OAK RIDGE NATIONAL LABORATORY
operated by
UNION CARBIDE CORPORATION
NUCLEAR DIVISION
for the
U.S. ATOMIC ENERGY COMMISSION

ORNL-TM-1919

INDICES OF QUALITATIVE VARIATION

Allen R. Wilcox

DISCLAIMER

ORNL-TM-1919

Contract No. W-7405-eng-26

Director's Division
CIVIL DEFENSE RESEARCH PROJECT*

INDICES OF QUALITATIVE VARIATION

Allen R. Wilcox

OCTOBER 1967

OAK RIDGE NATIONAL LABORATORY
Oak Ridge, Tennessee
operated by
UNION CARBIDE CORPORATION
for the
U.S. ATOMIC ENERGY COMMISSION

* This research was supported in part by the Advanced Research Projects Agency, Office of the Secretary of Defense. The views expressed in this article are the author's own and do not necessarily reflect those of ORNL or the USAEC.

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK

CONTENTS

Prefatory Note

The statistical exploration reported in this memorandum was undertaken to help us to better implement our responsibilities for the analysis of American public opinion on defense policy in general, and continental defense in particular. As familiarity with surveys will certify, many of the most interesting questions and sets of responses only meet nominal criteria, i.e., cannot be placed in relative positions on a scale. Accordingly, the analyst who wishes to treat these questions with a desirable degree of rigor and sophistication needs measures appropriate to them, in the author's phrase, measures of qualitative variation. This memorandum is a modest effort to assist the analyst who desires to improve his ability to explore essentially qualitative data. As such, we hope it will be helpful to a variety of social scientists who confront such data in the course of empirical research.

Davis B. Bobrow

INDICES OF QUALITATIVE VARIATION

Allen R. Wilcox

I. INTRODUCTION

In textbook treatments of measures of variation, statistics are always mentioned which measure the variation of a univariate distribution when the variable under consideration satisfies the requirements of an ordinal, interval, or ratio scale. Most commonly, the range, the semi-interquartile range, the average deviation, the standard deviation, and the variance are presented and discussed. However, the presentation and discussion of measures of variation suitable for use with variables that satisfy only the requirements of a nominal scale is often completely absent. In addition, there appears to be no unified discussion of those few appropriate measures scattered through the literature. Based on the assumption that such measures may have considerable utility for the statistical handling of qualitative data, this paper represents a first attempt to gather together and to generate alternative indices of qualitative variation. The treatment is introductory throughout. A more intensive mathematical and empirical treatment of these measures will hopefully be provided by others who wish to probe more deeply into the characteristics and utility of the statistics.

II. FORMAL PROPERTIES OF THE INDICES

The discussion is limited to measures that satisfy certain formal conditions. The first three of these conditions represent what are considered to be desirable formal properties of such indices. First and second, the maximum and/or minimum values that such an index may obtain should not depend on the magnitude of either of the two basic parameters of a qualitative distribution--the number of cases and the number of categories. These conditions facilitate comparison of the values of a particular index, even when they are derived from radically different distributions. Third, they must all have a standard range of values: in this case, from 0 to 1. This condition facilitates comparison <u>among</u> indices for the <u>same</u> distribution.

The final condition specifies the forms that a distribution must have when an index based on it obtains the maximum and minimum values. With the number of cases placed at 100 and the number of categories at 4, Fig. 1 illustrates with histograms the minimum-related and maximum-related forms, respectively. Thus, the minimum value of 0 occurs if, and only if, all cases fall within one category. Conversely, the maximum value of 1 occurs if, and only if, an identical number of cases fall within each category.[1] This condition delineates the general "type" of measure to which this paper is being addressed. It is assumed that an index that does not satisfy this condition can be more usefully related to some other concept or idea. Measures that satisfy this condition can be thought of as indices of "generalized qualitative variation."[2]

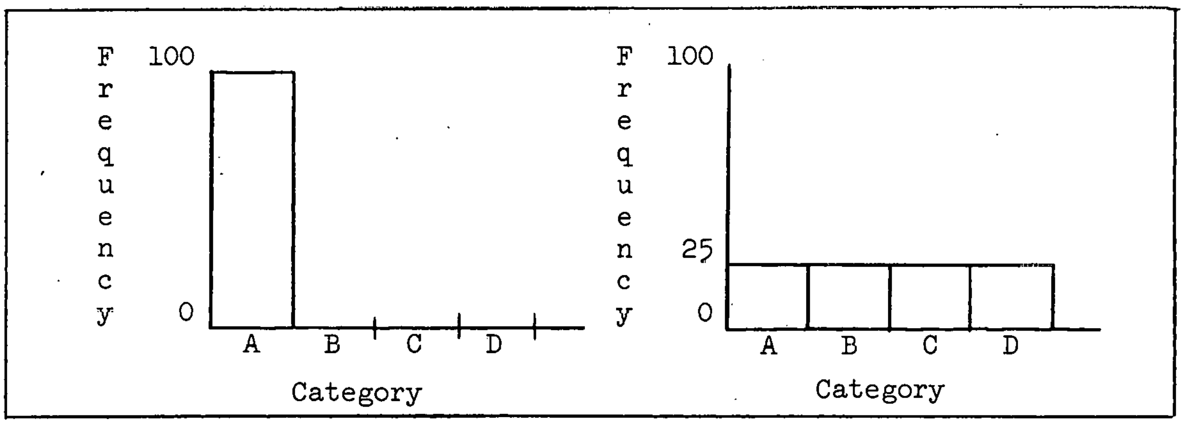

Fig. 1

[1] As a consequence of this condition, the indices discussed herein reach 1 only when the number of categories divides evenly into the number of cases. This circumstance impares the comparability of the indices in certain cases but does not diminish their general usefulness.

[2] For a description of a measure of qualitative variation that does not satisfy this condition, see George J. McCall and J. L. Simmons, "A New Measure of Attitudinal Opposition," Public Opinion Quarterly, 30:2 (Summer 1966) 271-278. A critique of this measure can be found in David Gold, "Critical Note on a New Measure of Attitudinal Opposition," Public Opinion Quarterly, 31:1 (Spring 1967) 76-79.

III. DESCRIPTION OF THE INDICES

MODVR

The first index that we have developed, MODVR, is an index of deviation from the mode, analogous to the standard deviation as a measure of deviation from the mean. The core of this index is the sum of the differences between the non-modal frequencies and the modal frequency, i.e.,

$$\sum_{i=1}^{K} (f_m - f_i)$$

where f_i = the frequency of the ith category, and

f_m = the frequency of the modal category.

Since $\sum_{i=1}^{K} f_i = N$ and $\sum_{i=1}^{K} f_m = Kf_m$, the formula can be simplified to $Kf_m - N$.

Dividing this expression by N(K-1) standardizes the range of the index (K = the number of categories and N = the number of cases).

This expression is directly related to Linton Freeman's variation ratio, v.[3] The formula for v is $1 - \frac{f_m}{N}$. According to Freeman, the variation ratio is to be used as an aid in judging the adequacy of the mode as a summarizer of the distribution. It is simply the proportion of nommodal cases. The higher its value, the less adequate the mode is as a summarizer. As inspection of its formula will confirm, v can achieve a minimum of 0, but it can never achieve a maximum of 1 because the modal frequency can never be 0. However, by dividing and subtracting by appropriate correction factors, $\frac{f_m}{N}$ can be standardized to

$$\frac{\frac{f_m}{N} - \frac{1}{K}}{\frac{\frac{N}{K}(K-1)}{N}}$$ which reduces to $$\frac{Kf_m - N}{N(K-1)}$$, the core of MODVR. MODVR, in other words, is essentially a standardized variation ratio.

---

[3]Linton C. Freemen, Elementary Applied Statistics, New York: John Wiley and Sons, 1965, pp. 40-43.

Finally, the core of this index (and of most of the others which follow) is subtracted from 1 so that low values will stand for low variation and high values for high instead of vice versa. Thus, MODVR[4] equals

$$1 - \frac{Kf_m - N}{N(K-1)}$$

The next five indices are all variations on a theme, for they are all basically analogs of measures that are used on ordinal or interval data. This is done by treating the frequencies of the K categories as values of a quantitative variable, calculating the particular statistic, and finally standardizing the result to conform to the criteria listed above. In this way indices are constructed that are analogous to the range, average deviation, mean difference, variance, and standard deviation.

RANVR

The first measure of this type that we have developed is based on the range and is designated RANVR. Its basic component is the difference between the highest or modal frequency and the lowest frequency. The formula[5] follows:

$$1 - \frac{(f_m - f_\ell)}{f_m}$$

where f_m = the modal frequency, and
f_ℓ = the lowest frequency.

AVDEV

The third index that we have generated, AVDEV, is an analog of the average or mean deviation, which is defined as the arithmetic mean of the absolute differences of each value from the mean. Utilizing this

---

[4] A computational formula is $\frac{K(N-f_m)}{N(K-1)}$.

[5] A computational formula is $\frac{f_\ell}{f_m}$.

concept on qualitative data, the formula[6] is

$$1 - \frac{\sum_{i=1}^{K} \left| f_i - \frac{N}{K} \right|}{2 \frac{N}{K} (K-1)} .$$

MNDIF

The next index that we have developed, MNDIF, is an analog of the mean difference, a measure of variation that is discussed much less frequently than the average deviation or standard deviation. It is defined as "the average of the differences of all the possible pairs of variate-values, taken regardless of sign."[7] The mean difference differs from the average and standard deviations in that it is "dependent on the spread of the variate-values among themselves and not on the deviations from some central value."[8] For qualitative data, this notion can be formulated as

$$1 - \frac{\sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \left| f_i - f_j \right|}{N(K-1)} .$$

---

[6] A somewhat simpler formula for computational purposes is

$$1 - \frac{K \sum_{i=1}^{K} \left| f_i - \frac{N}{K} \right|}{2N(K-1)} .$$

[7] Maurice G. Kendall and Alan Stuart, The Advanced Theory of Statistics, Vol. 1, New York: Hafner Publishing Company, 1958, p. 46.

[8] Ibid., p. 47.

MNDIF can be approached in other ways mathematically. Suppose that the numbers 1 to K are assigned to the K categories in ascending order of magnitude. Given this weighting, two indices of variation can be derived. The first is based on subtraction and essentially sums the category weights multiplied by the differences between category proportions and the average proportion, i.e.,

$$1 - \frac{\sum_{i=1}^{K} i \left(\frac{f_i}{N} - \frac{1}{K}\right)}{K\left(1-\frac{1}{K}\right) - \sum_{i=1}^{K-1} \frac{i}{K}}$$

The second is similar,[9] but is based on division, i.e.,

$$1 - \frac{\left(\frac{\sum_{i=1}^{K} i \frac{f_i}{N}}{\frac{1}{K}}\right) - \frac{K(K+1)}{2}}{\frac{K(K-1)}{2}}$$

We derived both of these indices from an intuitive conception of what seemed, on inspection of a number of distributions, to be useful measures of variation--specifically, measures that take into account differences among the frequencies of all K categories. Given the

[9] A computational formula for both this and the previous formula is

$$1 - \frac{2\sum_{i=1}^{K} i\, f_i - N(K+1)}{N(K-1)}$$

similarity of this conception to the definition of a mean difference, it is in retrospect not surprising that they and MNDIF are algebraically equivalent.[10]

VARNC

The fifth index, VARNC, is an analog of the variance, which is defined as the arithmetic mean of the squared differences of each value from the mean. For use with qualitative data, its formula[11] is

$$1 - \frac{\sum_{i=1}^{K}\left(f_i - \frac{N}{K}\right)^2}{\frac{N^2 (K-1)}{K}}$$

This formula is quite similar to AVDEV, the basic difference being that the differences are squared instead of having their absolute values taken.

---

[10] It might be noted that another measure, Gini's index of concentration, is also equivalent to MNDIF when it is modified for qualitative data. For a discussion of Gini's index, see Hayward R. Alker and Bruce M. Russett, "On Measuring Inequality," Behavioral Science, 9:3 (July 1964) 207-218.

[11] A computational formula is

$$1 - \frac{\sum_{i=1}^{K} f_i^2 - \frac{N^2}{K}}{N^2 (K-1)}$$

which can be reduced even further to

$$\frac{K\left(N^2 - \sum_{i=1}^{K} f_i^2\right)}{N^2 (K-1)}$$

If MNDIF is approached in like manner and a squared version derived, the resulting formula is

$$1 - \frac{\sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \left(f_i - f_j\right)^2}{N^2(K-1)}$$

and is equal to VARNC.[12]

VARNC can be approached in another way through a discussion of Mueller and Schussler's Index of Qualitative Variation (IQV).[13] They base this index on the idea that the total number of differences among the categories (or in their terminology, items) of a given variable provides a means of measuring qualitative variation. This total is obtained by counting the differences between each category and every other category and summing these differences. Thus, using sex as an example of qualitative variable, "if there were nine boys and three girls, each of the nine boys would differ from each of the three girls, producing 27 differences."[14] They reduce this counting procedure to the following rule: multiply every category frequency by every other category frequency and sum these products. Although the authors do not

---

[12] This is similar to the case for quantitative data. As Kendall and Stuart note, a squared mean difference in the quantitative case is equal to twice the variance. Op cit., p. 47.

[13] John H. Mueller and Karl F. Schuessler, Statistical Reasoning in Sociology, Boston: Houghton Mifflin Company, 1961, pp. 177-179.

[14] Ibid., p. 177.

present a general formula,[15] one can be constructed as follows:

$$\frac{\sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \left(f_i f_j\right)}{\frac{K(K-1)}{2}\left(\frac{N}{K}\right)^2}$$

This formula is also algebraically equivalent to VARNC.

Finally, VARNC can also be derived by standardizing χ^2. A typical formula for χ^2 is

$$\sum_{i=1}^{K} \frac{(O_i - E_i)^2}{E_i}$$

where O_i = the observed frequency for the *i*th category, and

E_i = the expected frequency for the *i*th category.

Substituting the notation used in this paper, the formula is

$$\sum_{i=1}^{K} \frac{(f_i - \frac{N}{K})^2}{\frac{N}{K}}$$

---

[15] A computational formula:

$$\frac{2\,K \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} (f_i f_j)}{N^2(K-1)}$$

and, when standardized, it becomes

$$1 - \frac{\dfrac{\sum_{i=1}^{K}\left(f_i - \frac{N}{K}\right)^2}{N}}{\dfrac{N(K-1)}{K}}$$

which reduces to VARNC.

STDEV

STDEV, the final analog index that we have developed, is based on the standard deviation, which is defined as the square root of the variance. As with VARNC, a formula can be derived by starting either with AVDEV or with MNDIF. From AVDEV, the formula[16] is

$$1 - \sqrt{\frac{\sum_{i=1}^{K} (f_i - \frac{N}{K})^2}{(N - \frac{N}{K})^2 + (K-1)(\frac{N}{K})^2}}$$

From MNDIF, the formula is

$$1 - \sqrt{\frac{\sum_{i=1}^{K-1} \sum_{j=i+1}^{K} (f_i - f_j)^2}{N^2(K-1)}}$$

[16] A computational formula:

$$1 - \sqrt{\frac{\sum_{i=1}^{K} f_i^{\,2} - \frac{N^2}{K}}{\dfrac{N^2(K-1)}{K}}}$$

HREL

The seventh and last index of qualitative variation is HREL, a measure originally developed by engineers for use in specifying the properties of communications channels. The rationale for HREL is presented in terms of guessing by Virginia Senders (supplementing the mode as best guess):

> What we need is a measure of uncertainty, or of "poorness of a guess," which will be high when the number of alternative possibilities is high, and low when some of the possibilities are much more likely than others. One possible measure is the _average number of questions we have to ask to specify the correct alternative_.[17]

A formula for such a measure of uncertainty, H, is

$$-\sum p_i \log_2 p_i$$

where p_i = the proportion of cases in the _ith_ category.[18] This is a measure of actual uncertainty. To obtain a measure of uncertainty which can be compared across distributions, actual uncertainty must be divided by the maximum possible uncertainty. Maximum possible uncertainty occurs when the alternatives are equally likely (i.e., when all category frequencies are the same) and is equal to $\log_2 N$. The resulting formula for a measure of relative uncertainty is

---

[17] Virginia L. Senders, _Measurement and Statistics_, New York: Oxford University Press, 1958, p. 79. For additional material on informational measures of this type, see Fred Attneave, _Applications of Information Theory to Psychology_, New York: Henry Holt and Co., 1959, and Henry Quastler (ed.), _Information Theory in Psychology_, Glencoe: The Free Press, 1955.

[18] The units in which H is measured are called bits. Senders has this to say about the term:

> In (communications) channels and networks, relays and tubes are among the most important components. A typical relay can be either open or closed, a tube can fire or not. Thus, relays and tubes correspond to men giving answers to true-false or yes-or-no questions, and a binary (base 2) system seems to be the most convenient one to use. H is therefore measured in binary digits, which has been shortened to "bits."

$$\frac{-\sum p_i \log_2 p_i}{\log_2 K} \text{.}$$ [19]

IV. SUMMARY AND CONCLUSION

The statistics described in the preceding pages provide a considerable range of alternative measures of generalized qualitative variation. Of these, only HREL and VARNC (differently formulated) have been taken entirely from the statistical literature. The idea behind MODVR was present in the form of a non-standardized index. The remaining four indices (RANVR, AVDEV, MNDIF, and STDEV) have not, to our knowledge, previously appeared in the literature. Because of this and because our objective has been one of preliminary presentation, a number of tasks remain for those who would continue the examination of indices of this type. First, inquiry into sampling distributions has evidently been conducted only for HREL. Second, further exploration into mathematical interrelationships might provide insights into the comparative theoretical relevance of the indices. Finally, and most importantly, an accumulation of knowledge on the comparative utility of these statistics in various research contexts would be highly desirable.

[19] For ease in computation using mathematical tables, one can use either of the following formulas:

$$\frac{-\sum p_i \log_{10} p_i}{\log_{10} K} \qquad \frac{-\sum p_i \log_e p_i}{\log_e K}$$

APPENDIX A

Input Preparation, Sample Problem, and Output for the NOMSTAT Computer Program

This appendix presents a description of and sample input and output from NOMSTAT, a computer program which calculated the seven statistics described in the body of the text. Input--the input variables read by NOMSTAT are described in Table 1. The sequence in which these variables are read and the format by which they are read is given in Table 2.

Table 1. Nomstat Input Variable Description

Input Variable		Description
1.	TITLE	Alphanumeric title identification
2.	NDIST	Number of distributions
3.	NPROP	1 if statistics are to be in proportion form, any other number if in percentage form
4.	NPCNT	1 if category sizes are to be put out in frequency form, any other number if in percentage form
5.	NPUNCH	1 if statistics are to be punched as well as printed, any other number if not
6.	NSUM	1 if summary statistics (means, standard deviations, and coefficients of variation) are not to be calculated, any other number if they are
7.	NPCT	1 if input is in percentage form, any other number if in frequency form
8.	NCAT	Vector containing the number of categories in each distribution
9.	NN	Vector containing the number of observations in each distribution--used only if category sizes are in percentage form
10.	IVAR	Input FORMAT under which data matrix is read
11.	FREQ, or PCNT	Data Matrix. Which of these is read in depends on input variable 7

Table 2. Nomstat Input Variable Sequence and Format

	Input Variables	Card Format	Number of Cards
1.	TITLE	FORMAT (20A4)	1
2.	NDIST, NPROP, NPCNT, NPUNCH, NSUM, NPCT	FORMAT (6I3)	1
3.	NCAT	FORMAT (40I2)	5
4.	NN, only if NPCT = 1	FORMAT (20I4)	10
5.	IVAR	FORMAT (20A4)	1
6.	PCNT (If NPCT = 1) or FREQ (If NPCT ≠ 1)	IVAR	Variable

Output - The output variables printed by NOMSTAT are described in Table 3. The sequence in which these variables are printed is given in Table 4.

Table 3. Nomstat Output Variable Description

Output Variable		Description
1.	TITLE	Alphanumeric title identification
2.	NDIST	Number of distributions
3.	NPROP	1 if statistics are to be in proportion form, any other number if in percentage form
4.	NPCNT	1 if category sizes are to be put out in frequency form, any other number if in percentage form
5.	NPUNCH	1 if statistics are to be punched as well as printed, any other number if not
6.	NSUM	1 if summary statistics (means, standard deviations, and coefficients of variation) are not to be calculated, any other number if they are
7.	NPCT	1 if input is in percentage form, any other number if in frequency form
8.	NCAT	Vector containing the number of categories in each distribution
9.	NN	Vector containing the number of observations in each distribution--used only if category sizes are in percentage form
10.	IVAR	Input FORMAT under which data matrix is read
11.	FREQ, or PCNT	Data matrix. Which of these is read in depends on input variable 7.
12.	AVDEV	Nominal statistic 1
13.	MNDIF	Nominal statistic 2
14.	STDEV	Nominal statistic 3
15.	VARNC	Nominal statistic 4
16.	HREL	Nominal statistic 5
17.	MODVR	Nominal statistic 6
18.	RANVR	Nominal statistic 7
19.	NN	Number of observations in each distribution
20.	FREQ, or PCNT	Data matrix. Which of these is read in depends on input variable 7.
21.	MEAN	Vector of the means of the nominal statistics
22.	SIGMA	Vector of the standard deviations of the nominal statistics
23.	COVAR	Vector of the coefficients of variation of the nominal statistics

Table 4. Nomstat Output Variable Sequence and Sense Switch Control

Sequence	Output Variable	Sense Switch Control
1.	TITLE	
2.	NDIST, NPROP, NPCNT, NPUNCH, NSUM, NPCT	
3.	NCAT	
4.	NN	
5.	IVAR	
6.	FREQ, or PCNT	
7.	AVDEV, MNDIF, STDEV, VARNC, HREL, MODVR, RANVR	
8.	NN	
9.	FREQ, or PCNT	FREQ (If NPCNT=1), otherwise PCNT
10.	MEAN, SIGMA, COVAR	If NSUM $\neq$ 1

Sample Problem - A sample problem has been used which is based on the responses of a cross-section of the American public to a question on foreign policy asked on 29 surveys over a period of several years.

```
NOMSTAT TEST CASE - ALLEN R. WILCOX


CONTROL CARD AND VARIABLE FORMAT CARD INPUT

29  2  2  2  2  2
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
(5X,3F5.0)


DATA CARD INPUT
     629.  549.  127.
     623.  497.  153.
     691.  411.  109.
     816.  344.  103.
     960.  241.   86.
     755.  434.  106.
     849.  297.  114.
     834.  297.  127.
     796.  371.  121.
     624.  506.  128.
     724.  427.  151.
     681.  475.  125.
     639.  527.  116.
     613.  535.  136.
     657.  490.  125.
     734.  388.  137.
     704.  448.  131.
     722.  422.  126.
     782.  390.  104.
    1049.  166.   87.
     947.  249.   88.
     867.  333.  104.
    1009.  163.  103.
    1000.  164.  105.
    1035.  128.   89.
     916.  237.   81.
     798.  327.  112.
     904.  263.  121.
     860.  297.  125.
```

	AVDEV	MNDIF	STDEV	VARNC	HREL	MODVR	RANVR	N	PERCENTAGES		
1	64.60	61.53	64.20	87.18	85.81	77.70	20.19	1305.	48.	42.	10.
2	68.03	63.08	66.90	89.04	88.44	76.59	24.56	1273.	49.	39.	12.
3	63.50	51.94	58.37	82.67	82.25	64.41	15.77	1211.	57.	34.	9.
4	53.09	43.55	50.26	75.26	76.54	53.09	12.62	1263.	65.	27.	8.
5	38.11	32.09	37.24	60.61	64.92	38.11	8.96	1287.	75.	19.	7.
6	62.28	49.88	56.60	81.16	80.63	62.55	14.04	1295.	58.	34.	8.
7	48.93	41.67	47.40	72.34	75.01	48.93	13.43	1260.	67.	24.	9.
8	50.56	43.80	49.19	74.18	76.90	50.56	15.23	1258.	66.	24.	10.
9	57.30	47.59	54.11	78.94	79.93	57.30	15.20	1288.	62.	29.	9.
10	65.26	60.57	64.33	87.27	86.17	75.60	20.51	1258.	50.	40.	10.
11	66.59	55.99	61.88	85.47	85.73	66.59	20.86	1302.	56.	33.	12.
12	64.64	56.60	61.99	85.56	84.73	70.26	18.36	1281.	53.	37.	10.
13	63.57	59.20	62.80	85.16	84.64	75.23	18.15	1282.	50.	41.	9.
14	66.89	62.85	65.48	88.09	86.98	78.39	22.19	1284.	48.	42.	11.
15	64.74	58.18	62.95	86.27	85.26	72.52	19.03	1272.	52.	39.	10.
16	62.55	52.58	58.76	82.99	83.62	62.55	18.66	1259.	58.	31.	11.
17	65.32	55.34	61.25	84.98	84.62	67.69	18.61	1283.	55.	35.	10.
18	64.72	53.07	59.36	83.48	83.41	64.72	17.45	1270.	57.	33.	10.
19	58.07	46.87	53.80	78.65	78.89	58.07	13.30	1276.	61.	31.	8.
20	29.15	26.11	28.95	49.52	56.21	29.15	8.29	1302.	81.	13.	7.
21	39.37	33.10	38.40	62.06	66.11	39.37	9.29	1284.	74.	19.	7.
22	50.27	41.49	47.99	72.95	74.79	50.27	12.00	1304.	66.	26.	8.
23	31.29	28.94	31.17	52.63	59.29	31.29	10.21	1275.	79.	13.	8.
24	31.80	29.47	31.68	53.32	59.93	31.80	10.50	1269.	79.	13.	8.
25	26.00	24.44	25.95	45.16	52.65	26.00	8.60	1252.	83.	10.	7.
26	38.65	32.33	37.69	61.17	65.25	38.65	8.84	1234.	74.	19.	7.
27	53.23	44.54	50.87	75.86	77.55	53.23	14.04	1237.	65.	26.	9.
28	44.72	39.21	43.90	68.53	72.37	44.72	13.38	1288.	70.	20.	9.
29	49.38	42.67	48.06	73.02	75.88	49.38	14.53	1282.	67.	23.	10.
MEAN	53.16	46.16	51.09	74.64	76.36	55.68	15.06				
SIGMA	12.88	11.56	11.98	12.64	10.01	15.67	4.42				
COVAR	0.24	0.25	0.23	0.17	0.13	0.28	0.29				

APPENDIX B

Fortran List of the NOMSTAT Computer Program

```
**FTN,G,L,A,P,E.
      PROGRAM NOMSTAT
C
C         PURPOSE
C            COMPUTE SEVEN NOMINAL MEASURES OF VARIATION FOR USE WITH
C            UNIVARIATE DISTRIBUTIONS.
C
C         DESCRIPTION OF PARAMETERS
C            IVAR      - VARIABLE FORMAT VECTOR
C            NCAT      - NUMBER OF CATEGORIES VECTOR
C            FREQ      - FREQUENCY MATRIX
C            PCNT      - ORIGINAL PERCENTAGE MATRIX
C            PCT       - WORKING PERCENTAGE MATRIX
C            NN        - NUMBER OF OBSERVATIONS VECTOR
C            SUM       - STATISTICS SUMMATION VECTOR
C            SUMSQ     - STATISTICS SQUARED SUMMATION VECTOR
C            STAT      - STATISTICS VECTOR
C            AVER      - MEAN VECTOR
C            VAR       - VARIANCE VECTOR
C            SIG       - STANDARD DEVIATION VECTOR
C            COVAR     - COEFFICIENT OF VARIATION VECTOR
C            NCT       - NUMBER OF CATEGORIES
C            XCUM      - NUMBER OF OBSERVATIONS
C            XMEAN     - AVERAGE NUMBER OF OBSERVATIONS PER CATEGORY
C            AVDEV     - NUMERATOR TERM USED IN STAT(1)
C            XMN       - NUMERATOR TERM USED IN STAT(2)
C            XSQ       - NUMERATOR TERM USED IN STAT(3) AND STAT(4)
C            H         - NUMERATOR TERM (ABSOLUTE UNCERTAINTY) USED IN
C                        STAT(5)
C            HMAX      - DENOMINATOR TERM (MAXIMUM UNCERTAINTY) USED IN
C                        STAT(5)
C
      DIMENSION IVAR(100),NCAT(200),FREQ(50,200),PCNT(50,200),NN(200),
     1PCT(50,200),SUM(7),SUMSQ(7),AVER(7),VAR(7),SIG(7),COVAR(7),STAT(7)
      DIMENSION TITLE(20)
C
C         READ IN TITLE CARD
C
      READ 111, (TITLE(I),I=1,20)
  111 FORMAT(20A4)
      PRINT 112, (TITLE(I),I=1,20)
  112 FORMAT(1H1,20A4)
      PRINT 57
   57 FORMAT (1H1,43HCONTROL CARD AND VARIABLE FORMAT CARD INPUT)
C
C         READ IN CONTROL CARD WITH SIX PARAMETERS IN 6I3 FORMAT
C            1 NDIST   = NUMBER OF DISTRIBUTIONS
C            2 NPROP   = 1 IF STATISTICS ARE TO BE IN PROPORTION FORM,ANY
C                          OTHER NUMBER IF IN PERCENTAGE FORM
C            3 NPCNT   = 1 IF CATEGORY SIZES ARE TO BE PUT OUT IN
C                          FREQUENCY FORM, ANY OTHER NUMBER IF IN PERCENTAGE
C                          FORM
C            4 NPUNCH  = 1 IF STATISTICS ARE TO BE PUNCHED AS WELL AS
C                          PRINTED, ANY OTHER NUMBER IF NOT
C            5 NSUM    = 1 IF SUMMARY STATISTICS (MEANS, STANDARD
C                          DEVIATIONS, AND COEFFICIENTS OF VARIATION) ARE
C                          NOT TO BE CALCULATED, ANY OTHER NUMBER IF THEY
C                          ARE
C            6 NPCT    = 1 IF INPUT IS IN PERCENTAGE FORM, ANY OTHER
C                          NUMBER IF IN FREQUENCY FORM
C
```

```
      READ 15,NDIST,NPROP,NPCNT,NPUNCH,NSUM,NPCT
   15 FORMAT (6I3)
      PRINT 56, NDIST, NPROP, NPCNT, NPUNCH, NSUM, NPCT
   56 FORMAT (1H0,6I3)
C
C          READ IN FIVE CONTROL CARDS IN 40I2 FORMAT CONTAINING THE NUMBER
C          OF CATEGORIES IN EACH DISTRIBUTION - LIMIT OF 200
C
      L = 1
      M = 40
      DO 16 I = 1,5
      READ 17,(NCAT(K),K=L,M)
   17 FORMAT(40I2)
      PRINT 59, (NCAT(K),K=L,M)
   59 FORMAT (1H ,40I2)
      L = L+40
   16 M = M+40
C
C          IF PERCENTAGE INPUT IS CHOSEN, READ IN TEN CONTROL CARDS IN
C          20I4 FORMAT CONTAINING THE NUMBER OF OBSERVATIONS IN EACH
C          DISTRIBUTION - LIMIT OF 200
C
      IF(NPCT.EQ.1)44,45
   44 L = 1
      M = 20
      DO 46 I = 1,10
      READ 47,(NN(K),K=L,M)
   47 FORMAT(20I4)
      PRINT 62, (NN(K),K=L,M)
   62 FORMAT (1H ,20I4)
      L = L + 20
   46 M = M + 20
   45 CONTINUE
C
C          READ IN ONE VARIABLE FORMAT CARD
C
      READ 20, (IVAR(I),I=1,20)
   20 FORMAT(20A4)
      PRINT 63, (IVAR(I),I=1,20)
   63 FORMAT (1H ,20A4)
C
C          READ IN DATA ACCORDING TO VARIABLE FORMAT IN EITHER FREQUENCY
C          OR PERCENTAGE FORM
C
      PRINT 61
   61 FORMAT (1H1,15HDATA CARD INPUT)
      IF (NPCT .EQ. 1) GO TO 41
      DO 40 L = 1,NDIST
      NCT = NCAT(L)
      READ IVAR, (FREQ(I,L),I=1,NCT)
   40 PRINT IVAR, (FREQ(I,L),I=1,NCT)
      GO TO 43
   41 DO 42 L = 1,NDIST
      NCT = NCAT(L)
      READ IVAR, (PCNT(I,L),I=1,NCT)
   42 PRINT IVAR, (PCNT(I,L),I=1,NCT)
   43 CONTINUE
C
C          PRINT OUT HEADING
C
      IF(NPCNT.NE.1)GO TO 23
      PRINT 13
```

```
   13 FORMAT(1H1,5X,69H  AVDEV   MNDIF   STDEV   VARNC    HREL   MODVR   RANVR
     1     N     FREQUENCIES//)
      GO TO 24
   23 PRINT 25
   25 FORMAT(1H1,5X,69H  AVDEV   MNDIF   STDEV   VARNC    HREL   MODVR   RANVR
     1     N     PERCENTAGES//)
C
C          INITIALIZE TO 0 TWO VECTORS - SUM (SUMMATION OF THE NOMINAL
C          STATISTICS) AND SUMSQ (SUMMATION OF THE NOMINAL STATISTICS
C          SQUARED)
C
   24 DO 19 I = 1,7
   19 SUM(I) = SUMSQ(I) = 0
C
C          BEGIN MAJOR DO LOOP - CALCULATION OF STATISTICS
C
      DO 50 K = 1,NDIST
C
C          CALCULATE NUMBER OF OBSERVATIONS (XCUM) AND PERCENTAGES (PCNT)
C          IF FREQUENCIES ARE INPUT
C
      NCT = NCAT(K)
      IF(PCT.EQ.1)GO TO 49
      XCUM = 0
      DO 5 I = 1,NCT
    5 XCUM = XCUM + FREQ(I,K)
      GO TO 66
   49 XCUM = NN(K)
   66 DO 30 I = 1,NCT
   30 PCNT(I,K) = FREQ(I,K)/XCUM*100.
C
C          SUM NUMERATOR TERMS FOR STATISTICS 1-5
C
      XMEAN = XCUM/NCT
      AVDEV = XMN = XSQ = H = 0
      DO 6 M = 1, NCT
      XSQ = XSQ+(PCNT(M,K)*XCUM/100-XMEAN)*(PCNT(M,K)*XCUM/100-XMEAN)
      IF (PCNT(M,K) .EQ. 0) 8,65
   65 H = H+(-1*PCNT(M,K)/100)*(LOGF(PCNT(M,K)/100)/LOGF(2.))
    8 AVDEV = AVDEV+ABSF(PCNT(M,K)*XCUM/100-XMEAN)
      MP = M+1
      DO 6 MPP = MP,NCT
    6 XMN = XMN+ABSF(PCNT(M,K)*XCUM/100-PCNT(MPP,K)*XCUM/100)
C
C          CALCULATE STATISTICS 1-5
C
      NCT1 = NCT-1
      STAT(1) = 100-50*NCT*AVDEV/(XCUM*NCT1)
      STAT(2) = 100-100*XMN/(XCUM*NCT1)
      STAT(3) = 100-100*SQRTF(XSQ)/SQRTF(XCUM**2*NCT1/NCT)
      STAT(4) = 100-100*NCT*XSQ/(XCUM**2*NCT1)
      XCT = NCT
      HMAX = LOGF(XCT)/LOGF(2.)
      STAT(5) = H*100/HMAX
C
C          CALCULATE WORKING VECTOR TO PRESERVE ORDER OF PERCENTAGES FOR
C          OUTPUT PURPOSES
C
      DO 64 I = 1,NCT
   64 PCT(I,K) = PCNT(I,K)
C
C          FIND LARGEST AND SMALLEST CATEGORY PERCENTAGES FOR USE IN
```

```
C          STATISTICS 6-7
C
      DO 100 I = 1,NCT1
      IF (PCNT(I+1,K) .GE. PCNT(I,K)) 100, 101
  101 EXCH = PCNT(I+1,K)
      PCNT(I+1,K) = PCNT(I,K)
      PCNT(I,K) = EXCH
  100 CONTINUE
      NCT2 = NCT-2
      DO 102 I = 1,NCT2
      IF (PCNT(I+1,K) .LE. PCNT(I,K)) 102,103
  103 EXCH = PCNT(I+1,K)
      PCNT(I+1,K) = PCNT(I,K)
      PCNT(I,K) = EXCH
  102 CONTINUE
C
C          CALCULATE STATISTICS 6-7
C
      STAT(6) = NCT*(100-PCNT(NCT,K))/NCT1
      STAT(7) = 100*PCNT(NCT1,K)/PCNT(NCT,K)
C
C          CALCULATE PROPORTIONS IF REQUESTED
C
      IF(NPROP.NE.1)GO TO 82
      DO 80 I = 1,7
   80 STAT(I) = STAT(I)/100
C
C          SUM THE STATISTICS AND THE STATISTICS SQUARED
C
   82 DO 21 I = 1,7
      SUM(I) = SUM(I)+STAT(I)
   21 SUMSQ(I) = SUMSQ(I)+STAT(I)**2
C
C          PRINT MAIN OUTPUT IN EITHER FREQUENCY OR PERCENTAGE FORM
C
      IF(NPCNT.NE.1)GO TO 71
      IF(NPCT.EQ.1)51,52
   51 DO 53 I = 1,NCT
   53 FREQ(I,K) = PCT(I,K)*XCUM/100
   52 PRINT 73,K,(STAT(I),I=1,7),XCUM,(FREQ(I,K),I=1,NCT)
   73 FORMAT(I6,7F7.2,F6.0,2X,10F5.0,4(/1H+,62X,10F5.0))
      GO TO 72
   71 PRINT 12,K,(STAT(I),I=1,7),XCUM,(PCT(I,K),I=1,NCT)
   12 FORMAT(I6,7F7.2,F6.0,2X,10F5.0,4(/1H+,62X,10F5.0))
C
C          PUNCH STATISTICS IF REQUESTED
C
   72 IF(NPUNCH.NE.1) GO TO 50
      PUNCH 14,K,(STAT(I),I=1,7)
   14 FORMAT(I4,7(F7.2))
   50 CONTINUE
C
C          CALCULATE AND PRINT MEANS (AVER), STANDARD DEVIATIONS (SIG),
C          AND COEFFICIENTS OF VARIATION (COVAR) IF REQUESTED
C
      IF(NSUM.EQ.1)GO TO 110
      DO 22 I = 1,7
      AVER(I) = SUM(I)/NDIST
      VAR(I) = (SUMSQ(I)/NDIST) - AVER(I)**2
      SIG(I) =SQRTF(VAR(I))
   22 COVAR(I) = SIG(I)/AVER(I)
      PRINT 9,(AVER(I),I=1,7),(SIG(I),I=1,7),(COVAR(I),I=1,7)
    9 FORMAT(1H0,5H MEAN,7F7.2/X,5HSIGMA,7F7.2/X,5HCOVAR,7F7.2)
  110 CONTINUE
      END NOMSTAT
```

ACKNOWLEDGEMENTS

The author gratefully acknowledges the careful perusal of and helpful commentary on this paper by M. A. Kastenbaum and W. E. Lever of the Statistics Section and by Pamela Wilcox of the Civil Defense Research Project.

THIS PAGE

WAS INTENTIONALLY

LEFT BLANK

INTERNAL DISTRIBUTION

1. T. H. Atkinson
2. D. B. Bobrow
3. S. B. Bobrow
4. M. C. Bell
5. W. J. Boegly, Jr.
6. J. C. Bresee
7. F. R. Bruce
8. C. V. Chester
9. G. A. Cristy
10. N. E. Cutler
11. L. Dresner
12. R. C. Durfee
13. M. Feliciano
14. D. A. Gardiner
15. C. M. Haaland
16. A. S. Householder
17. M. A. Kastenbaum
18. C. H. Kearny
19. W. E. Lever
20. J. S. Levey
21. J. L. Liverman
22. H. G. MacPherson
23. W. M. McMullen
24. C. Nader
25. D. L. Narver, Jr.
26. D. B. Nelson
27. A. F. Shinn
28. J. W. Strohecker
29. R. A. Uher
30. V. R. Uppuluri
31. A. M. Weinberg
32. E. P. Wigner
33-57. A. R. Wilcox
58. P. B. Wilcox
59-60. Central Research Library
61. ORNL-Y-12 Technical Library, Document Reference Section
62-66. Laboratory Records Department
67. Laboratory Records, ORNL RC
68. ORNL Patent Office
69. Laboratory and University Division, ORO
70-84. Division of Technical Information Extension
85-134. Civil Defense Project Library

EXTERNAL DISTRIBUTION

135. Chadwick F. Alger, Department of Political Science, Northwestern University, Evanston, Illinois
136. Lee F. Anderson, Department of Political Science, Northwestern University, Evanston, Illinois
137. Douglas Bwy, Department of Political Science, Western Reserve University, Cleveland, Ohio
138-147. L. J. Deal, Civil Effects Branch, Division of Biology and Medicine, U.S. Atomic Energy Commission, Washington, D.C.
148. Harold Guetzkow, Department of Political Science, Northwestern University, Evanston, Illinois
149. Kenneth Janda, Department of Political Science, Northwestern University, Evanston, Illinois
150. David W. Minar, Department of Political Science, Northwestern University, Evanston, Illinois
151-160. Col. Ralph Pennington, Advanced Research Projects Agency, Department of Defense, Washington, D.C.
161. Rudolph J. Rummel, Department of Political Science, University of Hawaii, Honolulu, Hawaii
162. William Scott, Department of Psychology, University of Colorado, Boulder, Colorado

163-172. W. E. Strope, Office of Civil Defense, Department of the Army, Washington, D.C.

173-177. Raymond Tanter, Advanced Research Projects Agency, The Pentagon, Washington, D.C.