MISCELLANEA.

CONTENTS.

SOME EXPERIMENTS TO TEST THE THEORY OF GOODNESS OF FIT.

By JOHN BROWNLEE, M.D., D.Sc.

SOME time ago in 1911, in discussing the possibility of applying a Mendelian analysis to Dr. Beddoe's observations on hair and eye colour in Europe, I came across a distribution of error as estimated by χ^2 which was incompatible with the theory at that time developed. Later, Greenwood and Yule, discussing the statistics relating to inoculation in enteric fever, found a similar discrepancy. Explanations of this discrepancy were given by Mr. Fisher and Mr. Yule in the *Journal of the Royal Statistical Society*, January, 1922, where Mr. Fisher discussed the subject from a mathematical, and Mr. Yule from an experimental, point of view.

With a view of testing the possibilities of application of the method as it appeared in some of my work, I have had a series of coin-tossing experiments made. The method was as follows. A cylinder of about 6 inches in diameter, which ended in a cone and a funnel of diameter convenient in size to hold a penny, was constructed. Sixty-four pennies were placed in this cylinder, shaken thoroughly up, and allowed to fall into the funnel. From this they were removed by means of a movable cap and placed in a column on the table. The pennies were taken successively from this column in pairs, and the nature of the result, two heads, a head and a tail or two tails, recorded. These sixty-four pennies were shaken 512 times. When the records were analysed it was found that two heads occurred 4,010 times, a head and a tail 8,191 times, and two tails 4,183 times. When the pennies were grouped in fours, it was found that four heads occurred 468 times, three heads and a tail 2,026 times, two heads and two tails 3,072 times, one head and three tails 2,121 times, and four tails 505 times. In all a head has occurred 16,215 times and a tail 16,553 times, suggesting a slight bias in favour of a tail, which might be expected, as the centre of gravity of a penny must be slightly nearer the face of the coin containing the head than the tail. Absolute equality would give 16,384 heads, the difference between this and the number found being 169. As the standard error is 90·5, the deviation from equality, however, is not sufficiently great to make the presence of a bias certain.

Having ascertained that the numbers found in the experiment were sensibly those expected *à priori*, the observations were grouped in different ways, fitted to a variety of formulæ and the distribution of χ^2 examined. In the first place, the records of successive tosses of two pennies were taken in groups of 256, so that there were in all 64 instances. The theoretical fit of these, on the assumption of no bias, is 64 double heads, 128 heads and tails and 64 double tails.

TABLE I.—Comparison of distribution of χ^2 in fitting the experiments to 64, 128, 64, and the theoretical distribution $n' = 3$.

χ^2	< 1	1	2	3	4	5	6
Experiment	25	17	7	6	5	1	3
Theory	25	15	9	5	4	2	3

The comparison of the distribution of χ^2 found when the experiments were tested against the framework and of the distribution given by the theory of chance is shown in the above table. In this case $n' = 3$. It is obvious at once that the theory and experiment are in very close accord.

The 64 instances were next fitted, not to the point binomial $(\frac{1}{2} + \frac{1}{2})^2$, but to a point binomial which allowed for a slight bias on the one side or other, namely $\{(\frac{1}{2} + \varepsilon) + (\frac{1}{2} - \varepsilon)\}^2$, when ε is a small quantity. If α, β, γ denote the number of two heads, a head and a tail and two tails, the value of ε calculated by least χ^2 is given by $\{\beta^2 + 6\overline{\alpha^2 + \gamma^2}\}\ \varepsilon = (\alpha^2 - \gamma^2)$. In this case one degree of freedom has been removed. The observations cannot deviate so far from the framework, consequently there is a great increase in the number of good fits. The distribution of χ^2 corresponds now, not to the value of $n' = 3$, but $n' = 2$, as may be seen by the comparison in the accompanying table.

TABLE II.—Comparison of distribution of χ^2 in fitting observations to the term of $256(\frac{1}{2} + \varepsilon + \frac{1}{2} - \varepsilon)^2$, and the theoretical distribution $n' = 2$.

χ^2	<1	1	2	3	4	5	6–
Experiment	43·0	11·0	5·0	2·0	1·0	1·0	1·0
Theory	43·7	10·2	4·7	2·4	1·3	0·7	1·0

The fitting in the preceding paragraph comes into direct line with the case which occurred when a Mendelian analysis was applied to Dr. Beddoe's record of hair colour.* Assuming in this case that the number of brown- and fair-haired persons in a race is a and the number of jet black-haired persons b, if the population mate freely and fertility be equal, it will be distributed in the form of a^2 brown- and fair-haired persons, $2ab$ dark-haired persons (the cross), and b^2 jet black-haired persons. In any case, even with only moderately free mating, stability would be practically accomplished in the course of five or six generations. Taking Dr. Beddoe's records from the country districts where such conditions may be assumed, the values of χ^2 were calculated after each instance had been fitted to a framework of $a^2 + 2ab + b^2$. The distribution of these values of χ^2 are given in the accompanying table in the first column, in the second the distribution expected when $n' = 2$, and the third when $n' = 3$. Here it is obvious again that the comparison is with $n' = 2$; one degree of freedom has been taken away. The population from which the random sample was taken is not known, and the only framework with which the data can be compared is one calculated from the data themselves. This was kindly pointed out to me by Mr. Fisher. The resemblance to the coin-tossing experiments seems complete.

* Dr. Beddoe made observations on the colour of hair for a large number of rural districts in England. In 169 instances the number of persons examined exceeded 100 and to these observations a Mendelian analysis was applied.

TABLE showing the distribution of χ^2 when a Mendelian analysis of the hair colour of 169 English villages is compared with the theoretical distributions.

χ^2.	Observations.	Theoretical.	
		$n' = 2$.	$n' = 3$.
0–1	125	115·4	66·5
1–2	21	27·0	40·4
2–3	12	12·5	24·4
3–	11	14·1	37·7

Turning now to experiments with four coins, there are 32 examples of 256 tossings. The distribution of χ^2 has been calculated for 3 cases: Firstly, against the known theoretical value, namely, four heads occurring 16 times, three heads and one tail 64 times, etc. Secondly, with the framework determined from the individual sample by the method already mentioned, so that the comparison is made with the successive terms of

$$256\ (\overline{\tfrac{1}{2} + \eta} + \overline{\tfrac{1}{2} - \eta})^4$$

If the individual numbers in the tossings are denoted by α, β, γ, δ, ε, and the value of η obtained by the least χ^2 method, the first approximation is

$$\{60\ (\alpha^2 + \varepsilon^2) + 6\ (\beta^2 + \delta^2) + \gamma^2\}\,\eta = 24\ (\varepsilon^2 - \alpha^2) + 3\ (\delta^2 - \beta^2).$$

Thirdly, the normal curves with the use of Sheppard's correction for the second moment was calculated for each instance. The results are shown in the accompanying table. The distribution of the values of χ^2 expected when n' is 5 and n' is 4 are given in the first two columns of the table and then successively the distribution of χ^2 in the three experimental fittings. It will be observed that when the rigid framework is taken, the values when $n' = 5$ give a very close correspondence between theory and experiment. When the framework is derived from the sample, the correspondence is very close to the distribution in which n' is equal to 4 and very divergent from that in which n' is equal to 5, while when the normal curve is used something between the two is found. Using the normal curve, Mr. Fisher, as I understand him, would say that two degrees of freedom have been taken away, the mean point and the standard deviation being adjusted in each case to the sample. The comparison should thus be to the distribution $n' = 3$, but in reality, judging by the χ^2 test, it is nearest to the distribution $n' = 5$. I probably, therefore, misunderstand Mr. Fisher upon this point.

TABLE showing the distribution of χ^2 in 33 samples, theoretically and experimentally.

χ^2.	$n' = 5$.	$n' = 4$.	Rigid Framework.	Adjusted Framework.	Normal Curve.
0–2...	8·4	13·7	7	12	10
2–4....	10·6	9·9	10	12	12
4–6....	6·6	4·8	10	6	5
6–	6·4	3·6	5	2	5

Some fittings to a parabola have also been made in the case of the tossing of two coins. Denoting the number of cases of two heads, etc., by α, β, γ, these numbers have been graduated to the expression $a + bx + cx^2$. In this case the origin has been taken in the middle of β, and the distance of α and γ from this respectively positive and negative unity. In order that the total may be 256, the condition $3a + 2c = 256$ must be fulfilled, as in the χ^2 method the sum of the positive and negative errors must be zero. Two cases are possible. In the first, b is taken equal to zero and a and c vary subject to the condition; in the second, a is taken equal to 128, and c equal to $- 64$, and b varies. If b and c are allowed to vary together, the condition determines the whole three exactly and χ^2 is equal to zero. In the two cases considered the fitting has been made by the method of least χ^2, and the results between theory and experiment are given in the accompanying table. Again, it is seen that $n' = 2$ is the accurate usage.

TABLE showing the distribution of χ^2 in the 64 instances under the conditions referred to in the text.

χ^2.	$n' = 2$.	$b = 0$. $3a + 2c = 256$.	$a = 128$. $c = - 64$. b variable.
0–1	43·7	40	41
1–2	10·2	15	14
2–3	4·7	6	3
3–	5·3	3	6

The χ^2 test has been applied by others as well as myself, without the necessary condition, however, being fulfilled, in the hope that the error introduced would be inconsiderable. It has, therefore, been thought advisable to test some cases to determine the extent of the error introduced. The experiments of tossing two coins have been fitted to the expression $a + bx + cx^2$. If α, β and γ denote the three terms, then $\alpha = a - b + c$, $\beta = a$, $\gamma = a + b + c$; allow some of these to vary and determine the best values by least squares or least χ^2. If the whole quantities be allowed to vary, of course a perfect fit is obtained and χ^2 is equal to zero. Some of the cases are identical with those which have already been discussed, and only those in which the positive and negative errors are not equal are described.

Case I.—$a = 128$, $b = 0$, c varies, the value given by least squares being $c = \frac{1}{2}(\alpha + \gamma) - 128$. In this case, as will be seen from the table, n' is equal to 2·5.

Case II.—a is constant, b and c vary. The values found by least squares are $a = 128$, $b = \frac{1}{2}(\gamma + \alpha)$, $c = \frac{1}{2}(\gamma + \alpha) - 128$. In this case the distribution of χ^2 becomes equivalent to $n' = 1$.

Case III.—b constant, a and c vary. This gives $a = \beta$, $b = 0$, and $c = \beta - \frac{1}{2}(\alpha + \gamma)$. The distribution in this case corresponds to $n' = 2$.

If the χ^2 method is to be used, it is obvious then that the fundamental condition that the sum of the errors be zero must be fulfilled.

TABLE showing the distribution of χ^2 in cases of tossing of two coins when the numbers are fitted to the expression $a + bx + cx^2$, without the fundamental condition being fulfilled.

	Theory. $n' = 2 \cdot 5$.	Actual.	Theory. $n' = 1$.	Actual.	Theory. $n' = 2$.	Actual.
0–1 ..	33·8	33	48·0	54	43·7	41
1–2 ...	17·8	16	8·1	6	10·2	14
2–3 ...	6·8	6	3·6	3	4·7	5
3–4 ...	3·3	6	1·6	1	2·4	3
4–5 ..	1·4	1	1·3	—	1·3	—
5– ...	0·9	2	—	—	1·7	1

In conclusion, it is evident that a new criterion of the value of n' that is to be taken will require to be worked out. Taking the case of tossing four coins where the fit has been made to three different frameworks, it seems evident enough that the comparison with the theoretical population known *à priori* should be made with the value $n' = 5$; that if the form of the population be known. the introduction of a corrected term will increase the number of good fits. In this case the terms of a binomial, if a corrected term be introduced so as to derive the framework from the actual sample, a comparison should be made with $n' = 4$. In practice, however, these conditions are not often fulfilled, and some curve like the normal curve is used to graduate the statistics. In the case considered here the normal curve does not give such good results as the point binomial, and it seems difficult *à priori* to decide what value of n' should be used.

There is a further point, however, which requires to be discussed, and that is the legitimacy of assuming the fundamental condition, namely, that the positive and negative errors are of equal value. In graduating, for instance, for some diseases, the number of deaths at each year of age from 1 year to 10, it is found that the death-rate decreases in a geometrical progression. Taking the case of scarlet fever in Glasgow during the end of last century, this ranges from 20 per cent. to 2 per cent. It does not seem to me legitimate to make the total number of deaths according to the graduated curve equivalent to the total number of deaths found in the statistics when such extreme variations in the death-rate occur. A case like this is better tested by calculating the probable error of each observation at each year of life, and noting whether the grouping of the actual errors with reference to the probable error agrees with the theory of chance.