INTERPRETATION OF MENDELIAN CLASS FREQUENCIES[1]

By WALTER A. HENDRICKS, formerly *assistant poultry husbandman*, JOSEPH P. QUINN, *poultry geneticist*, and ALBERT B. GODFREY, *assistant poultry husbandman, Animal Husbandry Division, Bureau of Animal Industry, United States Department of Agriculture*

INTRODUCTION

The application of the chi-square test to Mendelian class frequencies constitutes a familiar illustration of the general problem of testing a hypothesis by means of observed data. Since the test is almost universally employed by geneticists, it appears desirable to present a few principles that may serve as a guide to its interpretation.

The investigator should have no illusions regarding the ability of the test to prove or disprove the hypothesis under consideration. No such test exists. Neither should he harbor the mistaken idea that the test measures the probable truth of the hypothesis. Attempts to measure the probable truth of a hypothesis have always become involved in the controversial subject of inverse probability and have not been generally accepted by mathematical statisticans. The philosophy underlying the chi-square test is foreign to such a concept and avoids the difficulties to which it leads, but only at the expense of providing the investigator with a test that falls short of what is actually desired. The test merely shows whether or not the observed data are likely to be obtained under conditions of random sampling if the hypothesis under consideration is true. The investigator's conclusions in regard to the truth or falsity of the hypothesis can be nothing more than inferences based on the results given by the test and such other information relative to his particular problem as may be available.

THE CONCEPT OF RANDOM VARIATION

It is regrettable that the small amount of information which the chi-square test is capable of contributing to problems involving the testing of hypotheses has in the past been still further restricted by the failure of some investigators to comprehend the concept of random variation and its implications in experimental work. At least a part of the blame for this failure must be borne by the statistician, for the older chi-square tables are not adequate for an efficient application of the chi-square test. These tables tend to create the impression that only values of chi square larger than one would expect to arise by chance should be regarded as lying outside the normal range of variation.

The reason for this impression is a subject for speculation. One can hardly suppose that such statisticians as Pearson (*7*)[2] and Elderton (*1*) had any mistaken ideas as to what constitutes random variation, although it appears unlikely that any particular importance would

[1] Received for publication October 25, 1938.
[2] Italic numbers in parentheses refer to Literature Cited, p. 759.

have been attached to a small value of chi square at a time when the test was used principally to measure the goodness of fit of frequency curves. If any objection to too good a fit occurred to them, they appear to have made no mention of it. Fisher (*3*) was fully aware of the implications of unusually good agreement between observation and theory, but proposed no better criterion for determining the normal range of variation of chi square than was provided by Elderton's table. This circumstance may have been due to the lack of a suitable criterion for establishing limits of variability in the case of skew frequency distributions.

In order to establish a criterion that will show whether or not an observed value of chi square is within the normal range of variation, it is necessary to arrive at an agreement as to what constitutes the normal range of variation. The senior author has found that the shortest range of values which will include 95 percent of the observations may be regarded as the normal range of variation, and has computed the following chi-square table (table 1):[3]

TABLE 1.—*Critical values of chi square for 5-percent level of significance*

Degrees of freedom	Value		Degrees of freedom	Value		Degrees of freedom	Value	
	Lower	Upper		Lower	Upper		Lower	Upper
1	0.000	3.841	11	2.954	20.305	21	9.267	33.921
2	.000	5.991	12	3.516	21.729	22	9.958	35.227
3	.003	7.816	13	4.099	23.135	23	10.656	36.526
4	.085	9.530	14	4.700	24.525	24	11.361	37.818
5	.298	11.191	15	5.318	25.900	25	12.073	39.103
6	.607	12.802	16	5.948	27.263	26	12.791	40.383
7	.990	14.369	17	6.591	28.615	27	13.514	41.658
8	1.425	15.897	18	7.245	29.955	28	14.243	42.927
9	1.903	17.393	19	7.910	31.285	29	14.977	44.192
10	2.414	18.861	20	8.584	32.607	30	15.716	45.451

The principle on which table 1 is based is theoretically sound and lends itself well to the problem of establishing limits of variability for a large variety of statistics that are distributed according to skew frequency curves. It so happens that the ordinates of the chi-square distribution at the two extremities of this range are equal when the number of degrees of freedom is greater than 2. This property, which can be demonstrated mathematically, is the necessary and sufficient condition that a particular value of chi square regarded as lying outside the normal range of variation cannot have a probability of occurrence equal to, or greater than, the probability of occurrence of some value on the other side of the mode that would be accepted as being within the normal range of variation. For 1 and 2 degrees of freedom, the lower bound of admissible values of chi square lies at zero, which agrees with the known fact that observed values in the neighborhood of zero are expected to occur more frequently than any other. For an extremely large number of degrees of freedom, the chi-square distribution approaches a symmetrical form resembling the normal curve of errors. In such a case, the normal range of variation, as defined in this paper, would exclude approximately 2.5 percent of the values at

[3] Acknowledgment is made for the assistance of Kate W. Robey, of the Bureau of Animal Industry, in the construction of this table.

each extremity of the distribution, a phenomenon that has a well-known parallel in the usual treatment of the normal curve. However, this condition is not realized until the number of degrees of freedom becomes very large. For 30 degrees of freedom, the criterion presented in this paper excludes only 1.5 percent of the values of chi square at the lower extremity of the curve and 3.5 percent of the values at the upper extremity.

It is evident that the table leads to results that are perfectly consistent with the nature of the frequency distributions involved for various numbers of degrees of freedom. A similar method of approach has recently been used by Fertig and Proehl (*2*) in a consideration of Neyman and Pearson's likelihood function. In fact, the table that Fertig and Proehl (*2*) have presented for testing an observed standard deviation may be used to test an observed value of chi square in accordance with the theory embodied in Hendricks and Robey's table. For an observed value of chi square, one may compute Fertig and Proehl's k by means of the formula

$$k=0.434294\,\frac{\chi^2}{n-2}-\log_{10}\frac{\chi^2}{n-2},$$

in which n is the number of degrees of freedom involved, and determine the probability associated with that value of k by deducting 2 units from the correct number of degrees of freedom before entering the table.

Table 1 and the table presented by Fertig and Proehl (*2*) provide more satisfactory criteria than have previously been available for deciding whether or not an observed value of chi square, as stated by Pearson (*7*), "is such that it can be reasonably supposed to have arisen from random sampling."

An investigator obtaining a value outside the normal range of variation should reconsider his hypothesis and subject his data to further scrutiny. The conclusions to be drawn in the case of a value larger than one would expect to arise by chance, are familiar to most geneticists, but the interpretation to be placed on a value smaller than one would expect may not be so apparent. Such a value does not necessarily imply either that the hypothesis is wrong or that there is something wrong with the experimental technique, for unusual sets of data occasionally arise by chance. However, any investigator who consistently obtains values of chi square smaller than would be expected to arise by chance should infer, in the case of Mendelian class frequencies, that he is working with data that were consciously or unconsciously selected to agree with the hypothesis under consideration.

DETECTION OF SELECTED DATA

In some cases, unusually close agreement between observation and theory occurs because of a natural tendency on the part of the investigator to stop his experiment when the data seem to agree with his hypothesis. The fact that such a procedure results in selected data may not have occurred to some geneticists. In this connection, it may be pointed out that, according to the law of chance, an investigator may obtain data that will satisfy any one of several different hypotheses regarding the proportion of observations to be expected in each of the various classes. If he stops his experiment when the results seem to agree with his chosen hypothesis, his experiment cannot be regarded as a fair test of the hypothesis.

Data derived from such an experiment generally yield unusually small values of chi square because of the tendency of investigators to underestimate the amount of variability that can be accepted as arising by chance. This fact is illustrated in some data reported by Quinn and Godfrey (*8*). These data, considered en masse, consist of 44 classes of observations in which the method of selection just referred to is known to have been used. The 44 classes yield 22 degrees of freedom for which chi square has the value 6.398. By referring to table 1 it will be observed that the normal range of variation of chi square for 22 degrees of freedom is 9.958 to 35.227. The small value of chi square obtained from the data, therefore, reflects the artificial agreement between observation and theory that was imposed by the selection of the data. However, in this case, supplementary evidence is available to confirm the conclusions reported and there is no question of their validity.

The older literature in the field of genetics contains reports of data that seem to have been obtained under similar conditions. For example, Lippincott (*6*) reported three sets of data obtained from backcross matings in a study of plumage-color inheritance in chickens. From the use of only the totals in his tables, the three sets of data yield 7 degrees of freedom for which chi square has the value 0.550. The normal range of variation of chi square for 7 degrees of freedom is 0.990 to 14.369. In this illustration, there is also no question of the validity of the conclusions reported. However, such selected data are undesirable and may at times retard the discovery of genetic relationships.

A critique of Mendel's classic work recently reported by Fisher (*4*) shows how selected data may appear to substantiate an incorrect conclusion. Fisher (*4*) found unusually close agreement between observation and theory in Mendel's data. For 84 degrees of freedom, chi square had the value 41.6. In order to determine the normal range of variation, one may make use of the approximation to the chi-square distribution for large numbers of degrees of freedom suggested by Fisher (*3*). The range is 60 to 111. Therefore, the observed value, 41.6, is smaller than one would expect under conditions of random sampling and indicates that the data are a selected set. In one phase of his work, Mendel's data agree closely with the 2 : 1 ratio which he expected to obtain, but differ significantly from the correct ratio.

One other example may be cited because it illustrates another method of detecting selected data. Smith (*9*) reported data on the inheritance of annual versus biennial growth habit in sweetclover plants and concluded that the 2 factors segregated in a 3:1 ratio in the F_2 generation. The totals of his 2 classes of observations, obtained from 91 families, yield 1 degree of freedom for which chi square has the value 2.592. This result is within the normal range of variation, 0.000 to 3.841, and appears to substantiate his hypothesis. However, when the chi square test is applied to the totals of 9 subgroups of his observations, 8 containing the data for 10 families each and 1 containing the data for the remaining 11 families, chi square has the value 42.345. The normal range of variation of chi square for 9 degrees of freedom is 1.903 to 17.393. This result indicates that the data are more variable than would be expected if the hypothesis were correct and that the agreement of the totals with the hypothesis, as reported by Smith (*9*), is spurious. The same conclusion in regard to this set of data has been reported by Kirk and Immer (*5*), who made a study of the results reported for individual families.

REPORTING DATA ON POULTRY GENETICS

Poultry geneticists, as a group, apparently do not report data that agree too closely with the hypotheses tested. Warren[4] has summarized a large amount of data obtained by various investigators studying autosomal factors in the fowl.[5] Omitting data for factors involving linkage, the summary yields 213 degrees of freedom for which chi square has the value 284.4. The normal range of variation is 174 to 255. The observed value of chi square is thus larger than one would expect, but a careful examination of the data indicates that this is due to the presence of a few highly discordant observations rather than to any systematic bias in the data.

An investigator need not hesitate to report data in which the agreement between observation and hypothesis differs significantly from what is expected under conditions of random sampling if he is convinced that the discrepancy is due to chance. To avoid criticism, an investigator may be tempted to report only such results as appear to be favorable to his hypotheses and to withhold results that exhibit either less or more variability than is expected. Such a procedure, if practiced consistently, will inevitably lead to an accumulation of fictitious data that accomplishes little more than to reflect the personal opinions of the investigators who obtained it.

SUMMARY

The principles involved in the interpretation of Mendelian class frequencies are discussed. The application of the chi-square test to such data is considered as a means of detecting data which do not agree with what is expected under conditions of random sampling. A new chi-square table, which appears to be more useful for this purpose than the tables in use at the present time, is presented.

LITERATURE CITED

(1) Elderton, W. Palin.
1902. TABLES FOR TESTING THE GOODNESS OF FIT OF THEORY TO OBSERVATION. Biometrika 1: [155]–163.

(2) Fertig, John W., with the assistance of Proehl, Elizabeth A.
1937. A TEST OF A SAMPLE VARIANCE BASED ON BOTH TAIL ENDS OF THE DISTRIBUTION. Ann. Math. Statis. 8: 193–205, illus.

(3) Fisher, R. A.
1934. STATISTICAL METHODS FOR RESEARCH WORKERS. Ed. 5. 319 pp., illus. Edinburgh and London.

(4) ———
1936. HAS MENDEL'S WORK BEEN REDISCOVERED? Ann. Sci. 1: [115]–137.

[4] Warren, D. C. TESTS FOR LINKAGE AMONG AUTOSOMAL FACTORS IN THE FOWL. Kansas State College. 1937. (Mimeographed).

[5] The authors are indebted to D. C. Warren and F. B. Hutt for permission to use the portions of their data, from Dr. Warren's summary, which have not yet been published.

(5) KIRK, L. E., and IMMER, F. R.
1928. APPLICATION OF GOODNESS OF FIT TESTS TO MENDELIAN CLASS FREQUENCIES. Sci. Agr. 8: 745–750.

(6) LIPPINCOTT, W. A.
1923. THE HEREDITARY RELATION OF DOMINANT WHITE AND BLUE IN CHICKENS. Poultry Sci. 2: 141–145.

(7) PEARSON, KARL.
1900. ON THE CRITERION THAT A GIVEN SYSTEM OF DEVIATIONS FROM THE PROBABLE IN THE CASE OF A CORRELATED SYSTEM OF VARIABLES IS SUCH THAT IT CAN BE REASONABLY SUPPOSED TO HAVE ARISEN FROM RANDOM SAMPLING. Phil. Mag. (5) 50: 157–175.

(8) QUINN, J. P., and GODFREY, A. B.
1937. COLOR-PRODUCING GENES IN WHITE SILKIE and WHITE ROSE COMB BANTAM. Poultry Sci. 16: 340–344.

(9) SMITH, HUGH BURNICE.
1927. ANNUAL VERSUS BIENNIAL GROWTH HABIT AND ITS INHERITANCE IN MELILOTUS ALBA. Amer. Jour. Bot. 14: 129–146, illus.