Proceedings of the Iowa Academy of Science

Volume 37 | Annual Issue Article 66

1930

A Statistical Test of Experimental Technique

George W. Snedecor
Iowa State College

Recommended Citation
Snedecor, George W. (1930) "A Statistical Test of Experimental Technique," *Proceedings of the Iowa Academy of Science, 37(1)*, 279-287.
Available at: https://scholarworks.uni.edu/pias/vol37/iss1/66

A STATISTICAL TEST OF EXPERIMENTAL TECHNIQUE

GEORGE W. SNEDECOR

The phrase "experimental technique" is used in a largely inclusive sense in this paper. A more precise title would be "A Statistical Test of Homogeneity." The reason for the more flexible phrase will be apparent, I trust, as the argument progresses.

The test to be described is applicable to that type of data known as "homograde statistics" (2) or the "statistics of attributes" (11). This is the type which arises when the individuals in a sample are classified in alternate catagories such as male or female, dead or alive, infested or free, wrinkled or smooth, yellow or green. Although it is one of the oldest and most highly developed branches of statistics, it was for years somewhat overshadowed among biologists by Karl Pearson's appealing presentation of his findings in "heterograde statistics" or the "statistics of variables." Due to the increasing interest in genetics, in studies of immunity, in the relative potency of disinfectants, etc., the method is being brought into even greater prominence than before. Its practical usefulness has been greatly extended through its recent adaptation by R. A. Fisher (3 and 5) to the requirements of small numbers of observations.

For purposes of introduction, I shall illustrate the older form of the test, applicable to large numbers of observations. Table I contains a distribution published by McPhee in 1927 (6) of litters of 8 pigs arranged according to the number of males per litter.

Table I — Distribution of Litters of Eight Swine According to Number of Males in Litter

Number of Males	Number of Litters	Expected Number of Litters	Deviation	(Deviation)²	(Deviation)² / Expected
0	0	0.3	—0.3	0.09	0.30
1	5	2.6	2.4	5.76	2.22
2	9	9.8	—0.8	0.64	0.07
3	22	21.3	0.7	0.49	0.02
4	25	28.8	—3.8	14.44	0.50
5	26	24.9	1.1	1.21	0.05
6	14	13.5	0.5	0.25	0.02
7	4	4.2	—0.2	0.04	0.01
8	1	0.6	0.4	0.16	0.27
	106	106.0	0.0	(1-13)	3.46

After calculating the average percentage of males per litter, 51.77%, the third column in the table is computed on the assumption of a binomial distribution. This involves the assumption that the probability or chance that any individual be a male is uniform throughout the sample. The next column headed "Deviation," contains the differences between actual and expected frequencies. Our problem is to test whether this set of deviations from expected is just such a set as might be expected to occur in random sampling, or whether the deviations are too great to warrant such an assumption. The mechanics of the test consists in squaring each deviation, dividing each square by the expected frequency, and adding the results. The sum is a statistic known as "Chi Square," written χ^2, originated by Karl Pearson (8) in 1900. By consulting a table of values of χ^2, we find that if the number of classes is nine, such as is the case in our table, any value of χ^2 lying between 1 and 13 may occur merely as a result of chance variation from the theoretical distribution. Since we have χ^2=3.46, we conclude that this distribution of males deviates from expected no more than is usual in samples of such size; in other words, we conclude that this is just such a sample as may normally be drawn from a homogeneous population, in which there is a uniform probability of maleness. This is the test of homogeneity. Experimental technique is comparatively simple in this case. The experimental animals are well chosen, the records are accurately kept, and the size of the sample is adequate.

In the next table (II) there is shown a similar distribution

Table II — Distribution of Litters of Eight Swine According to Number of Males in Litter

Number of Males	Number of Litters	Expected Number of Litters	Deviation	(Deviation)^2	(Deviation)^2 / Expected
0	1	1.8	— 0.8	0.64	0.36
1	8	14.2	— 6.2	38.44	2.71
2	37	47.0	—10.0	100.00	2.13
3	81	90.9	— 9.9	98.01	1.08
4	162	109.6	52.4	2,745.76	25.04
5	77	84.8	— 7.8	60.84	.72
6	30	41.1	—11.1	123.21	2.99
7	5	11.2	— 6.2	38.44	3.42
8	1	1.4	— 0.4	0.16	0.11
	402	402.0	0.0	(1-13)	38.57

published by Parkes in 1923 (7). The probability of maleness is only 49.16%. The column of deviations indicates clearly the lack of agreement between actual and expected numbers of litters. This disagreement is evaluated numerically in the resulting

χ^2=38.57, a value far beyond the range (1-13) allowable in random sampling. Evidently this sample was not drawn at random from a homogeneous population. The experimental technique is called in question. Did the probability of maleness vary from litter to litter or from pig to pig? Were the data accurately reported? These and others which may present themselves to you are questions for the experimenter. The statistician asserts that the variation of this distribution from the binomial is greater than can be accounted for by the exigencies of random sampling. Either the population was not homogeneous for maleness, or the sampling was not properly done.

We now turn from the classical method of applying this test in large samples to that devised by R. A. Fisher (3), applicable to small numbers of sub-samples. The illustration (table III) is

Table III — Numbers of Medium-smooth Ears of Corn in Eight Samples of One Thousand

Sub-sample number	1	2	3	4	5	6	7	8
No. M-S ears, X	449	448	539	580	500	520	498	572

$p = 51\%$ $\quad$ $q = 49\%$

$\Sigma X = 4{,}106$ $\quad$ $\Sigma X^2 = 2{,}124{,}814$

$M_x = 513.2$ $\quad$ $(\Sigma X) M_x = 2{,}107{,}404$

$qM_x = 249.8$ $\quad$ Difference $= 17{,}410$

$$\chi^2 = \frac{17{,}410}{249.8} = 70. \quad (2-14)$$

taken from Burnett's investigation (1) of the vitality of corn. We are now dealing with only eight sub-samples. If this distribution were mathematically comparable to those already considered, we should have a thousand and one classes, requiring many thousands of sub-samples to fill out the distribution. The method presented yields a value of χ^2 which is used in the same way as that already described. As indicated in the table, we obtain here the value χ^2=70 which lies far outside of the allowable range (2-14). The conclusion is that this sample is not typical of random samples drawn from a field of corn homogeneous as to medium-smoothness of ears. From a statistical standpoint, the sample is not homogeneous for this character. We cannot subject such a sample to further analysis with confidence that it is representative of the field so far as smoothness of ear is concerned.

The necessary computation includes (1) the mean number of medium-smooth ears per thousand (513.2), (2) the probability that an ear be not medium-smooth (q=0.49), (3) the sum of the squares of the deviations from the mean (17,410), and (4) the value of χ^2 obtained from the formula,

$$\chi^2 = \frac{\text{sum of squares of deviations from mean}}{(\text{probability against M-S}) \times (\text{mean no. M-S})}.$$

Confirming the foregoing results, we have other groups drawn from Burnett's experiments and displayed in the table IV. The

Table IV — Numbers of Ears of Corn in Eight Sub-samples of One Thousand in Each of Five Grades

Sub-sample No.	1	2	3	4	5	6	7	8	χ^2.(2-14)
Smooth	157	119	128	96	113	155	128	135	26
Medium-Smooth	449	448	539	580	500	520	498	572	70
Medium	246	236	175	195	240	198	246	214	23
Medium-Rough	98	127	101	85	94	88	91	53	35
Rough	50	70	57	44	53	39	37	26	29

sample of medium-smooth ears just discussed is re-entered in the second line. The values of χ^2 in the right hand column indicate that the experimental technique (in the sense I am using that phrase) was inadequate in all of the other four grades of smoothness. Perhaps the field was not homogeneous for these other grades. Perhaps the number of sub-samples was not adequate. Perhaps the method of classification was faulty. Whatever the explanation, the fact remains that any statistical conclusions must be tempered by the knowledge that either there was in the field no uniform probability of these several grades of smoothness, or else the sampling was not well done.

Drawing again upon the vast store of data in Burnett's thesis, we apply the χ^2 test (table V) to the results of the germination experiments, confining our attention to the sample of medium-smooth ears already found to be non-homogeneous as to smoothness. The question now is this: "Can we assume that this sample

Table V — Viability in Medium-smooth Ears of Corn, Eight Sub-samples

No. Sub-Sample	No. Kernals Tested	No. Germinating	Percentage
1	2,694	2,462	91
2	2.688	2,569	96
3	3,234	2,788	86
4	3,480	2,936	84
5	3,000	2,217	74
6	3,120	2,240	72
7	2,988	1,832	61
8	3,432	2,404	70
	24,636	19,448	79
Average No. = 3,080 $\sigma_B^2 = 0.54$		$p = 79\%$ $\sigma^2 = 117$	$q = 21\%$ $\chi^2 = 1727.(2—14)$

was homogeneous as to viability of seed?" The answer is an emphatic "No." χ^2=1727, whereas the allowable range in random sampling is from 2 to 14. Again using the phrase in the loose sense adopted in this paper, the "experimental technique" must be examined as to (1) homogeneity of the field in respect to viability of seed, (2) adequacy of the number of kernels (six) taken from each ear, (3) pattern used in drawing the kernels from the ear, (4) the technique of germinating the seeds, and perhaps other points which may suggest themselves to a biologist. On the face of the returns it is lack of homogeneity in the field which seems to be indicated. This could be determined by making a test of the aggregate of the other three alternatives; i.e., by examining the individual sub-samples of six kernels in each sample of medium-smooth ears per thousand. This would be done in the same way as the samples of swine litters were tested for homogeneity of maleness. If homogeneity were found in each thousand, the lack of homogeneity would thus be isolated in the field. The practical conclusion from such a result would be this: any comparisons of germination tests would have to be confined to samples drawn from only such areas in the field as proved to be homogeneous as to viability of seeds.

Up to the present point in this paper, we have discussed cases in which the size of the sub-sample was constant throughout any one sample. Now we shall consider the case, which insistently intrudes itself upon us, in which the number of individuals in the sub-sample varies. R. A. Fisher (3, page 89) mentions this possibility and proposes a method of procedure. We have, however, developed here at Iowa State College a method which we consider to be usually more satisfactory. We took our cue from Arne Fisher (2) making the necessary modifications and additions. Naturally, the lack of uniformity in sub-sample size detracts somewhat from the statistical validity of the test, but we have obtained results which seem worth while. So far as the value of χ^2 is concerned, it is identical with that obtained in R. A. Fisher's suggested method.

Necessary explanation of our method will now be given in connection with the data (from Burnett again) of table VI. Column three reveals the fact that the number of kernels tested in the sub-samples varied from 48 to 120. We are forced to use an average sized sub-sample (n=68), along with the usual average percentage of germinating seeds (p=90%). The mechanism of the computation is sufficiently well indicated in the table. The normal size of the resulting χ^2 (seven) indicates homogeneity as to via-

Table VI — Viability of Corn Stored in Six Different Locations

Place of Storage	No. Ears	No. Kernals Tested	No. Germinating	Percentage
Seed room	10	60	57	95
Garret (kitchen)	20	120	111	92
Toolshed (closed)	10	60	55	92
Toolshed (open)	10	60	55	92
Hung outdoors	8	48	41	85
Dry garret	10	60	50	83
	68	408	369	90

Average No. $= \frac{408}{6} = 68$, $p = \frac{369}{408} = 90\%$, $q = 10\%$

$\sigma_B^2 = \frac{90 \times 10}{68}$, $= 13.$

$\frac{100 \text{ (Sum of products)}}{\text{no. of kernels}} = \frac{3,343,590}{408} = 8,195$

$p^2 = (90.44)^2 = 8,179$

Difference $= 16$

$$\chi^2 = \frac{6 \times 16}{13} = 7. (1—11)$$

bility in the sample. The phrase "experimental technique" is now still further stretched to include the variation in locality of storage — an experimentally controlled factor. Since therefore the sample has remained homogeneous under such control, our conclusion also may be enlarged. We are now warranted in deducing the fact that any differences in viability due to locality of storage are statistically non-significant. Thus in this case our test covers not only the homogeneity of the experimental material and adequacy of technique, but also the non-significance of controlled differences in treatment.

The next table (VII) with its value of χ^2=145 (allowable χ^2

Table VII — Germless Seeds from a Certain Cross in Maize, Twenty-one Ears

Total Seeds	Germless Seeds	Percentage	Total Seeds	Germless Seeds	Percentage
442	97	22	378	53	14
408	89	22	534	38	7
290	66	23	239	52	22
311	34	11	205	48	23
357	99	28	469	94	20
569	117	21	172	14	8
220	29	13	415	92	22
276	42	15	510	99	19
370	69	19	492	72	15
276	51	18	213	17	8
448	81	18	7,594	1,353	17.8

$\chi^2 = 145.$ (11 — 32)

from 11 to 32) illustrates a case of non-homogeneous experimental material. The data are taken from a new bulletin by Wentz (10). The probability of the event, "germless seed," cannot be considered as uniform throughout the sample, and no conclusions would be valid if based on the assumption of such uniformity.

In the next set, (table VIII) the test again covers not only the character of the variation in the experimental material, but also

Table VIII — Sex-Ratio in Guinea Pigs — by Months, Seven Year Averages

Month	Number Born	Number Males	Percentage
January	114	65	57
February	122	64	52
March	146	65	45
April	89	41	46
May	134	72	54
June	160	80	50
July	183	88	48
August	232	114	49
September	174	80	46
October	233	129	55
November	256	112	44
December	171	86	50
	2014	996	49.45
	$\chi^2 = 14.$	(5 — 20)	

the effect of a certain type of experimental control — a very mild type which consists in merely tabulating the results by months. The data are taken from a recent article by Schott and Lambert (9). The question raised is this: "Do the different sex-ratios as tabulated indicate a significant effect of season on sex-ratio, or are they merely such as would normally arise in random sampling from a population homogeneous as to sex-ratio?" Since $\chi^2=14$ with a sampling range of 5 to 20, the conclusion is that the seasonal effect, if any, is masked by the usual experimental variation in samples of this size. No lack of homogeneity is evident.

The non-homogeneous character of the two series in table IX (12) is obvious. The values of χ^2 are far in excess of those normally found in random sampling from a homogeneous population. The test reveals the changing probability of injury not only from plot to plot, but also from tree to tree. The latter variability is so great that no valid conclusion may be drawn as to the effects of the different methods of treating the plots. So long as the technique of choosing the trees, applying the chemicals, and sorting the fruits yields such variable probability of injury, differences due to experimental control are submerged in the greater differences due to experimental variation.

Table IX — Technique of Spraying and Dusting Apples

Sprayed Plot				Dusted Plot			
Tree Number	Number Apples Examined	Number Apples Injured	Percentage Injured	Tree Number	Number Apples Examined	Number Apples Injured	Percentage Injured
1	1804	102	5.6	1	1083	118	10.9
2	1811	88	4.9	2	1011	48	4.8
3	860	2	0.2	3	946	128	13.5
4	1671	7	0.4	4	840	37	4.4
5	1078	11	1.0	5	2347	41	1.8
6	1204	9	0.8	6	2404	69	2.9
7	1199	17	1.4	7	2548	38	1.5
8	2149	13	0.6	8	2376	38	1.6
	11,776	249	2.1		13,555	517	3.8
$\chi^2 = 256. (2 — 14)$				$\chi^2 = 481. (2 — 14)$			

Table X — Mortality Among Chicks Inoculated with Tuberculosis, Two Generations

1927			1928		
Number Inoculated	Number Dying	Percentage	Number Inoculated	Number Dying	Percentage
42	14	33	91	22	24
93	28	30	75	17	23
80	34	42	48	8	17
44	23	52	28	8	29
25	9	36	71	29	41
			42	9	21
			42	10	24
284	108	38	397	103	26
$\chi^2 = 7. (1 — 9)$			$\chi^2 = 12. (2 — 13)$		

In the last table (X), we observe the results of a successful technique applied under difficult circumstances. The investigation was reported by Irwin (4). The chicks tested in 1928 were the offspring of those surviving the 1927 inoculation. The samples proved to be homogeneous for probability of death despite the high variability of chicks under experimentation, the differences in season, the changing culture, and the actual technique of applying the injection.

BIBLIOGRAPHY

1. Burnett. The vitality of corn. Thesis, Iowa State College (1906).
2. Fisher, Arne. The mathematical theory of probabilities. Macmillan (1922).
3. Fisher, R. A. Statistical methods for research workers. Oliver and Boyd (revised, 1928).
4. Irwin. Disease resistance and its inheritance with reference to its application in animal breeding. Thesis, Iowa State College (1925).
5. "Mathetes". Statistical study on the effect of manuring on infestation of barley by gout fly. Annals of Applied Biology, 11:220 (1924).

6. McPhee. The swine herdbook as a source of data for the investigation of the sex ratio and frequency of sex combinations in pig litters. Journal of Agricultural Research, 34:714 (1927).
7. Parkes. Studies on the sex-ratio and related phenomena. Biometrika 12:373 (1923).
8. Pearson. 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. Philosophical Magazine, Series 5, 50:157 (1900).
9. Schott and Lambert. The normal variability of the sex-ratio in the Guinea Pig. Iowa State College Journal of Science, 4:343 (1930).
10. Wentz. The inheritance of germless seeds in maize. Iowa State College Research Bulletin No. 121 (1930).
11. Yule. An introduction to the theory of statistics. Chas. Griffin and Company, Ltd. (eighth edition, 1927).
12. Hartzell. Application of one of Pearson's probability theorems and some special probability equations to entomological data. Journal of Economic Entomology, 22:202 (1929).

Iowa State College,
Ames, Iowa.