PSYCHOMETRIKA—VOL. 21, NO. 3
SEPTEMBER, 1956

# THE RELATION BETWEEN INFORMATION AND VARIANCE ANALYSES*

W. R. GARNER

THE JOHNS HOPKINS UNIVERSITY

AND

WILLIAM J. MCGILL

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Analysis of variance and uncertainty analysis are analogous techniques for partitioning variability. In both analyses negative interaction terms due to negative covariance terms that appear when non-orthogonal predictor variables are allowed may occur. Uncertainties can be estimated directly from variances if the form of distribution is assumed. The decision as to which of the techniques to use depends partly on the properties of the criterion variable. Only uncertainty analysis may be used with a non-metric criterion. Since uncertainties are dimensionless (using no metric), however, uncertainty analysis has a generality which may make it useful even when variances can be computed.

## I. *Introduction*

Shannon (5) has defined amount of information by the formula

$$H(y) = -\sum_{k=1}^{r} p(k) \log_2 p(k), \tag{1}$$

where $y$ has $r$ discrete values, and $p(k)$ is a probability distribution defined over $y$. In communication theory, $y$ is considered a source of signals, and the measure $H$ represents the average number of binary digits required to code or store one of the signals. A broader interpretation, however, makes $H$ a parameter which measures the non-metric variability of any probability distribution. $H$ has a value of zero when the probability is concentrated in a single category and is maximum when the probability is uniformly distributed over all categories.

Psychologists have been attracted by the non-metric character of this measure and the obvious application to situations where variances cannot be computed. Since this use of the measure is concerned only with its statistical properties and not with its interpretation in communication theory, we

*The work of the senior author was supported by Contract N5ori-166, Task Order 1, between the U.S. Office of Naval Research and The Johns Hopkins University. This is Report No. 166-I-192, Project Designation No. NR 145-089, under that contract.

shall use the more general term uncertainty, $U$, to refer to the measure. We shall show that uncertainty has many of the properties of variance and can be partitioned into components as variance can.

## II. *The Analysis Problem*

The relations discussed apply when a criterion is predicted from one or more predictors. The development will be presented for the three-variable case, where the problem is to determine to what extent values of the criterion variable can be predicted from two predictor variables.

Our notation is as follows: The criterion variable $y$ can assume any value $y_k$. The two predictor variables $w$ and $x$ can assume values $w_i$ or $x_j$. We assume that all three variables are categorized in order that the formulas for uncertainty and variance analysis may have equivalent notations. This assumption does not limit any of the principles demonstrated.

In the three-dimensional matrix, $n_{ijk}$ refers to the number of cases in a single cell; $n_{ij.}$ refers to the total number of cases having the $i$th value of $w$ and the $j$th value of $x$; and $n_{i..}$ refers to the total number of cases having the $i$th value of $w$. Similar subscripts indicate other combinations of the three variables; $n$ with no subscript indicates the total number of cases in the matrix. In analysis of variance formulas, $\bar{y}$ indicates a mean value, and the subscript notation just illustrated is used for mean values of the subclassifications.

## III. *The Nature of Uncertainty Analysis*

Analysis of variance can be considered as two separate processes. First, the variance of the criterion variable is partitioned into its several identifiable components—components which add up to the total variance. This process is a simple descriptive one; there are no probability assumptions involved in its use. One describes the components of a total variance, making no assumptions about the distributions from which the data are drawn. The second process, which is not a necessary consequence of the first, involves using these partitioned components to obtain estimates of population variances and to make inferences about the parent population. For this process, the actual data provide sample estimates of population distributions; here assumptions about the population distributions become critical.

Uncertainty analysis likewise has both processes. The first process is purely descriptive: it is intended to allow the partitioning of the uncertainty of the criterion variable $U(y)$ into components. Since this process is entirely descriptive, there are no underlying probability assumptions. All that is required for its use is that a data matrix of the type described above is available. The primary purpose of this paper is to demonstrate the nature of uncertainty partitioning and to compare it to variance partitioning. This process is illustrated and explained in Table 1. The results of uncertainty partitioning specify sources and magnitudes of variabilities as well as amount of categorical discrimination available. These uses are explained more fully by Garner and Hake (**1**) and by McGill (**3**).

The $n_{ijk}$ can be considered as sample estimates of $p(i, j, k)$; we can use these sample estimates to test various hypotheses about the parent distribution. For example, suppose we wish to test the hypothesis that both predictors are independent of the criterion, i.e.,

$$p(i, j, k) = p(i, j)p(k). \tag{2}$$

It can be shown (**3**, **4**) by using the likelihood ratio, that when hypothesis (2) is true, [1.3863 $nU(y{:}\ w, x)$] is distributed approximately as chi square. Independent tests can be constructed in the same way for each of the predictors separately as well as for the interaction between predictors. The approximation to chi square is of the same order as the familiar chi-square contingency test so that, in effect, uncertainty analysis is analysis of contingency chi square. Miller and Madow (**4**) discuss this aspect of uncertainty analysis more thoroughly.

## IV. *The Orthogonal Case*

Usually in analysis of variance and in uncertainty analysis, the experimenter tries to set up orthogonal predictions. Orthogonality is defined as zero association between the predictors. This requirement is met when the cell frequencies in the matrix of the $n_{ij.}$ can be predicted correctly from the row and column marginal frequencies, i.e., when

$$n_{ij.} = \frac{n_{i..}n_{.j.}}{n}. \tag{3}$$

### *Uncertainty Analysis*

The partitioning of $U(y)$ in uncertainty analysis is illustrated by

$$U(y) = U(y{:}\, w, x) + U_{wx}(y), \tag{4}$$

where the uncertainty measures have the definitions given in Table 1. The second term on the right-hand side of (4) is the error uncertainty, i.e., the amount of uncertainty in the criterion $y$ remaining after the predictable uncertainty has been eliminated. The first term on the right-hand side of (4) is the predictable uncertainty; it in turn can be partitioned into components

$$U(y{:}\, w, x) = U(y{:}\, w) + U(y{:}\, x) + U(y{:}\, \overline{wx}). \tag{5}$$

These terms are also defined in Table 1. A feature of uncertainty analysis is the interaction term $U(y{:}\, \overline{wx})$. This is the uncertainty in $y$ predictable from unique combinations of $w$ and $x$.

Equation (5) describes a process that is identical in form with the

TABLE 1

Symbols, Formulas, and Definitions Used in Three-variable Uncertainty Analysis

The criterion variable, $y$, is assumed to be non-metric. The predictor variables, $w$ and $x$, are categorized and may or may not be metric variables.

| | Symbol | Formula | Definition |
|---|---|---|---|
| (1) | $U(y)$ | $-\sum_{k}\left[\frac{n_{..k}}{n}\right]\log_2\left[\frac{n_{..k}}{n}\right]$ | *Total uncertainty:* The amount of uncertainty in the criterion variable, $y$. |
| (2) | $U_w(y)$ | $-\sum_{i}\left[\frac{n_{i..}}{n}\right]\sum_{k}\left[\frac{n_{i\cdot k}}{n_{i..}}\right]\log_2\left[\frac{n_{i\cdot k}}{n_{i..}}\right]$ | *Conditional uncertainty:* The amount of uncertainty in $y$ when one predictor variable, $w$, is held constant. |
| (3) | $U_{wx}(y)$ | $-\sum_{i,j}\left[\frac{n_{ij\cdot}}{n}\right]\sum_{k}\left[\frac{n_{ijk}}{n_{i\ \cdot}}\right]\log_2\left[\frac{n_{ijk}}{n_{ij\cdot}}\right]$ | *Error uncertainty:* The amount of uncertainty remaining when both predictor variables, $w$ and $x$, are held constant. |
| (4) | $U(y{:}w)$ | $U(y)-U_w(y)$ | *Contingent uncertainty:* The uncertainty in $y$ due to the predictor variable, $w$. |
| (5) | $U_w(y{:}x)$ | $U_w(y)-U_{wx}(y)$ | *Partial contingent uncertainty:* The uncertainty in $y$ due to the predictor variable, $x$, when the predictor variable, $w$, is held constant. |
| (6) | $U(y{:}w,x)$ | $U(y)-U_{wx}(y)$ | *Multiple contingent uncertainty:* The uncertainty in $y$ due to the joint influence of the predictor variables, $w$ and $x$. |
| (7) | $U(y{:}\overline{wx})$ | $-U(y)+U_w(y)+U_x(y)-U_{wx}(y)$, *or* $U_w(y{:}x)-U(y{:}x)$ | *Interaction uncertainty:* The uncertainty in $y$ due to unique combinations of the predictor variables, $w$ and $x$. |

TABLE 2

Symbols, Formulas, and Definitions Used in Three-variable Analysis of Variance

The criterion variable, $y$, is assumed to be metric. The predictor variables, $w$ and $x$, are categorized but not necessarily metric.

| | Symbol | Formula | Definition |
|---|---|---|---|
| (1) | $V(y)$ | $\sum_k \left[\frac{n_{..k}}{n}\right] (y_k - \bar{y})^2$ | *Total variance:* The variance of the criterion variable, $y$. |
| (2) | $V_w(y)$ | $\sum_i \left[\frac{n_{i..}}{n}\right] \sum_k \left[\frac{n_{i\cdot k}}{n_{i..}}\right] (y_k - \bar{y}_{i..})^2$ | *Conditional variance:* The variance of $y$ when one predictor variable, $w$, is held constant. |
| (3) | $V_{wx}(y)$ | $\sum_{i,j} \left[\frac{n_{ij\cdot}}{n}\right] \sum_k \left[\frac{n_{ijk}}{n_{ij\cdot}}\right] (y_k - \bar{y}_{ij\cdot})^2$ | *Error variance:* The variance remaining in $y$ when both predictor variables, $w$ and $x$, are held constant. |
| (4) | $V(y{:}w)$ | $V(y) - V_w(y)$ | *Main effect:* The variance of $y$ due to the predictor variable, $w$. |
| (5) | $V_w(y{:}x)$ | $V_w(y) - V_{wx}(y)$ | *Partial main effect:* The variance of $y$ due to the predictor variable $x$, when the predictor variable, $w$, is held constant. |
| (6) | $V(y{:}w,x)$ | $V(y) - V_{wx}(y)$ | *Total predictable variance:* The variance of $y$ due to the joint influence of the predictor variables, $w$ and $x$. |
| (7) | $V(y{:}\overline{wx})$ | $V(y) + V_w(y) + V_x(y) - V_{wx}(y)$, *or* $V_w(y{:}x) - V(y{:}x)$ | *Interaction variance:* The variance of $y$ due to unique combinations of the predictor variables, $w$ and $x$. |

partitioning of variance in analysis of variance; in the orthogonal case the interaction uncertainty can be interpreted by analogy with interaction variance. This is true despite the fact that interaction uncertainties are sometimes negative (**3**). This problem will be discussed in detail in Section V.

*Analysis of Variance*

Uncertainty analysis is generally appropriate when the criterion variable $y$ is a categorical variable, i.e., one allowing only nominal scale values (cf. **6**). The predictor variables may be categorical, or they may be metric variables which are categorized for purposes of analysis. If the criterion is a true metric variable, i.e., one having at least the properties of an interval scale, we can compute variances and perform analysis of variance. The predictor variables must be categorized in any simple form of the analysis of variance.

Equations describing analysis of variance are essentially identical to those of uncertainty analysis. The defining equations are given in Table 2; except for the fact that variances are computed from squared deviations, whereas uncertainties are computed from log-probabilities, the equations are identical to those in Table 1. The partition of the variance of the criterion can be written:

$$V(y) = V(y\colon w, x) + V_{wx}(y). \tag{6}$$

Again the two parts on the right-hand side of the equation are the predictable and the error components of the total variance. The predictable variance can be broken down as before:

$$V(y\colon w, x) = V(y\colon w) + V(y\colon x) + V(y\colon \overline{wx}). \tag{7}$$

The terms in (7) are explained in detail in Table 2.

Normally the analysis of variance in (7) is called double classification; the variances are generally identified in terms of the two predictors. This shorthand procedure is convenient for most purposes. However, it obscures the fact that the data array is three-dimensional. The analysis is identical to the one treated in uncertainty analysis in every respect, except that in the analysis of variance the criterion variable has a metric, whereas it does not in uncertainty analysis.

## V. *The Non-Orthogonal Case*

In Section IV it was mentioned that the interaction term in uncertainty analysis can assume negative values under certain conditions. It is equally true that the interaction term in analysis of variance can be negative, if it is defined as in Table 2. The negative interaction term is due to non-orthogonality and can be thought of as due to a negative covariance term that may attenuate or exceed the positive interaction effect.

*Uncertainty Analysis*

It is not difficult to show that the interaction uncertainty in (5) can be written

$$U(y: \overline{wx}) = U_y(w: x) - U(w: x). \tag{8}$$

This form of the interaction term shows at once that interaction cannot be negative with orthogonal predictors since orthogonality requires that $U(w: x) = 0$.

In the non-orthogonal case, however, $U(w: x)$ will be greater than zero. With certain combinations of cell frequencies, the contingent uncertainty between $x$ and $w$ can be larger than the partial contingent uncertainty—resulting in negative interaction. A simple illustration of this principle is provided when each value of $w$ is paired uniquely with each value of $x$. Now $U(w: x)$ is as large as it can be. Furthermore, $U_y(w: x)$ cannot be greater than $U(w: x)$ since $U(w: x)$ is the maximum contingent uncertainty that can be obtained from a contingency table involving $w$ and $x$. Equation (8) shows that the interaction will never be greater than zero. An identical result is obtained in the variance analysis when the predictors are completely confounded.

*Analysis of Variance*

It is usually assumed that the components of the total variance in analysis of variance must be positive. This is true only in the orthogonal case; if an analysis of variance is carried out with a non-orthogonal experimental design, using the equations given in Table 2, negative interaction terms can occur.

To show how this happens, we now analyze the components of the interaction variance for the general case. The equation is

$$\begin{aligned} V(y: \overline{wx}) = \frac{1}{n} \sum_{i,j} n_{ij.}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y})^2 \\ - \frac{2}{n} \sum_{i,j} \left(n_{ij.} - \frac{n_{i..}n_{.j.}}{n}\right)(\bar{y}_{i..} - \bar{y})(\bar{y}_{.j.} - \bar{y}). \end{aligned} \tag{9}$$

It can be seen that the interaction variance is composed of two parts: the first part is essentially the interaction variance in the orthogonal case; the second part is a negative covariance term. This term must be zero in the orthogonal case [see equation (3)], but in the non-orthogonal case it cannot be ignored. The redundancy introduced by non-orthogonality is illustrated clearly in multiple regression. No interaction term is permitted, but a correction for non-orthogonality must be introduced whenever the predictor variables are correlated (cf. **2**).

## VI. *Effects of Non-Orthogonality*

Our discussion of non-orthogonality shows that it is best to design experiments with orthogonal predictor variables. The analysis is simplified, and the uninterpretable interaction components are eliminated.

Clearly the covariance in (9) is not just part of the interaction variance. In fact, when predictors are non-orthogonal, the concept of interaction is almost meaningless. For example, consider an analysis of variance in which the predictors $w$ and $x$ are completely confounded. The two main-effect variances and the interaction will all be identical. The covariance term in (9) must be large enough to cancel out two of these variances, but we do not know which two of the variances should be cancelled out. In a sense, the covariance term is a correction factor which must be applied to the entire set of variances. Thus, a covariance term (whether or not it is large enough to produce a negative interaction) renders an exact interpretation of the component variances impossible.

The multiple contingent uncertainty or the total predictable variance can be computed directly as shown in the defining equations in Tables 1 and 2. The negative covariance term is included; there is no over-estimation of the total predictable variance or uncertainty. However, the interpretation of results should be made only in terms of combinations of the two predictors —no valid statements can be made about them independently.

Sometimes it is impossible to obtain orthogonal predictor variables, particularly when there are more than two. In time series successive events are usually not orthogonally related because no independent control of these events is possible. If the time series has serial dependencies, preceding events cannot be orthogonal. Consequently, the total predictability of events in a time series cannot in general be computed by adding up the separate predictabilities obtained from preceding events displayed by one or more units in the time series.

## VII. *Estimation of Uncertainties from Variances*

It is clear that uncertainty analysis and analysis of variance are analogous analytic techniques. In fact, variances may be used to estimate uncertainties if we assume that $y$ is normally distributed.

Shannon (5) has shown that the uncertainty of a normal distribution can be specified as

$$\text{est } U(y) = \tfrac{1}{2} \log_2 2\pi e V(y) - \log_2 m, \tag{10}$$

where est $U(y)$ is the estimated total uncertainty of the criterion variable on the assumption of a normal distribution of values of $y_k$, and where $m$ is the width of the category interval on the $y$ continuum.

We can write similar equations for any of the variances obtained in analysis of variance. For example,

$$\text{est } U_{wx}(y) = \tfrac{1}{2} \log_2 2\pi e V_{wx}(y) - \log_2 m \tag{11}$$

is the error uncertainty estimated from error variance. From definition (6) in Table 1, and from equations (10) and (11), we can write

$$\text{est } U(y\colon w, x) = \tfrac{1}{2} \log_2 [V(y)/V_{wx}(y)]. \tag{12}$$

Thus, it is relatively simple to estimate the multiple contingent uncertainty from the appropriate variances. The expression on the right-hand side of this equation is reminiscent of the multiple correlation ratio ($\eta$). We can, in fact, write

$$\text{est } U(y\colon w, x) = -\tfrac{1}{2} \log_2 [1 - \eta^2(y\colon w, x)]. \tag{12–A}$$

Estimated uncertainties have the properties of additivity observed in computed uncertainties. Consequently, the expression on the right-hand side of (12) can be partitioned into three components, each of which is based on its equivalent variances as follows:

$$\text{est } U(y\colon w) = \tfrac{1}{2} \log_2 [V(y)/V_w(y)], \tag{13}$$

$$\text{est } U(y\colon x) = \tfrac{1}{2} \log_2 [V(y)/V_x(y)], \tag{14}$$

$$\text{est } U(y\colon \overline{wx}) = \tfrac{1}{2} \log_2 \{[V_w(y) \cdot V_x(y)]/[V(y) \cdot V_{wx}(y)]\}. \tag{15}$$

These estimating equations point out some of the differences between uncertainty and variance. If (15) is used to estimate the interaction uncertainty when the interaction variance is zero, cases can be found in which the estimated interaction uncertainty (and the computed interaction uncertainty) will not be zero. Converse cases (i.e., zero uncertainty interactions with finite variance interactions) can also be found. These apparent contradictions are due to the fact that variances and uncertainties, while analogous, do not measure exactly the same characteristics of probability distributions. Uncertainty analysis depends on the number of categories occupied by a distribution. Variance analysis depends on the weights or values attached to these categories.

## VIII. *Application of the Measures*

We have now shown that uncertainty analysis and analysis of variance are equivalent in many respects; the question naturally arises as to when one should be used in preference to the other. This decision depends on the properties of the data and the assumptions the experimenter is willing to make. If the criterion variable $y$ has only the properties of a nominal or ordinal scale, then only uncertainty analysis is permissible. Uncertainty analysis has the greater generality and requires no assumptions about metric properties of the criterion.

On the other hand, uncertainty analysis does not give any information about the metric if it exists. If the criterion variable is metric with at least the properties of an interval scale, then analysis of variance must be used to retain information about the metric. The variance measure in retaining the metric sacrifices generality since the variances obtained from one experiment are not directly comparable to those obtained from another. Thus, the fact that the uncertainty measure is dimensionless gives it a generality which allows direct comparison of experimental results which differ in their metric.

To summarize, the measures are similar in many respects, but they are not identical. The uncertainty measure has greater generality and the advantages of generality. The variance measure is more specific but retains information about the metric. The decision as to which to use depends not only upon the properties of the criterion variable but also upon the gain expected from being more sensitive instead of more general. In many applications it is reasonable to use both measures and compare them.

## REFERENCES

1. Garner, W. R. and Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446-459.
2. Garner, W. R. and McGill, W. J. Relation between uncertainty, variance, and correlation analyses. Mimeographed report, Department of Psychology, The Johns Hopkins University, Baltimore, Maryland.
3. McGill, W. J. Multivariate information transmission. *Psychometrika*, 1954, **19**, 97-116.
4. Miller, G. A. and Madow, W. G. On the maximum likelihood estimate of the Shannon-Wiener measure of information. Report AFCRC-TR-54-75. Operational Applications Laboratory, ARDC, August 1954.
5. Shannon, C. E. A mathematical theory of communication. *Bell syst. tech. J.*, 1948, **27**, 379-423, 623-656.
6. Stevens, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951, pp. 1-49.

*Manuscript received 10/9/54*

*Revised manuscript received 9/23/55*