$

Equation [4] might be used as an approximation for calculating the uncertainty interval in the result by simply substituting w_i for v_i. This yields

$$w_R = \left|\frac{\partial R}{\partial v_1}w_1\right| + \left|\frac{\partial R}{\partial v_2}w_2\right| + \ldots\ldots + \left|\frac{\partial R}{\partial v_n}w_n\right| \qquad [5]$$

This equation will be referred to as the linear equation. If it is employed, the odds on the uncertainty interval in the result will be much higher than the odds used in the variables. This is because of the fact that the errors in each variable can have a range of values, and it is quite unlikely that all of them will have the most adverse values at the same time.

Theorem 2. If R is a linear function of n independent variables, each of which is distributed with a standard deviation σ_i, then the standard deviation of R is given by

$$\sigma_R = \left[\left(\frac{\partial R}{\partial v_1}\right)^2\sigma^2{}_1 + \left(\frac{\partial R}{\partial v_2}\right)^2\sigma^2{}_2 + \ldots\ldots + (\partial R/\partial v_n)^2\,\sigma^2{}_n\right]^{1/2} \qquad [6]$$

We have seen, however, that the best measure of uncertainty is neither the maximum value nor the standard deviation, but some interval based on certain odds. For the special case in which the variables are distributed normally the distribution of the result also will be normal, and the following theorem applies:

Theorem 3. If R is a linear function of n independent variables, each of which is normally distributed, then the relation between the interval for the variables w_i, and the interval for the result w_R, which gives the same odds for each of the variables and for the result is

$$w_R = \left[\left(\frac{\partial R}{\partial v_1}w_1\right)^2 + \left(\frac{\partial R}{\partial v_2}w_2\right)^2 + \ldots\ldots + \left(\frac{\partial R}{\partial v_n}w_n\right)^2\right]^{1/2} \qquad [7]$$

Equation [7] might be used directly as an approximation for calculating the uncertainty interval in the result. Equation [7] will be referred to as the second-power equation.

Examples were calculated to compare the accuracy of the linear and second-power equations for predicting the appropriate interval in the result in the case of different distributions.

Three different frequency-distribution functions were chosen, one normal, one corresponding to one wave length of a sine curve, and the last corresponding to an isosceles triangle. Since the latter two distributions have finite limits, and are considerably less normal than the distributions usually considered in connection with uncertainties, they constitute a severe test of the generality of the second-power equation. Odds of 9 to 1, 19 to 1, and 99 to 1, were chosen as being of interest in experimental engineering work. The mathematics employed in calculating the exact distribution function of the result are outlined in the Appendix.

For Table 1 the result was considered to be proportional to the sum of two variables. According to this table, the second-power equation gives odds nearer to the desired odds in every case. The second-power equation predicts the uncertainty intervals for the result to within ±10 per cent of the correct value, while the linear equation predicts uncertainty intervals varying from the correct value by as much as 40 per cent.

The difference between intervals given by the linear and second-power equations increases as the square root of the number of variables if each variable has about equal effect on the result.

For Table 2 the result was taken to be proportional to the sum of an infinite number of variables. This table shows even more clearly the superiority of the second-power equation. The odds given by the second-power equation for the uncertainty interval in the result are still reasonable while the linear equation gives infinite odds. The error in the interval introduced by use of the second-power equation increases to no more than 15 per cent, while the error due to use of the linear equation becomes infinite. Since the state of knowledge of uncertainty intervals for the variables is of the order of ±50 per cent, it seems entirely reasonable to use the second-power equation for the calculation of the uncertainty interval for the result.

The second-power equation, Equation [7], is also of importance in planning instrumentation. It applies in this sense both to single- and multiple-sample experiments since statistics cannot be applied to a multiple-sample experiment in advance of the tests. It shows that uncertainties in individual variables add into the uncertainty in the result by the square. Consequently, the effect of large uncertainties in the variables is emphasized, and a given reduction in a large uncertainty is far more important than the same numerical reduction in a small uncertainty. The second-power equation is thus a useful tool in the selection of instrumentation for experiments.

EXAMPLE

As an illustration of the ideas presented in the foregoing, consider the measurement of velocity with a Pitot tube in an air stream. If c denotes the velocity, Δp the pressure difference between the Pitot tube and the atmosphere, and p_a and T_a the pressure and temperature of the air, respectively, then Bernoulli's equation and the perfect-gas equation of state give

$$c = \sqrt{\frac{2(\Delta p)RT_a g_0}{p_a}} \qquad [8]$$

Here the result is the velocity and the variables are Δp, p_a, and T_a. Even in this apparently simple measurement there are a great many possible sources of error which give rise to the uncertainties. Some of them are alignment of the Pitot tube in the flow, leaks in the pressure tubing, changes in bore, or differences in surface condition of the manometer measuring the pressure difference, fluctuations in atmospheric and stream pressure, and so on. In this case the Mach number must not be too high nor the Reynolds number too low. In either case the assumption of incompressible frictionless flow inherent in Bernoulli's equation is violated.

Let us suppose that the Mach number and the Reynolds number have proper values. In order to use Equation [7] all uncertainty intervals for the variables must be based on the same odds. Twenty to one will be used. If T_a is measured by a calibrated mercury-in-glass thermometer, p_a with a Bourdon gage, and Δp with a U-tube manometer, a description of the readings might be

$\Delta p = 8.0 \pm 0.1$ in. H_2O (20 to 1)
$T_a = 67.4$ F $= 527.1 \pm 0.2$ deg F abs (20 to 1)
$p_a = 14.7 \pm 0.3$ psia (20 to 1)

Evaluating the $(\partial R/\partial v_i)$ terms and substituting into Equation [7]

$$w_R = \left[\frac{1}{4}\frac{2RT_a g_0}{(\Delta p)p_a}(w_{\Delta p})^2 + \frac{1}{4}\frac{2(\Delta p)RT_a g_0}{p_a{}^3}(w_{p_a})^2 + \frac{1}{4}\frac{2(\Delta p)Rg_0}{p_a T_a}(w_{T_a})^2\right]^{1/2} \qquad [9]$$

Equation [9] is greatly simplified upon dividing by Equation [8] to nondimensionalize

$$\frac{w_c}{c} = \frac{w_R}{R} = \left[\left(\frac{1}{2}\frac{w_{\Delta p}}{\Delta p}\right)^2 + \left(\frac{1}{2}\frac{w_{p_a}}{p_a}\right)^2 + \left(\frac{1}{2}\frac{w_{T_a}}{T_a}\right)^2\right]^{1/2} \qquad [10]$$

This nondimensional form is simpler in most cases. It can

TABLE 1 DISTRIBUTION OF THE RESULT: $R = (v_1 + v_2)/\sqrt{2}$

	Normal	Sinusoidal	Triangular
Distribution of variable	$f(v) = \frac{1}{\sqrt{2\pi}}e^{-v^2/2}$ for all v	Define $A = \sqrt{\pi^2/3 - 2}$; then $f(v) = A(1 + \cos Av)/2\pi$ for $\lvert v\rvert \le \pi/A$	$f(v) = 1/\sqrt{6} - \lvert v\rvert/6$ for $\lvert v\rvert \le \sqrt{6}$
Distribution of result	$f(R) = \frac{1}{\sqrt{2\pi}}e^{-R^2/2}$ for all R	$f(R) = \frac{A}{2\pi^2\sqrt{2}}[(3/2)\sin\sqrt{2}AR + (\pi - AR/\sqrt{2})(2 + \cos\sqrt{2}AR)]$ for $0 \le \lvert R\rvert \le \sqrt{2}\pi/A$	$f(R) = \frac{2}{3\sqrt{3}} - \frac{R^2}{3\sqrt{3}} + \frac{\lvert R\rvert^3}{18}$ for $0 \le \lvert R\rvert \le \sqrt{3}$ and $f(R) = \frac{4}{3\sqrt{3}} - \frac{2\lvert R\rvert}{3} + \frac{R^2}{3\sqrt{3}} - \frac{\lvert R\rvert^3}{54}$ for $\sqrt{3} \le \lvert R\rvert \le 2\sqrt{3}$
Desired odds	9:1	19:1	99:1	9:1	19:1	99:1	9:1	19:1	99:1
Actual odds for interval given by second-power equation	9:1	19:1	99:1	8.3:1	16.6:1	51:1	9.4:1	17.2:1	42:1
Actual odds for interval given by linear equation	49:1	178:1	370:1	54:1	257:1	6620:1	75:1	300:1	7500:1
Correct interval of result for desired odds	1.64	1.96	2.58	1.68	1.95	2.44	1.66	1.94	2.44
Interval given by second-power equation	1.64	1.96	2.58	1.62	1.88	2.26	1.68	1.90	2.20
Interval given by linear equation	2.32	2.77	3.64	2.29	2.66	3.20	2.37	2.69	3.12

TABLE 2 DISTRIBUTION OF THE RESULT: $R = \lim_{n\to\infty}(v_1 + v_2 + \ldots + v_n)/\sqrt{n}$

	Normal	Sinusoidal	Triangular
Distribution of variable	$f(v) = \frac{1}{\sqrt{2\pi}}e^{-v^2/2}$ for all v	Define $A = \sqrt{\pi^2/3 - 2}$; then $f(v) = A(1 + \cos Av)/2\pi$ for $\lvert v\rvert \le \pi/A$	$f(v) = 1/\sqrt{6} - \lvert v\rvert/6$ for $\lvert v\rvert \le \sqrt{6}$
Distribution of result	$f(R) = \frac{1}{\sqrt{2\pi}}e^{-R^2/2}$ for all R	$f(R) = \frac{1}{\sqrt{2\pi}}e^{-R^2/2}$ for all R	$f(R) = \frac{1}{\sqrt{2\pi}}e^{-R^2/2}$ for all R
Desired odds	9:1	19:1	99:1	9:1	19:1	99:1	9:1	19:1	99:1
Actual odds for interval given by second-power equation	9:1	19:1	99:1	8.5:1	15.7:1	41:1	9.6:1	16.4:1	35:1
Actual odds for interval given by linear equation	∞:1	∞:1	∞:1	∞:1	∞:1	∞:1	∞:1	∞:1	∞:1
Correct interval of result for desired odds	1.64	1.96	2.58	1.64	1.96	2.58	1.64	1.96	2.58
Interval given by second-power equation	1.64	1.96	2.58	1.62	1.88	2.26	1.68	1.90	2.20
Interval given by linear equation	∞	∞	∞	∞	∞	∞	∞	∞	∞

be obtained as shown previously, or alternatively by use of logarithmic differentiation on Equation [8], or by substituting $c + w_R$ for c, $p_a + w_{pa}$ for p_a, and so on, multiplying out and neglecting terms in w^2.

When the numbers are substituted into Equation [10] using consistent units, one obtains

$$w_c/c = 1/2\,[1.08 \times 10^{-4} + 4.15 \times 10^{-4} + 1.44 \times 10^{-7}]^{1/2} = 1.1 \text{ per cent}$$

This calculation illustrates that if the $\partial R/\partial v$ terms are neglected, as is sometimes done, serious errors occur. In this case w_c/c would have been in error by a factor of 2. It also shows that improving the measurement in the temperature would not change w_c/c appreciably, but a 50 per cent reduction in w_c/c could be obtained simply by using a manometer instead of a Bourdon gage to measure p_a. This illustrates the utility of the second-power equation in determining which variables need most attention in improving the accuracy of an experiment.

Presentation of Uncertainties in a Report

The questions, "How should uncertainties in the variables be estimated and described?" and "How do the uncertainties propagate into the results?" have now been discussed and conclusions reached. The answer to the final question, "What should be presented in a report?" cannot be given definitively since it is subject to the demands of space and time as well as the practices of societies and publishers. The authors feel that presentation of uncertainty intervals for the results as found by Equation [7] along with the odds used should be sufficient in most cases. For more elaborate reports the uncertainty intervals assigned to each variable also might be useful.

Conclusion

The method suggested here can be summarized as follows:

1 Describe the uncertainty in each variable as

$$\text{mean} \pm \text{uncertainty interval (odds of } b \text{ to 1)} \quad \ldots\ldots [1]$$

2 Compute the uncertainty interval in each result as

$$w_R = \sqrt{\left(\frac{\partial R}{\partial v_1} w_1\right)^2 + \left(\frac{\partial R}{\partial v_2} w_2\right)^2 + \ldots + \left(\frac{\partial R}{\partial v_n} w_n\right)^2} \quad \ldots [7]$$

3 Present at least the value of w_R and the chosen odds for each result as an integral part of a report or paper.

The value of w_R found in this way will be based on essentially the same odds as the uncertainty intervals in the variables. The only important restriction is that the uncertainties in each of the variables must be independent.

The method thus summarized provides a means for describing and analyzing the uncertainties in single-sample experiments. In this method the actual estimation of the uncertainty intervals must still depend on the judgment of the experimenter. At present this judgment can be acquired only by laboratory experience since data on the total uncertainty interval in most instruments are unavailable. A great many engineering experiments are part single-sample and part multiple-sample.

In such cases the available repeated measurements should be analyzed by statistical methods to supplement the judgment of the experimenter. In either case this method should provide a simple and useful tool by which the experienced investigator can more accurately describe and analyze experimental uncertainty in both the laboratory and design stages of his work.

BIBLIOGRAPHY

1 "ASTM Manual on the Presentation of Data," Third Printing, February, 1949.

2 "On the Mathematical Theory of Errors of Judgment With Special Reference to the Personal Equation," by Karl Pearson, Philosophical Transactions of the Royal Society of London, series A, vol. 198, 1902, pp. 235–299.

3 Discussion of reference (12), by F. D. Tuemmler.

4 "Theory of Errors of Observation," by Sir George Bidell Airy, Macmillan and Company, London, England, 1879.

5 "Introduction to Mathematical Statistics," by P. G. Hoel, John Wiley & Sons, Inc., New York, N. Y., 1947.

6 "Experimental Mechanical Engineering," by H. Diederichs and W. C. Andrae, John Wiley & Sons, Inc., New York, N. Y., vol. 1, 1930, chapter 1.

7 "Numerical Mathematical Analysis," by J. B. Scarborough, second edition, Johns Hopkins Press, Baltimore, Md. 1950.

8 "Applied Experimental Psychology," by A. Chapanis, et al., John Wiley & Sons, Inc., New York, N. Y., 1949.

9 "Tolerance Management," by P. R. Pike and E. W. Silverberg (Raytheon Corporation), presented before Boston Section ASME, October 25, 1951.

10 "Theory of Measurements," by A. DeF. Palmer, McGraw-Hill Book Company, Inc., New York, N. Y., 1912.

11 "Introduction to the Theory of Probability and Statistics," by N. Arley and K. R. Buch, John Wiley & Sons, Inc., New York, N. Y., 1950.

12 "Precision and Accuracy of Test Methods," by G. Wernimont, Symposium on Applications of Statistics, ASTM, San Francisco, Calif. October 11, 1949.

13 "Accuracy of Visual Interpolation . . . ," by M. Leyzoreck, *Journal of Applied Psychology*, vol. 33, April, 1949, p. 270.

14 "Theory and Methods for Analysis of Errors in Machine Systems," by A. Chapanis, Annals New York Academy of Sciences, vol. 51, 1951, article 7.

15 "On Estimating Precision of Measuring Instruments and Product Variability," by F. E. Grubbs, *Journal of the American Statistical Association*, vol. 43, 1948, pp. 243–264.

16 "The Design of Experiments," by R. A. Fisher, Oliver & Boyd, Edinburgh, Scotland, 1949.

APPENDIX

The three different frequency-distribution functions chosen were all assumed to have zero mean and unit standard deviation. The equations for these distributions are presented in the headings of Tables 1 and 2. Outside the regions specified, the frequency-distribution functions are zero.

For Table 1 the result was chosen to be proportional to the sum of two variables

$$R = (v_1 + v_2)/\sqrt{2} \quad \ldots\ldots [11]$$

The factor $1/\sqrt{2}$ is introduced simply to give the result a standard deviation of unity; it does not affect the comparison between the linear and second-power equations. The frequency-distribution function of this result can be found by evaluating the following integral[6]

$$f(R) = \int f\left(\sqrt{2}R - v_1\right) f(v_1)\, \sqrt{2}\, dv_1 \quad \ldots\ldots [12]$$

Calculation of the distribution function of the sum of more than two variables becomes quite tedious. For the limiting case of an infinite number of variables, however, the resulting distribution is normal,[7] and a comparison becomes easy to make. The result will have a unit standard deviation if one takes R to be

$$R = \lim_{n \to \infty} (v_1 + v_2 + \ldots + v_n)/\sqrt{n} \quad \ldots\ldots [13]$$

[6] Reference (11), pp. 40–48.
[7] Ibid., p. 108.