

A GLANCE AT STATISTICAL PROCEDURE

Thomas W. McKern

To the student who depends upon printed source material for further scientific research, the problem of accurate statistical methodology is vital. The statistical methods of many authors rarely include the proper testing of face value significance; statistical deficiencies easily can be demonstrated throughout the literature of fields which rely on quantitative research methods.

An interest in this subject was acquired during the preparation of a Master's thesis concerning differential human growth when an analysis of the statistical methods applied by other authors to similar data was attempted. It was surprising to find that most of the work surveyed had not been tested further than by a compilation of standard deviations and coefficients of variation. In many cases, conclusions were formulated on the sole basis of mean measurements.

It must be pointed out that standard deviation, computed from a single sample, is itself a variable and subject to error since it depends upon the size of the mean or median in each case, and upon the particular unit used. For this reason the standard deviations of two frequency distributions are usually not directly comparable. Standard deviations may, therefore, be made comparable by expressing them as percentages of their means. This percentage is called the coefficient of variation and is used as a measure of the representativeness of an average; it may be applied in direct comparison of the variability and dispersion of variates of different absolute mean sizes. However, the mere statement of the value of "V" (coefficient of variation) computed from a single sample is not necessarily conclusive evidence of the degree of association, because it too must depend on the sample and its expressed mean.

It is possible to check the significance of variability of material that is represented by the usual statistical tables by means of the "t" test. If the standard deviation is lacking, the sigma of a similar study may be substituted. The quantity "t" may be used for testing the significance of the difference between two mean values. If N_1 and N_2 represent the number of items in each of the two samples being compared, M_1 and M_2 are the respective sample means, and σ_1 and σ_2 are the standard deviations of the two samples, then

$$t = \frac{(M_1 - M_2) \sqrt{\frac{N_1 \cdot N_2}{N_1 + N_2}}}{\sqrt{\frac{N_1 \sigma_1^2 + N_2 \sigma_2^2}{N_1 + N_2 - 2}}}$$

The above formula corresponds to that used by Simpson and Roe for the comparison of the means of two small samples.(1) It has sometimes been called the coefficient of admodality. The "t" test, by including the three major variables of any statistical experiment, i.e., the sample, the mean and the standard deviation, measures the probability that a difference observed between two means could have been equaled or exceeded by chance sampling of one population. The reader is directed to both Simpson and Roe(2) and Fisher(3) since the "t" test as used in this paper is composed of elements from both works.

* * * * *

The relationship between man and his environment has been a subject of considerable speculation, argument, and scientific research. In recent years a widespread interest in quantitative statement of facts has started a trend toward the statistical analysis of the problems of the social scientist. As a result of this interest, there has been extensive publication in the name of statistical analysis, while few attempts have been made to examine the accuracy of the broad conclusions derived from these studies.

It has been concluded by most contemporary authors that environmental factors exert a certain amount of influence upon human physique. The amount of influence of environmental factors has never been determined, for heredity and environment cannot be divided arbitrarily into specific percentages.

The several studies which have been subjected to the "t" test in this paper were selected from the literature of the field. In the absence of any original computations of variance, the standard deviation of my own urban series(4) was used.

As early as 1877, H. P. Bowditch(5) made a study of the growth of children in Massachusetts. In his report he included material on the effect of occupation on the height of children. His table(6) has been duplicated in part, with the addition of a "t" test column:

COMPARATIVE MEAN HEIGHTS

<u>Age</u>	<u>No.</u>	<u>Employed in Factories</u>	<u>No.</u>	<u>Not Employed in Factories</u>	<u>"t" Tests*</u>
9	17	122.2 cm.	41	123.3 cm.	.06
10	48	127.0 "	28	128.6 "	1.19
11	53	130.2 "	25	129.6 "	.44
12	42	135.5 "	20	134.6 "	.65
13	45	138.3 "	22	139.6 "	.52
14	51	143.7 "	16	144.0 "	.18

* 1.96 is significant (7)

It was from data such as these that Bowditch claimed a marked association between the occupational status of the parents and the height of the children. The children of professional parents were found to have the highest average height while the children of the unskilled laborers had the lowest.

The inadequacy of the above values as a basis for such a conclusion is well illustrated by the results of the test for significance applied to the variations of the means for each age group. The fact that there are differences between the means of the two groups, and that these differences show a slight advantage in height for those not employed in factories, is not valid evidence with which to make conclusive statements. The differences must have a significant value.

Roberts(8) investigated the comparative heights of the high and the low social classes in England. His sample included adults from ages twenty to thirty. The following table(9) illustrates the type of material used as a basis for his conclusions:

<u>Occupational Sub-Groups</u>	<u>Mean Height</u>
Professional	69.0 in.
Commercial	68.0 "
Laborers	65.5 "

Roberts concluded:(10)

On the basis of this data, it is safe to say that there is a tangible correlation between high stature and high social class, on the one hand; between a low stature and low social class, on the other hand...

This is a further demonstration of the complete reliance upon the mean values of frequencies as an adequate basis for general con-

clusions. Testing the mean heights of the Professional and Laborer sub-groups, I obtained "t" of 1.90, which was just short of being significant.

In 1905, A. Niceforo(11) completed a study of the poor classes of Paris in which he included comparative material for the upper and lower socio-economic classes. In his summary, Niceforo concluded:(12)

The upper classes are taller, have a greater weight, greater cranial capacity, greater handsomeness and less serious and less numerous anomalies and defects than the lower classes.

Below is an example of the data from which his conclusions were formulated:(13)

<u>Social Groups</u>	<u>No.</u>	<u>Height</u>
Students	33	170.9 cm.
Wage Earners	33	168.0 "

Applying the "t" test to the mean heights, I found the "t" to equal 1.58 which is not significant at a 5% level. The 5% level is rather widely used at present and is adopted in the present study rather than the 1% level, which is preferred when there is need to be more conservative. All deviations greater than those with probabilities of 0.05 are regarded as real, or statistically significant. Admittedly, this method of testing the credibility of other studies is crude; however, it is the only means of testing such data where original research work is not obtainable.

Saller(14) published an anthropological text book which included statistical material relating to a correlation by height measurements between students and workers in several different countries. As an example, a part of one of his tables follows:(15)

<u>Stature</u>	<u>Italy</u>	<u>England</u>
Students	166.9cm	172.4cm
Workers	164.4 "	169.8 "

In testing Saller's conclusions, namely, that the students in both countries demonstrated a superior height to the workers, it was found that for Italians, "t" equaled 1.56; for English, "t" equaled 1.64. The variations of both means are insignificant.

In 1939, E. A. Hooton(16) published The American Criminal in which he attempted to correlate occupational status with the physical

measurements of his criminal types. Below is reproduced one of his tables which was tested in the same manner as the above studies.(17)

HEIGHT BY OCCUPATION GROUPING

	<u>No.</u>	<u>Range</u>	<u>Mean</u>	<u>S.D.</u>	<u>V.</u>
Extractive	1,301	146-190cm	172.41cm	6.18	3.58
Laborer	683	149-190 "	171.33 "	6.48	3.78
Factory	714	146-196 "	172.02 "	6.42	3.73
Transportation	302	152-193 "	172.29 "	6.42	3.73
Skilled trades	374	158-193 "	172.02 "	5.97	3.47
Trade	210	149-190 "	172.20 "	6.39	3.71
Public service	27	155-181 "	171.99 "	5.28	3.07
Semi-professional	69	149-190 "	170.73 "	7.50	4.39
Professional	27	155-187 "	171.06 "	6.60	3.84
Personal service	281	146-196 "	171.03 "	6.90	4.03
Clerical	117	149-190 "	170.97 "	7.38	4.32
	<u>4,102</u>	146-196cm	172.20cm	6.48	3.76

Dr. Hooton concludes:(18)

On the whole, occupational groups seem to be fairly strongly differentiated in stature since presumably significant deviations from the mean of the total series occur in 27.27 per cent.

From the above table, Hooton selected certain sub-groups, for comparative purposes, as deviating significantly. For example, the occupational sub-groups Clerical and Laborer were chosen to represent social contrast. The means of these two groups, when tested, yielded a "t" of .53 at a 5% level which is far from significant.

Examples such as this were not typical of Hooton's material. The point demonstrated is that the recorded results of other authors should be checked before their data are used as accurate bases for further contributions.

In 1943, G. Wartenweiler(19) published a group of tables correlating occupational status with many physical measurements. The conclusion drawn from these studies was the physical superiority of the upper socio-economic classes over the laboring classes. Below is a partial duplication of one of his tables with the present writer's own figures of significance in the last column:(20)

	<u>No.</u>	<u>Intellectual</u>	<u>No.</u>	<u>Labor</u>	<u>"t" test</u>
Stature	20	174.6cm	39	167.6 cm	4.54*
Lower leg	20	39.6 "	39	37.7 "	.14
Minimum frontal	20	10.4 "	39	10.65 "	1.70
Cephalic index	13	79.2 "	39	80.5 "	.08

* significant

Again, with the exception of mean stature, the mean values of the measurements show an insignificant variation despite the tendency for the upper classes to reveal slightly greater metric proportions.

The above studies are only a few that have been selected at random from the many works which are concerned with the relations of the organism to its environment. However, the problems which they present to the student are relevant to scientific research. They illustrate the type of material which is used by many as sufficient evidence for further studies along similar lines. The literature of the field of differential growth contains many statistical inadequacies which render many of the accompanying conclusions debatable.

It is not my intention to discard all studies that demonstrate small differences between the mean values of two samples. These differences, although slight, may be significant; however, they require appropriate analysis. This can be demonstrated easily by looking again at the table taken from Wartenweiler's study. The difference in stature between the socio-economic groups is not much greater than that of the lower leg measurements of the two groups and yet the former is highly significant when tested. It is realized that tests for significance are only verifications of observable differences, and that conclusions drawn from small samples may be valid manifestations of the problem under investigation. However, too many investigators have depended upon the raw results of such samples as adequate bases for their conclusions and have not attempted to test these results in any manner.

In the field of social scientific research, statistical recording and manipulation is playing a more and more decisive role in reducing our knowledge to a precise and objective form. If the methods which are initial in producing this emphasis are not adequately applied, the resulting inferences are useless as well as harmful.

FOOTNOTES

- (1) Simpson and Roe, 1939, p. 211.
- (2) Ibid., p. 211.
- (3) Fisher, 1941, p. 137.
- (4) McKern, 1948, p. 48.
- (5) Bowditch, 1877.
- (6) Ibid., p. 289.
- (7) Significance is derived by checking "t" test values in Shepard's Table (Fisher, 1941, Appendix) at the 5% level.
- (8) Roberts, 1892.
- (9) Ibid., p. 222.
- (10) Ibid., p. 222.
- (11) Niceforo, 1905.
- (12) Ibid., p. 250.
- (13) Ibid., p. 246.
- (14) Saller, 1930.
- (15) Ibid., p. 244.
- (16) Hooton, 1939.
- (17) Ibid., Appendix, Table VIII-13.
- (18) Ibid., p. 153.
- (19) Wartenweiler, 1943.
- (20) Ibid., p. 599.

BIBLIOGRAPHY

- Bowditch, H. P.
 1877 "Growth of Children", in 8th Annual Report of the State Board of Health, pp. 273-323. Boston, Mass.
- Fisher, R. A.
 1941 Statistical Methods for Research Workers. Oliver and Boyd, Edinburgh and London.
- Hooton, E. A.
 1939 The American Criminal. Harvard University Press, Cambridge, Mass.
- McKern, T. W.
 1948 The Problem of Analyzing the Effect of Environmental Factors on Human Physique. (Unpublished Thesis), University of Wisconsin.
- Niceforo, A.
 1905 Les Classes Pouvres. V. Giard and E. Brière, Paris.
- Roberts, C.
 1892 "On the Uses and Limits of Anthropometry", in Bulletin de l'Institut International de Statistique, Vol. VI. Rome.
- Saller, K.
 1930 Leitfaden der Anthropologie. J. Springer, Berlin.
- Simpson, G. G., and A. Roe.
 1939 Quantitative Zoology. McGraw-Hill, New York, N. Y.
- Wartenweiler, G.
 1943 "Wachstum und Formgestaltung des Menschlichen Fusses", in Ausgabe dem Anthropologischen Institute der Universität Zurich. Zurich.