

THE STATISTICAL INVESTIGATION OF SCHOOL GRADES.

By W. L. EIKENBERRY.

University High School, University of Chicago.

In the study of error and chance it has been shown that errors are amenable to a mathematical law. For instance, if one were to toss a penny at a crack a great many times it might be confidently predicted that he would not hit the crack every time. It might also be predicted that the coin would alight sometimes on one side of the crack and sometimes on the other. Indeed both theory and experiment agree in the conclusion that if there were no disturbing factors, the distribution of the casts would be equal on the two sides of the crack and the number of occurrences at any particular distance would bear an inverse relation to the distance. That is, the coin would fall most frequently near the crack and proportionally less frequently at greater distances. If the frequency of the fall of the coin at different distances to right and left be ascertained and these frequencies be represented graphically by plotting the curve it is found that the resultant curve is a symmetrical one. It reaches its highest point on the axis which represents the frequency upon the crack, and it descends equally to right and left. This curve is found to correspond very closely to the curve derived from the binomial equation (see dotted line, Fig. 3). Also, as the number of casts becomes greater the correspondence becomes closer and we are compelled to believe that for a very large number of casts the correspondence would be practically a coincidence. This curve is called the curve of the probability of error, or sometimes in biological statistics, the normal curve.

Study of the variation of various single characters in plants and animals has shown that these variations are in general distributed about a central mode in a fashion quite in harmony with the distribution of errors. There are of course many deviations from this standard distribution and each such deviation becomes a problem in itself. Both types of distribution are conveniently illustrated in the curves showing the variation of number of grains per head in two types of wheat, published in *SCHOOL SCIENCE AND MATHEMATICS*, Vol. II, No. 1, page 35, 1911. Curve "A" corresponds very closely with the binomial curve and is such a figure as one would ordinarily expect to find; curve "B" is very aberrant and suggests that it may be in fact composed of several regular but overlapping curves. The symmetrical curve,

"A". shows that in this variety of wheat there are a few small heads, a great many of medium size, and a few very large ones. That is, most of the individuals are mediocre and a few are exceptional in either direction. This law holds generally throughout the plant and animal kingdoms. With respect to any particular character it is found that most of the individuals of a species fall within mediocrity,—the central part of the curve—with a smaller number of exceptional individuals on either side of this central mode, and that the curve of distribution approximates the curve of probability of error. The exceptions are amenable to explanations which are consistent with the law.

If it be admitted that the probability curve is the normal variation curve in plants and animals, it follows as a corollary that variation in the physical characteristics of man should have a similar distribution. Data bearing upon this question have been collected and studied by many. To instance only one, Karl Pearson made a study of the stature of English women founded upon 1025 individuals. The resultant curve approximates the theoretical curve very closely indeed. (Fig. 1.) The correspondence is of as high degree as theoretical considerations would lead us to expect. Other studies of a similar character serve only to strengthen the conviction that in this respect man is amenable to the general biological law.

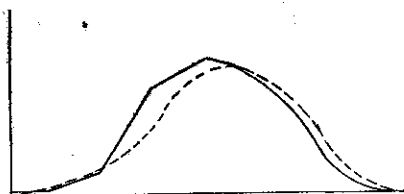


FIG. 1.

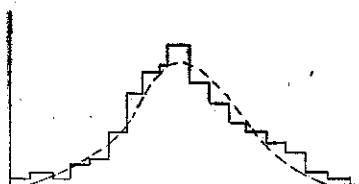


FIG. 2.

The notion that the law of biological variation may be properly applied to the study of mental and moral characters is almost inevitable. In a study of this sort there is of course a primary difficulty resulting from the imperfection of our methods of measuring such characters. Attempts at such measurement have however been made with, it is believed, a fair degree of accuracy, and considerable of the newer work in psychology is directed towards the accumulation of such measurements and the perfection of the technique needed in making them. Thorndike has

reported the results of a test of the efficiency of certain pupils in rapidity and accuracy of perception. The resultant curve again closely approximates the probability curve though as might have been expected not so closely as in the case of curves representing the more easily measurable physical characters. (Fig. 2.) The same thing is illustrated with a fair degree of correspondence in a study of memory for related words. Other similar studies have been made in sufficient numbers to lend great probability to the notion that mental characters do not differ in respect to the laws of their variations from physical characters. It is commonly assumed in all such studies that the normal distribution is that which corresponds to the probability curve.

School grades constitute a great body of data bearing upon the matter of individual mental abilities—possibly the greatest body of such data now available. Each grade represents an effort to measure the results of mental effort. In two particulars at least these grades are not ideal material. In the first place they do not pretend to measure ability, primarily, but achievement and it is quite possible that in a certain fraction of the cases the results achieved are not in close correspondence with the intrinsic ability of the individual; they may be due to unusual effort or to unusual carelessness. Doubtless both of these cases exist in any body of school records in considerable numbers but it is also probable that they occur in nearly equal numbers on both sides of the equation and practically counterbalance each other. At any rate, any investigation of school matters is quite as likely to be concerned with results as with abilities. The experience of investigators is that this factor does not in practice seriously modify the results.

A second objection lies in the possible unequal or wholly incommensurate standards of different teachers. If data are collected sufficiently widely to include a great many teachers this divergence also will disappear. Very frequently it is not sought to eliminate this deviation due to the individuality of the teacher; it is desired on the contrary to use the facts for the purpose of evaluating the teacher. The best evidence that the normal distribution of school grades is in accordance with the probability curve is the fact that whenever large numbers of typical records are studied they are found to be distributed in accordance with the theoretical considerations. As illustrations of this we may cite the study of Dearborn on the averages of 472 high school pupils (Fig. 3) and Johnson's study of the grades given dur-

ing two years in the University High School. The latter study includes all grades which are matters of record during the given period making a total of 13,726 grades. A later study, including over 25,000 grades is shown in Fig. 4. The close correspond-

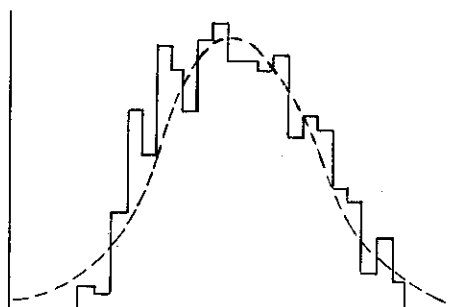


FIG. 3.

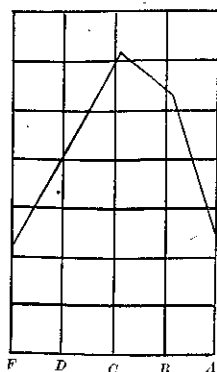


FIG. 4.

ence to the mathematical curve in all three cases is notable. The general correspondence in studies of this character is so close that the relation is now considered to be rather firmly established, and any serious deviations from this normal distribution are legitimate subjects of investigation.

Once it is granted that the law of variation holds in the main as to school grades, the way is open to investigate the work of departments, schools and individual teachers. A new and valuable tool has been gained. It is now possible to make quantitative and detailed comparisons between departments and between a department and the standard. Deviations from the theoretical distribution of marks may be entirely justifiable; indeed the circumstances may be such as to demand them as when the limitations of the course of study make the pupils of a class or department a highly selected group, but always a deviation from the normal raises a pertinent question calling for answer. Possibly the greatest value of such statistical study does not lie in the information conveyed by the simple curve of distribution but rather in the secondary investigations which are demanded in order to account for irregularities which appear. At the present stage in our knowledge of the exact condition of grading and indeed of most school affairs a new method of investigation, a new set of results, is likely to be of value precisely in proportion to the

stimulus which it gives to farther investigation and the definiteness with which it defines the starting point and direction of the needed investigation. Statistical examination of any set of grades does just this.

As a concrete example of such a study we may refer again to Johnson's "A Study of High School Grades" cited above. As examples of departmental studies there are presented here his figures showing conditions in four departments. (Fig. 5-8.)

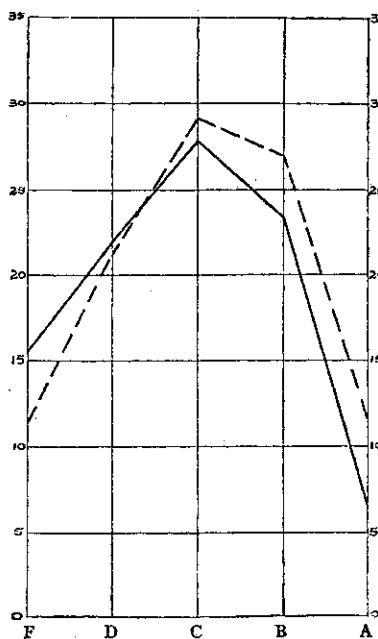


FIG. 5.

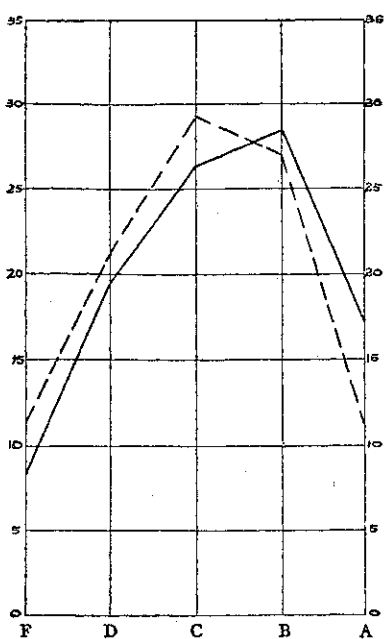


FIG. 6.

In each case the dotted line indicates the average of the school and the department curves may be compared directly with the same standard. All of these curves are constructed upon the basis of percentages and are therefore comparable. It will be noted that in fig. 5 there is an excess of D and F grades; in fig. 6 an excess of A and B; in fig. 7 a large deficiency of B grades and an excess of D and F; while in fig. 8 the most marked peculiarity is a very large proportion of B grades. In all of these cases the form of the curve relates it to the theoretical curve but it is shifted bodily to right or left. As to the interpretation

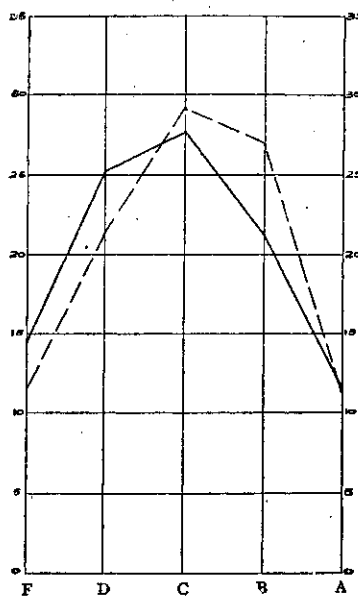


FIG. 7.

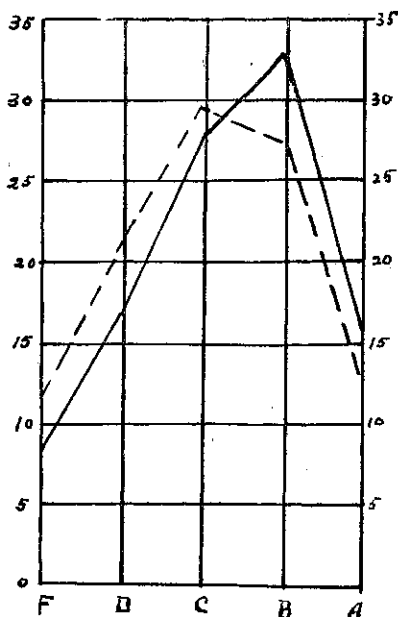


FIG. 8.

of this displacement, at least three explanations present themselves. If the curve is displaced toward the "F" end of the scale as in the case of fig. 7, it may be that the standard of the work expected of the pupils is higher than the age and ability of the pupils would justify; it may be that the material of the course is properly selected but that the instruction is relatively inefficient; or it may be that the standard of marking is too high. When the curve is displaced to the right as in fig. 8 the converse of these explanations would apply.

The comparison of individual teachers becomes a great deal more personal than the investigation of departments though even the latter is quite capable of arousing considerable individual interest. Two graphs from Johnson's paper may serve to illustrate individual peculiarities. One (Fig. 9) shows the divergence between the standards of two teachers within a single department when pupils and subject matter are as nearly uniform as possible and the other (Fig. 10) shows even greater divergence in a case where two departments are involved. One teacher, appearing in both figures, gives a maximum of high marks and a minimum of low ones while both of the other teachers group nearly half

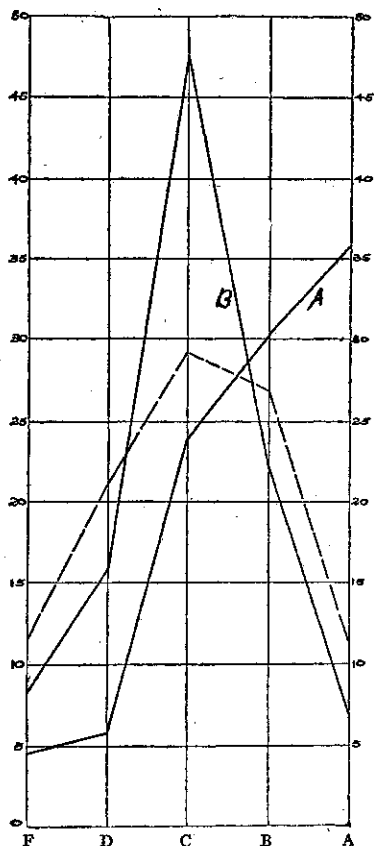


FIG. 9.

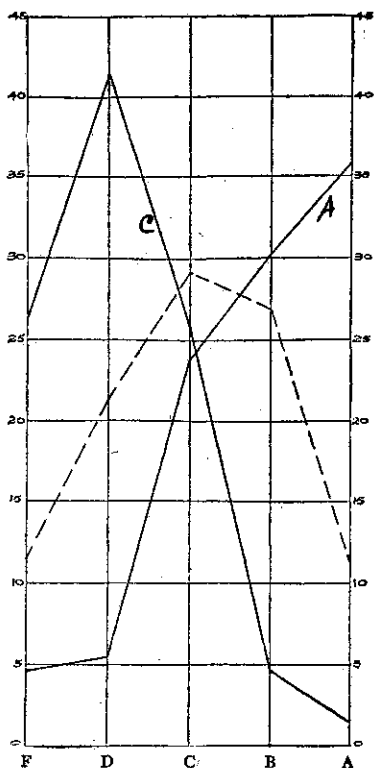


FIG. 10.

of the whole number in a single rather low grade. Evidently the subject matter cannot be so largely at fault for the matter of instruction is usually a department concern and there is not opportunity for such large variation, particularly in the case of two teachers in the same department. To take a particular case, let us consider the set of grades which find their greatest frequency in the D column (curve C). That this teacher's classroom is sought most largely by the lazy and incompetent is not probable and would doubtless be denied by the individual. The incompetents do not seek a teacher with a record for low grades. If then his pupils were of average ability it is not admissible to believe that only 2 per cent of them actually succeeded in accomplishing the full amount of work that might fairly be expect-

ed of them and therefore deserve A, nor is it easy to believe that more than 40 per cent of them succeeded in accomplishing only the minimum and 27 per cent failed entirely to measure up to expectations. To put it in another way, we may assume the grade "C" to express mediocrity, and in that case 80 per cent are alleged to be *below mediocrity*—rather a contradiction in terms. The only adequate explanation seems to involve either a correct standard of marking with inefficient teaching, or average teaching with an impossible standard. At any rate it can hardly be doubted that there is a large divergence in the standards of these several teachers and that this divergence works injustice to the pupils and to the teachers.

In certain institutions where this subject has been given a good deal of attention efforts are being made to so standardize the marking system that in the long run the distribution will be approximately the same for all teachers in the school. There seems to be a rather common agreement that the use of five grades the lowest of which indicates failure is convenient and scientifically acceptable. There is not yet such common agreement regarding the boundaries of the groups, particularly as to the number of pupils which we may expect to find the first and last groups.

Such statistical study is likely to prove of very distinct service to a new teacher entering a school for the new teacher is always at loss to know what difference there may be if any in the capacity of the new group compared with the old, and what difference in the school standard of marking. In the absence of any method of conveying this information from the administrative department, the first part of a new teacher's work becomes a sort of cut and try process. The statistical method will make it practicable for the school officers to give a new teacher some effective guidance in this rather difficult situation.

The writer may be pardoned for illustrating this point by reference to a study of the grades given by him in the University High School since other data are obviously not easily secured. Such study was particularly important to the writer because conditions in the University High School are very different in certain particulars from those obtaining in the public high school in which he had worked for some six years previously. In the latter school the author's courses were in the first year and were required of all pupils, the pupils came from widely divergent social conditions and were looking forward to many

different occupations with but few expecting to attend college. The University High School draws its pupils from a different and more homogeneous social group, the courses concerned are in part elective, and the majority of the pupils are expecting to enter college. The fact that the University High School is a tuition school is also a factor of considerable importance. Besides all this slightly different values are assigned to the letters used to indicate grades.

Recognizing these differences in the two schools but without positive data as to their effects, the author found considerable uncertainty in the matter of grading. Careful observation of the new conditions served to correct the more pronounced tendencies which were found in the grades given at first, but no amount of general observation would suffice to assure any very close approximation to the standard of the school. At the end of the second year an attempt was made to measure the degree of correspondence between the school standard and the individual standard. This study could not have been carried on successfully at an earlier time because there were not on record sufficient number of grades to make a normal distribution. This time was a favorable one for the study both because in the two years sufficient number of grades had accumulated to make a fairly typical distribution, and also because during these two years an intensive study of all the grades of the school for several years had been made by the Principal, and the distribution curve for the school and for each department was known.

The number of grades of record given by the writer during the two-year period studied is 295. Their distribution is shown graphically in Fig. 11, in which curve "C" represents the grades in question, while "B" represents the distribution in the science departments and "A" is the school average. It is of course perfectly obvious that while the number failing is almost the same as in other sciences and not widely different from the school average, the same does not hold true regarding other grades. There is a great preponderance in A and B grades with the corresponding deficiency in C and D grades, though the divergence from the practice of the science departments is not so great as the departure from the practice of the whole school. It is needless to say that this comparison has occasioned a very considerable revision of standards and methods as such a study may be expected to do in most cases.

While the writer is frank to admit that in this case the graph

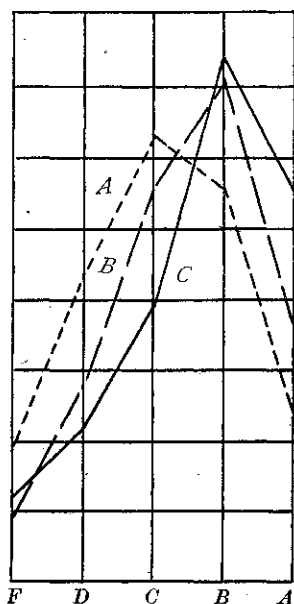


FIG. 11.

shown indicated a change in both standard of grading and material of instruction, it does not follow that this is always the case. There may be cases, particularly if the class be small, when anything else than a preponderance of high grades would be a great injustice. What it is intended to insist upon is that any great deviation from the normal should be the subject of careful consideration on the part of the teacher concerned, and that the teacher who gives an unusual number of high or low grades should make sure that he knows why he does so in each case. If it is certain that the reason lies in the characteristics of the pupil the correction must begin there, but if it lies in the subject matter or the standard of grading it is these that need correction. It can scarcely be

doubted that a careful statistical examination by every teacher of his own results would tend strongly toward more uniformity of grading.