1939

A critical study of analysis of variance

Earl B. Gardner

The University of Montana

A CRITICAL STUDY of ANALYSIS OF VARIANCE

by

Earl B. Gardner
B. A., State University of Montana, 1938

Presented in Partial fulfillment of the requirement for the degree of Master of Arts.

State University of Montana

1939

Approved:

A. S. Merrill
Chairman of Board of Examiners

W. G. Bateman
Chairman of Committee on Graduate Study

CONTENTS

Introduction.

In the mathematical literature of recent years there has appeared at times reference to a method of comparing statistical data, called Analysis of Variance. It is the purpose of this study to investigate the mathematical and historical developments of the method and to examine certain of its applications.

During the course of development, capital letters such as N, the number of items, and M, the mean, will be used to refer to the population. Small letters, n and m, will be used in reference to a sample. x is a variable item, and $\bar{x}$ with a subscript will represent the mean of a portion of the sample as denoted by the subscript. s^2 will be the variance of the population, while v^2 will be an estimate of s^2 based on a sample. S will be used to indicate summation, and when it appears with a subscript, it will denote summation with respect to that subscript only.

Other symbols will be explained as introduced.

MATHEMATICAL DEVELOPMENT

i. Variance of a Population.

If p represents the probability of success and q the probability of failure in any one trial for an event, then the probability of exactly r successes out of N trials is known to be ${}_NC_r\, p^r\, q^{N-r}$ where ${}_NC_r$ is the number of combinations of N things taken r at a time.[1] This expression is the $(N-r+1)^{th}$ term in the expansion of $(q+p)^N$.

If mathematical expectation is defined as the average return per trial in a large number of trials for a prize, then clearly if the prize be D and the probability of winning it in one trial is p, the ME of the person who makes one trial is pD. Also, if there exist a number of independent, mutually exclusive ways in which success may be obtained in a given trial, the probability of success in one trial is the sum of the probabilities of success of the independent and mutually exclusive events.[2]

Thus the mathematical expectation of the number of successes referred to in the first paragraph is the sum of all of the mutually exclusive possibilities, each multiplied by the number of successes of which it is the

---

1. Hall and Knight, Higher Algebra (London, 1936) p. 385.
2. Ibid., p. 381.

probability. That is, $ME = S[{}_NC_r\ p^r\ q^{N-r}\ r]$ where the summation extends from $r = 1$ to $r = N$. This equation may be written[3] $ME = S[\frac{N!}{r!(N-r)!}\ p^r\ q^{N-r}\ r]$

$$= S[\frac{N!}{(r-1)!(N-r)!}\ p^r\ q^{N-r}]$$

$$= S[\frac{(N-1)!}{(r-1)!(N-r)!}\ p^{r-1}\ q^{N-r}\ Np] = Np(q+p)^{N-1} = Np \ldots\ldots\ldots (1)$$

It may also be shown[4] that Np is the most probable or modal number of successes in N trials. Let this be represented by M. Let x represent an observed number of successes and d the discrepancy x-M: the mathematical expectation of the square of the discrepancy[5] will be

$$S[{}_NC_r\ p^x\ q^{N-x}\ (x-M)^2] = Np + N(N-1)p^2 - 2NpNp + N^2p^2$$
$$= Npq.$$

Variance is defined as the square of the standard deviation: that is, the square of the most probable deviation from the mean. Therefore, if s^2 be the variance of the population of N items which is being considered,

$$s^2 = Npq \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

This may be seen clearly if one remembers that each x is a certain number of successes, not M, and each x is accompanied by a d, therefore by a d^2. Then, as before, the ME of d^2 is given by the sum of the products of each d by its probability of occurrence. The method of reducing the

---

3. Rietz, H. L., Mathematical Statistics (Chicago 1927) p. 26.
4. Ibid., p. 25.
5. Ibid., pp. 26-7.

sum is the same as that used in deriving equation (1).

ii. Distribution of $(x-M)/s$.

In order to determine the probability of the occurrence of a deviation d_1 there must be set up a functional equation with d independent. Let the terms of $(q+p)^N$ be represented by ordinates, y_d. Then $d = x-Np$, and

$$y_d = \frac{N!}{(Np+d)!(Nq-d)!} p^{Np+d} q^{Nq-d} \quad \text{.........................(3)}$$

But since this expression will not readily admit itself to summation, use may be made of Stirling's formula[6] which reads:

"If the expression $N!$ be replaced by the expression $N^N e^{-N} \sqrt{2\pi N}$ the true value will have been divided by a number lying between 1 and $1 + \frac{1}{10N}$."

Upon making this substitution, equation (3) becomes

$$y_d = \frac{1}{\sqrt{2\pi Npq}} [1 + \frac{d}{Np}]^{-Np-d-\frac{1}{2}}[1 - \frac{d}{Nq}]^{-Nq-d-\frac{1}{2}} \quad \text{..........(4)}$$

By the use of logarithms,[7] a close approximation to y_d is found to be

$$y_d = \frac{1}{\sqrt{2\pi Npq}} \exp(-d^2/2Npq),$$

where the error introduced will vary approximately as $\exp(d/N)$, and since d is the difference between x and the mathematically expected value of the x's, this factor will be insignificantly small compared with x. By (2) $s^2 = Npq$,

6. Stirling, Methodus differentialis, p. 135, quoted in J. L. Coolidge, Introduction to Mathematical Probability, (Oxford 1925) pp. 39-41.
7. Rietz, op. cit., p. 27.

and equation (4) becomes[8]

$$y_d = (1/s\sqrt{2\pi}) \exp(-d^2/2s^2) \quad \text{.........................(5a)}$$

or, making the substitution $X = d/s$,

$$y_d = (1/s\sqrt{2\pi}) \exp(-X^2/2) \quad \text{..........................(5b)}$$

If the area under the whole curve (5b) be taken as one, then ydX will represent the relative frequency of the deviations lying within that infinitesimal interval, since it gives the fractional part of the area under the curve occupied by that interval. Therefore, the probability, dy, of any X picked at random falling in the interval X to X+dX is given by the differential equation of probability

$$dy = (1/\sqrt{2\pi}) \exp(-X^2/2)\, dX \quad \text{.........................(6)}$$

Equations (5) represent the bell-shaped normal curve of relative frequency, and the differential equation of probability (6) might be interpreted geometrically by the shaded area in figure 1, this infinitesimal area representing the relative frequency of the sum of all deviations between X_1 and X_1+dX in an infinite, normally distributed population.

It might be well to point out also that allowing N to increase reduces the error introduced in obtaining equations (4) and (5).

---

8. Rietz, op. cit., p. 34.

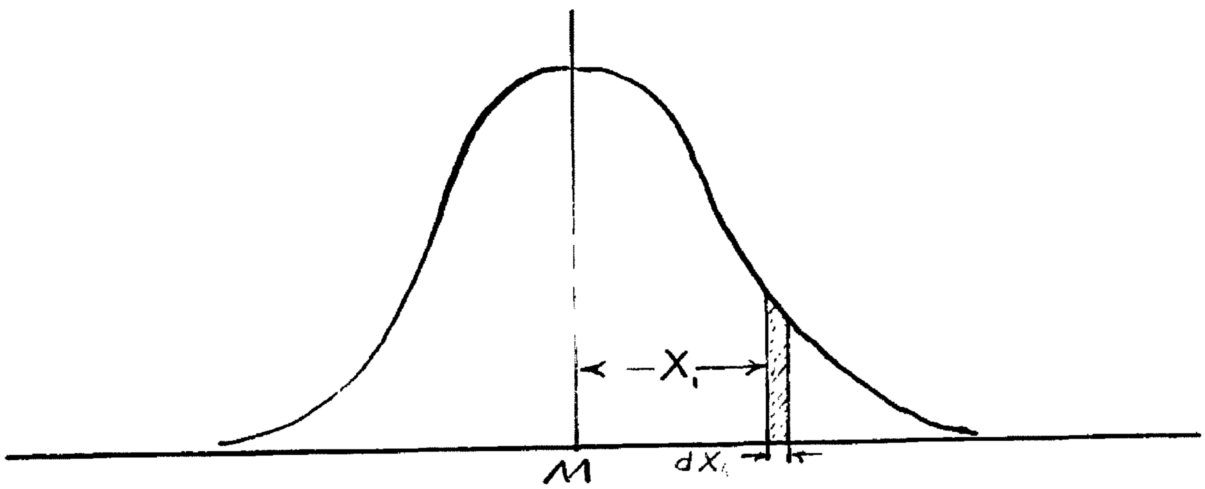

Figure 1.

iii. The Probability of n predetermined X's in a sample.

By using (6) it is easily seen that the total probability of getting a variable X_1 and another, X_2, and etc., ...X_n in a sample of n items is given by

$$df = (1/\sqrt{2\pi})^n \exp(-S[X^2/2])\, dX_1\, dX_2 \,....\, dX_n \quad ....(7)$$

the summation being over all items in the sample. This is true because each of these events is independent, and therefore the total probability is the product of their respective probabilities.

If m be the mean of the sample and v be an estimate of s from the sample, given by the formuli[9] as $m = [S(x)]/n$ and $v^2 = [S(x-m)^2]/(n-1)$, and if use is made of the facts that $S(x-M)^2 = S(x-m)^2 + n(m-M)^2$ and $X = (x-M)/s$, equation (7) may be written

9. That v as here given is the most efficient estimate of s will be shown and explained in section v.

$$df = (1/s\sqrt{2\pi})^n \exp(\frac{-n(m-M)^2}{2s^2})\exp(\frac{-S(x-m)^2}{2s^2})dx_1 \ldots dx_n$$

$$= (1/s\sqrt{2\pi})^n \exp(\frac{-n(m-M)^2}{2s^2})\exp(\frac{-(n-1)v^2}{2s^2})dx_1\ dx_2\ \ldots\ dx_n \ldots (8)$$

If $x_1+x_2+x_3+\ldots+x_n = mn$ be taken to represent the equation of a hyper-plane in n-dimensional space, the length of the radius vector drawn at right angles to the plane and intersecting it at a point Q will be[10] $OQ = mn/\sqrt{n}$. Let the distance from this point Q to a point P lying within the plane be $QP = \sqrt{S(x-m)^2} = v\sqrt{n-1}$. The plane on which P lies is of dimensions (n-1) and therefore P may take any position on the hyper-sphere whose surface is of (n-2) dimensions. Of the n parameters needed to describe the position of P we have two, namely m and v. The rest will necessarily be directional, and therefore may be taken as functions of the angles made by the radius vector with the axes. If they be $vf_1(\theta_1)$, $vf_2(\theta_2)$, ... $vf_{n-2}(\theta_{n-2})$, then the differential element of equation (8) becomes $f'_1(\theta_1)f'_2(\theta_2)\ldots f'_{n-2}(\theta_{n-2})$ $C'dmdv$. Since the functions of θ are independent of v and m when this expression is integrated over all the values of x, they give rise to a constant, and the value of the differential element becomes $C''v^{n-2}$ dmdv, and (8) becomes

$$df = C \exp(\frac{-(n-1)v^2}{2s^2}) \exp(\frac{-n(m-M)^2}{2s^2})\ v^{n-2}\ dm\ dv \ldots\ldots\ldots\ldots (9)$$

---

10. Love, Claud E. Elements of Analytic Geometry (New York 1935) pp. 39, 124.

This procedure may be thought of as an extension of the case of three dimensions. Here the plane may be represented as one with equal x, y, and z intercepts, its equation being $x+y+z = 3m$. Then $OQ = m/\sqrt{3}$, $QP = v\sqrt{2}$, and since to fix the position of P the quadrant in which the radius vector lies must be known, one angular function of the form $v\phi(\theta)$, where θ is independent of the values of m and v, is required. The differential element would then be $dx\ dy\ dz = v\phi(\theta)\ dm\ dv$, which completely describes the position of the point. A geometric interpretation of this may be had by examining the exaggerated figures 2 and 3. Here it may be seen that the differential is the volume of an infinitesimal cyllindrical shell, whose magnitude is dependent only on m and v, and not at all on the direction.

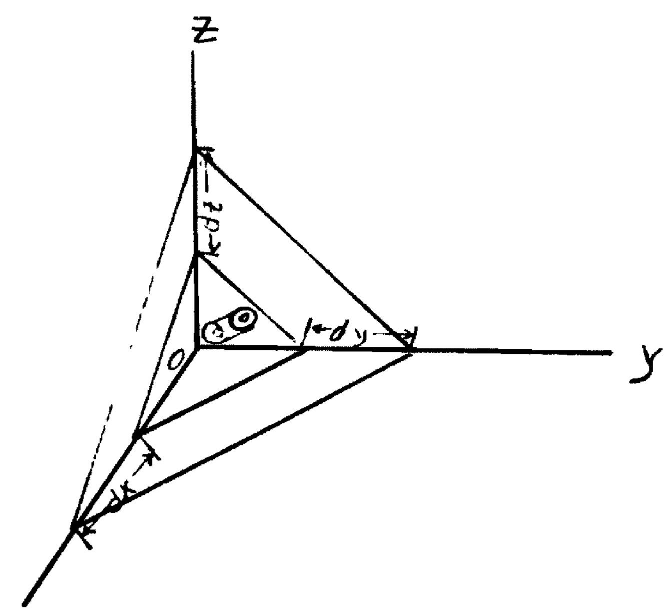

Fig. 2

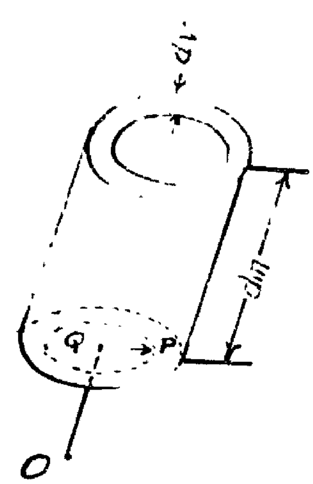

Fig. 3

The infinitesimal volume given by equation (9) represents, therefore, the relative frequency of the deviations which it describes.

iv. The Distribution of v .

Since m and v may be computed independently, m may be held constant and v allowed to vary, and in this way the frequency distribution of v may be found. If this is done in equation (9), it may be written

$$df = K(v/s)^{n-2}\exp[-\tfrac{n-1}{2}(v/s)^2]\ d(v/s)\ldots\ldots\ldots\ldots\ldots\ldots(10)$$

Integrating (10) over all possible cases gives the total possible probability, which is of course, one. By so doing, the value of K may be found. To make this integration, set

$$(v/s)^2 = 2y/(n-1) \qquad\qquad d(v/s) = dy/\sqrt{2(n-1)y}$$

$$(v/s)^{n-2} = 2^{(n-2)/2}\ (n-1)^{-(n-2)/2}\ y^{(n-2)/2}$$

and the integral becomes

$$K\ 2^{(n-3)/2}\ (n-1)^{-(n-1)/2}\int_0^{\infty} y^{(n-3)/2}\ e^{-y}\ dy$$

$$= K\ 2^{(n-3)/2}\ (n-1)^{-(n-1)/2}\ \Gamma[(n-1)/2] = 1\ldots\ldots\ldots\ldots(11)$$

$$\therefore K = \frac{(n-1)^{(n-1)/2}}{2^{(n-3)/2}}\ \frac{1}{\Gamma[(n-1)/2]}\ \ldots\ldots\ldots\ldots\ldots\ldots(12)$$

When v has n degrees of freedom,[11] its distribution is, by equations (10) and (12),

---

11. See section v.

$$df = \frac{r^{n/2}}{2^{(n-2)/2}} \frac{1}{\Gamma(n/2)} \exp[-\frac{n}{2}(\frac{v}{s})^2] (\frac{v}{s})^{n-1} d(\frac{v}{s}) \quad \ldots\ldots\ldots(13)$$

This method of approaching this problem is taken from a paper by J. O. Irwin.[12]

v. Degrees of Freedom

By (13), the frequency distribution of v is

$y = \frac{n^{n/2}}{2^{(n-2)/2}} \frac{1}{\Gamma(n/2)} \exp[-\frac{n}{2}(\frac{v}{s})^2] (\frac{v}{s})^{n-1} \frac{1}{s}$. The partial derivative of y with respect to s, when set equal to zero, will give the value of s for which y is a maximum, and therefore, the most probable value of v. The result is $s^2=v^2$ which proves that the value of v as used on page 6 gives the best estimate of s from a sample of n items.

The theory of degrees of freedom has been explained by Rider and Snedecor somewhat as follows. The number of degrees of freedom is defined as the number of independent variates. In determining variance this is one less than the total number of variates, for in any group of data, the mean having been calculated from all of the items, n-1 variates may be assumed at will, but the n^{th} will then be fixed in value by virtue of the fact that the deviates from the mean must all add up to be zero.[13] This should not be taken to indicate that the variance is entirely dependent upon the mean---

---

12. Irwin, J. O."Mathematical Theorems involved in the Analysis of Variance" Journal Royal Statistical Society, Vol. 94, pp. 284 ff.
13. Snedecor, Analysis of Variance and Covariance (Ames, Ia. 1934) p. 9.
    Rider, P. Modern Statistical Methods (New York, 1939) pp. 100, 133.

which is not true. For, as is clear from the development of equation (10), a number of samples may be had with the same mean, but different variances.

Degrees of freedom might be explained by saying that whereas the mean is dependent upon the values of the items, the estimate of variance is dependent upon the differences in the values of the adjacent items, there being n-1 such differences in a sample of n items.

vi. Distribution of w and z.

If there are two estimates of the same variance, v_1^2 and v_2^2, based respectively on n_1 and n_2 degrees of freedom, and w^2 is the ratio of the larger to the smaller, then $v_1^2 = w^2 v_2^2$. The distribution of v_1 will be, by equation (13),

$$df = \frac{n_1^{n_1/2}}{2^{(n_1-2)/2}\,\Gamma(n_1/2)} \exp(-n_1 v_1^2/2s^2)\,(v_1/s)^{n_1-1}\, d(\frac{v_1}{s}).$$

Or, since $v_1^2 = w^2 v_2^2$, equation (14),

$$df = \frac{n_1^{n_1/2}}{2^{(n_1-2)/2}\,\Gamma(n_1/2)} \exp(-n_1 w^2 v_2^2/2s^2)\,(wv_2)^{n_1-1}$$

$$s^{-n_1}\, v_2\, dw \quad \ldots\ldots\ldots\ldots(14)$$

gives the distribution of w for a given value of v_2^2. But the distribution of v_2 is, by equation (13),

$$df = \frac{n_2^{n_2/2}}{2^{(n_2-2)/2}\,\Gamma(n_2/2)} \exp(-n_2 v_2^2/2s^2)\,(v_2/s)^{n_2-1}$$

$$d(v_2/s) \;\ldots\ldots\ldots\ldots(15)$$

Therefore the complete distribution of w , as v_2 is allowed to vary over its whole range, will be given by the product of the second factors in equations (14) and (15), integrated over all possible values of v_2: that is, by the equation[14]

$$df = dw\int_0^\infty \frac{n_1^{n_1/2}\; n_2^{n_2/2}}{2^{[(n_1+n_2)/2]-2}\,\Gamma(n_1/2)\,\Gamma(n_2/2)}$$

$$\exp\,[(-n_1w^2+n_2)v_2^2/2s^2]\;\frac{v_2^{n_1+n_2-1}\; w^{n_1-1}}{s^{n_1+n_2}}\quad dv_2.$$

To integrate the right hand member of this equation the substitution $y = \dfrac{(n_1w^2+n_2)v_2^2}{2s^2}$ may be made. The equation reduces to

$$df = \frac{2\,n_1^{n_1/2}\; n_2^{n_2/2}\; w^{n_1-1}\quad dw}{(n_1w^2+n_2)^{(n_1+n_2)/2}\,\Gamma(n_1/2)\,\Gamma(n_2/2)}$$

$$\int_0^\infty y^{[(n_1+n_2)/2]-1}\quad e^{-y}\quad dy$$

---

14. To show this, let dy_w represent the probable frequency of **w** v_2 when w is variable and v_2 is constant. Then when v_2 is allowed to vary over all possible values, dy_v representing the probable frequency of v_2^2, the probable frequency of **w** will be $df = dy_w \int dy_v$ since with each value of dy_w may be associated any one value of dy_v and the new probability is the sum of all such possible, independent combinations.

Performing this last indicated integration yields

$$df = \frac{2n_1^{n_1/2} n_2^{n_2/2} \Gamma[(n_1+n_2)/2]}{\Gamma(n_1/2)\Gamma(n_2/2)} \frac{w^{n_1-1}}{(n_1w^2+n_2)^{(n_1+n_2)/2}} dw \quad \ldots\ldots\ldots\ldots\ldots\ldots(16)$$

This is the fundamental formula in the analysis of Variance. It says that if there are two samples from the same population, one having variance v_1^2 and n_1 degrees of freedom, and the other having variance v_2^2 and n_2 degrees of freedom, then the probability that the ratio of the larger variance to the smaller, $v_1/v_2 = w$ will lie between two values of w--say w_1 and w_2-- may be found by integrating the right hand side between the limits w_1 and w_2. In particular, the probability of getting a value of w greater than some number, w_0 say, may by found by integrating between w_0 and infinity.

Another form of this equation is had by making the substitution $z = \log_e w$, whence equation (16) becomes[15]

$$df = \frac{2n_1^{n_1/2} n_2^{n_2/2} \Gamma[(n_1+n_2)/2]}{\Gamma(n_1/2)\Gamma(n_2/2)} \frac{e^{n_1z}\, dz}{(n_1e^{2z}+n_2)^{(n_1+n_2)/2}} \quad \ldots(16a)$$

Tables have been compiled in terms of both w^2 and z which give the magnitude of this ratio that will ocurr 1 o/o

---

15. Development from Irwin, Op. cit. p. 237-8.

and 5 o/o of the time,[16] and it is by means of these tables that conclusions are arrived at concerning the significance of the difference between two estimates of the same variance. The tables are made by finding a value of w^2 [or z] such that the integral of equation (16) [or (16a)] between that value and infinity is 1/100 in the first case and 5/100 in the second. The test is made by comparing the computed value of w^2 or z with these two numbers from the table and deciding whether the computed value is significantly large. The final decision will depend upon the type of material being sampled. For example a difference ocurring between 1 o/o and 5 o/o of the time would not be nearly so significant in sampling the weight of logs for a rough check as in sampling the weight of diamonds supposed to be all of the same value.

Other more important uses of the method will be shown in later sections.

---

16. Fisher, R. A. Statistical Methods for Research Workers, (London 1934) Tables IV and VI.

APPLICATIONS

vii. Analysis of Variance within and among classes.

Suppose that the sample being considered divides itself into k classes with u individuals in each class. Let the measurement of the i^{th} individual in the j^{th} class be x_{ij}, the mean of the j^{th} class be $\bar{x}_j$, and m the mean of the sample.

The total sum of squares of deviations may be broken up as follows: (S_j being used to mean summation with respect to j)

$$S_iS_j(x_{ij} - m)^2 = S_iS_j[(x_{ij}-\bar{x}_j) + (\bar{x}_j-m)]^2$$

$$= S_iS_j(x_{ij}-\bar{x}_j)^2 + 2S_iS_j(x_{ij}-\bar{x}_j)(\bar{x}_j-m) + S_iS_j(\bar{x}_j-m)^2 \quad \ldots(17)$$

But $S_iS_j(x_{ij}-\bar{x}_j)(\bar{x}_j-m) = S_j[(\bar{x}_j-m)S_i(x_{ij}-\bar{x}_j)]$, and since $S_i(x_{ij}-\bar{x}_j)$ is zero, the middle term of (17) becomes zero, and (17) becomes

$$S_iS_j(x_{ij}-m)^2 = S_iS_j(x_{ij}-\bar{x}_j)^2 + S_iS_j(\bar{x}_j-m)^2 \quad \ldots\ldots\ldots\ldots(18)$$

Consider only the j^{th} class. As shown in Sec. v, the best estimate of the variance of the population will be $S_i(x_{ij}-\bar{x}_j)^2/(u-1)$, and the 1/k th part of the sum of these for all classes will be the mean of them throughout the sample. That is, an estimate of s^2 based on the sums of squares of deviations within classes only is

$$v^2 = S_iS_j(x_{ij}-\bar{x}_j)^2/k(u-1) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(19)$$

The other term in (18), $S_iS_j(\bar{x}_j-m)^2$ is plainly $uS_j(\bar{x}_j-m)^2$. The variance of the population is most closely approximated by the variance of the means when this variance is multiplied by the number of means,[17] and, as before, the variance of the means is most efficiently approximated when the sum of squares of deviations in the sample is divided by the number of degrees of freedom. Therefore, the efficient estimate of the population variance based on between-class deviations is[18]

$$v^2 = uS_j(\bar{x}_j-m)^2/(k-1) \qquad (20)$$

That these two estimates of s^2 are independent is seen in equation (18) where the sums of squares of deviations upon which they are based are shown to be respective components of the total sum of squares of deviations.

The purpose of this analysis is to determine whether the classes are sufficiently distinct to justify grouping the data in such manner. For example, suppose that the weights of immigrants were being tabulated by age and by country of origin. Such a test might then be made to find whether there is a significant difference in the weights of different nationalities. The test is made as was outlined in Sec. vi, using as the two estimates of variance the second terms of equations (19) and (20).

---

17. Jones, D. C. A First Course in Statistics (London 1924) p. 154.
18. Derivation from Rider, op. cit., p. 132.

viii. Analysis of Variance in Samples with two or more variables.

Suppose that a sample is being considered which seems to divide itself naturally into rows and columns, according to two criteria. Let k be the number of columns and u the number of rows, and let the measure of the individual in the i^{th} row and j^{th} column be x_{ij}, let all of the items in the sample--N(=ku) in number--have a mean m, where the population mean is M, and let the mean of the j^{th} column be $\bar{x}_j$ and the mean of the i^{th} row be $\bar{x}_i$.

The total sum of squares of deviations may be broken up as follows:

$$S_{ij}(x_{ij}-m)^2 = S_{ij}[(x_{ij}-\bar{x}_i-\bar{x}_j+m) + (\bar{x}_i-m) + (\bar{x}_j-m)]^2$$

$$= S_iS_j(\bar{x}_i-m)^2 + S_iS_j(\bar{x}_j-m)^2 + S_iS_j(x_{ij}-\bar{x}_i-\bar{x}_j+m)^2$$

$$= kS_i(\bar{x}_i-m)^2 + uS_j(\bar{x}_j-m)^2 + S_{ij}(x_{ij}-\bar{x}_i-\bar{x}_j+m)^2 \ldots\ldots\ldots\ldots(21)$$

It was shown in Sec. vii. that the first term of (21) gives, when divided by u-1, an efficient estimate of s^2 based on between-class relations only. Likewise, the second term gives an efficient estimate of s^2 based on inter-class relations only when divided by k-1.

The third term may be written

$$S_{ij}[(x_{ij}-M) - (\bar{x}_i-M) - (\bar{x}_j-M) + (m-M)]^2$$

$$= S_{ij}(x_{ij}-M)^2 - 2S_{ij}(x_{ij}-M)(\bar{x}_i-M) + S_{ij}(\bar{x}_i-M)^2 + \ldots.$$

Define E(x) as the expected value of x, considered over all possible values: that is, the absolute mean value. Then $E(x_{ij}-M)^2 = (1/N)S_{ij}(x_{ij}-M)^2 = s^2$, and so on for the other squared terms. $(x_{ij}-M)$ may be replaced by $(\bar{x}_i-M)$ when $E(x_{ij}-M)(\bar{x}_i-M)$ is being determined, because $\bar{x}_i$ is an estimate of x_{ij} determined as the mean of a number of x_{ij}'s and will, when summed over the population, give the same result.[19] This expected product then becomes $E(\bar{x}_i-M)^2 = s^2/u$. Continuing in this way, the following results are obtained.[20]

$E(x_{ij}-M)^2$	$= s^2$	$E(x_{ij}-M)(\bar{x}_j-M)$	$= s^2/k$
$E(\bar{x}_i-M)^2$	$= s^2/u$	$E(x_{ij}-M)(m-M)$	$= s^2/ku$
$E(\bar{x}_j-M)^2$	$= s^2/k$	$E(\bar{x}_i-M)(m-M)$	$= s^2/ku$
$E(m-M)^2$	$= s^2/ku$	$E(\bar{x}_j-M)(m-M)$	$= s^2/ku$
$E(x_{ij}-M)(\bar{x}_i-M)$	$= s^2/u$	$E(\bar{x}_i-M)(\bar{x}_j-M)$	$= s^2/ku$

Let R be the third term of (21), then adding the above values with proper signs and coefficients gives

$$E(R) = s^2[1 - 1/u - 1/k + 1/ku] = s^2(u-1)(k-1)/ku.$$

Therefore,

$$\frac{E(R)}{(u-1)(k-1)} = \frac{s^2}{ku}. \quad \text{And therefore} \quad \frac{E[S(R)]}{(u-1)(k-1)} = s^2.$$

That is, the value of R is an efficient estimate to s^2 when divided by $(u-1)(k-1)$. R is plainly an inter-action term and may be used to measure experimental error since it is

19. To prove this, let $x_{ij} = \bar{x}_i+d$. Then $E(x_{ij}-M)(\bar{x}_i-M) = E(\bar{x}_i-M)^2 + Ed(\bar{x}_i-M) = s^2/u + 0$.
20. This method may also be used to prove the other estimates of s^2.

made up of what is left after subtracting both the sum of squares within rows and that within columns from the total sum of squares.[21] Being thus independent of the other sums of squares, R may be used as the basis for an estimate of s^2, which estimate may be compared with those based on sums of squares within rows and within columns to determine the significance of classification into rows and columns in the same manner as was used in Sec. vii. It is evident that the estimates based on sums of squares within rows and within columns may be compared in the same way to answer questions pertinent to a particular sample.

Should a sample be found which may be classified on the basis of more than two variables, this same method is applicable. For example, suppose that there were k columns, u rows, and a sub-class consisting of c items at each intersection--the position of the item in the sub-class depending on a third variable. If m is the mean of the whole sample, $\bar{x}_i$ the mean of the i^{th} row, $\bar{x}_j$ the mean of the j^{th} column, $\bar{x}_b$ the mean of the items in the b^{th} sub-class, and the measure of the individual in the i^{th} row, j^{th} sub-class, and b^{th} position in the sub-classes, x_{ijb}, then as before, (S indicating summation over all items)

21. Rider, op. cit., p. 138.

$$S(x_{ijb}-m)^2 = S(\bar{x}_i-m)^2 + S(\bar{x}_j-m)^2 + S(\bar{x}_b-m)^2$$
$$+ S(\bar{x}_{ij}-\bar{x}_i-\bar{x}_j+m)^2 + S(\bar{x}_{ib}-\bar{x}_i-\bar{x}_b+m)^2 + S(\bar{x}_{jb}-\bar{x}_j-\bar{x}_b+m)^2$$
$$+ S(\bar{x}_{ijb}-\bar{x}_{ij}-\bar{x}_{jb}-\bar{x}_{ib}+\bar{x}_i+\bar{x}_j+\bar{x}_b-m)^2 \ldots\ldots\ldots\ldots\ldots\ldots (22)$$

The number of degrees of freedom to be associated with each of these terms in order to have an efficient estimate of s^2 are, respectively,

u-1	(u-1)(k-1)
k-1	(u-1)(c-1)
c-1	(k-1)(c-1)

(u-1)(k-1)(c-1)

these being determined as before.[22] Also, as before, comparisons of any two of these estimates may be made in order to answer a pertinent question about differences in the variances of these three groups, the sample, or the population. Comparing any estimate based on a group with that based on the last or inter-action term gives a test for the significance of the classification into that group.

22. Development from Irwin, op, cit., p. 289.

AN HISTORICAL NOTE ON ANALYSIS OF VARIANCE.

The real author of the normal curve (equation 6) seems to have been questioned in mathematical circles until Mr. K. Pearson found that our modern method of handling that function was first given to the world by De Moivre in 1733. The work of La Place came about fifty years later and that of Gauss some thirty years after La Place.[23] Bernoulli did some work on this problem in 1713 and Stirling in 1730. After De Moivre came Euler in 1738 and Maclaurin in 1742, but these seemed to Mr. Pearson to be of less importance than the three men first cited.

The distribution given by equation (13) was first attempted by Helmert in 1875-6, when he said:

"Given a normal parent population of x's with mean 0 and variance σ^2 from which are drawn at random each of N independent values, x_1, x_2, ... x_N, measured from the population mean as the origin, giving as the sample mean $\bar{x} = (x_1 + x_2 + x_3 + \ldots + x_N)/N$ and as the second moment of the sample from the population mean, $s^2 = \overline{\mu'} = (x_1^2 + x_2^2 + \ldots + x_N^2)/N$. Then the probability that the sum of squares of deviations, $U = x_1^2 + x_2^2 + \ldots + x_N^2$ will fall into the interval U to U+dU is

---

23. Pearson, K. "Historical note on the origin of the normal curve" Biometrika, Vol. 16 (1924) p. 402. Quoted in Rietz, H. L. Mathematical Statistics (Chicago 1927) p. 47.

given by $\frac{1}{2^{N/2}\ \sigma^N \Gamma(N/2)} U^{(N-2)/2} e^{-U/2\sigma^2} dU.$

. . . .

"...so that the frequency function of the sample variance $s^2 = \mu$ given by (3) is equal to[24]

$$(N/2)^{(N-1)/2} \frac{1}{\sigma^{(N-1)}\ \Gamma[(N-1)/2]} \mu^{(N-3)/2} e^{-N\mu/2\sigma^2}."$$

The next contribution was made by K. Pearson in 1900 when he published his chi² distribution.[25]

R. A. Risher says of these two distributions:

> "...Helmert's solution in 1875 of the distribution of the sum of the squares of deviations from a mean is in reality equivalent to the distribution of chi² given by K. Pearson in 1900. It was again discovered independently by Student in 1908, for the distribution of the variance of a normal sample. The same distribution was found by the author for the index of dispersion derived from small samples from a Poisson series.
>
> "What is even more remarkable is that although Pearson's paper of 1900 contained a serious error which vitiated most of the tests of goodness of fit made by this method until 1921, yet the correction of this error, when efficient methods of estimation are used,

---

24. Helmert, "Ueber die Wahrscheinlichkeit der Potenzsummen der Boebachtungsfehler und über einige damit im Zusammsenhange stehende Fragen" Zeitschrift für Mathematik und Physik Vol. 21, 1876, pp. 192-218. Quoted in Rietz, H. L. "Some topics in sampling Theory" Bulletin American Mathematics Society, Vol. 43, 1937 pp. 209-230.
25. Pearson, K. "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 V. 1, pp. 157-175.

leaves the form of the distribution unchanged, and only requires that some few units should be deducted from one of the variables with which the table of chi^2 is entered."[26]

What Mr. Fisher has reference to here when he speaks of the error is the fact that both of these methods divide the sum of squares of deviations by the number in the sample rather than by the number of degrees of freedom.

Student's distribution as given in 1908 is the distribution of s^2 (equation 13) in a form only very slightly different from that given here. It is, where s^2 is an estimate of the population variance,[27]

$$df = \frac{1}{\Gamma[(n-2)/2]} (n/2\sigma^2)^{n/2} (s^2)^{(n-2)/2} e^{-ns^2/2\sigma^2} d(s^2).$$

Fisher says again of this distribution, that it was intuitive. Fisher himself derives it by geometry of n-space (not that used in this paper), but his method is hard to follow and seems almost as intuitive as that of Student.[28]

It was in 1921 that Fisher took an active interest in this test and the name "Analysis of Variance" as changed from "Analysis of Variation" is due to him. He was working at the Rothamsted Experimental Station at the time and his method of separating the sum of squares of deviations

26. Fisher, R. A. Statistical Methods for Research Workers (London 1934) p. 17.
27. Student, "The Probable error of a mean" Biometrika Vol. 6, (1908-9) pp. 1-25.
28. Fisher, R. A. "Applications of Student's Distribution" Metron Vol. V No. 3 1925, p. 92.

within a sample as well as the z-distribution (equation 16a) and the tests pertinent thereto began to appear with increasing frequency in his reports of agricultural experiments, but it seems that he never bothered to write a formal paper on his methods or what he considered might be their generalizations. He did, however, show in 1924 that certain other distributions, notably that of Student, could be easily transformed into his z distribution,[29] and he outlined the method of procedure in his book "Statistical Methods for Research Workers," 1925 edition.

J. O. Irwin gave the first formal discussion of the general theory in his paper of 1931.[30] Another formal treatise was given by Wilks in 1932,[31] but essentially it has remained where Fisher left it—with the agriculturalist.

From the point of view of application, Snedecor has probably made the greatest use of Analysis of Variance and even outlined the various ways to get results and approximate results by using the distributions. His applications are to agriculture, and he gives no mathematical reason for their existence.[32]

---

29. Fisher, R. A. "On a distribution yielding the error function of several well-known statistics" Proceedings of the International Mathematical Congress Toronto 1924.
30. Irwin, J. O. op. cit.
31. Wilks, S. S. "Certain Generalizations in the Analysis of Variance" Biometrika Vol. 24, pp. 471-494.
32. Snedecor, G. W. op. cit.

CONCLUDING STATEMENT

Although 199 years have elapsed since the beginning of this method, the mathematical work upon it has been increasing in extent and scope and it seems entirely possible that it will soon be extended to apply to all frequency distributions and will be used in all branches of applied statistics.

An examination of a few of this year's texts on elementary statistics has disclosed the fact that in each of them there is a section outlining the method of dividing the total sum of squares within a sample and entering a table with the number of degrees of freedom applying to them. There would seem to be but little doubt that Analysis of Variance will soon become an important part of every college course in statistics.

BIBLIOGRAPHY

Coolidge, J. L. An Introduction to Mathematical Probability Oxford, 1925.

Fisher, R. A. "Applications of Student's Distribution" in Metron Vol. V, No. 3, 1925, pp. 90-93.

Fisher, R. A. "On a Distribution yielding the error function of several well-known statistics" in Proceedings of the International Mathematical Congress, Toronto, 1924.

Fisher, R. A. Statistical Methods for Research Workers London, second edition 1925, third edition 1934.

Hall and Knight Higher Algebra, London, 1936. Third edition.

Irwin, J. O. "Mathematical theorems involved in the Analysis of Variance" in Journal of Royal Statistical Society Vol. 94, 1931, pp. 284-300.

Jones, D. Caradog A First Course in Statistics London 1924

Love, Claud E. Elements of Analytic Geometry New York, 1935.

Rider, Paul R. An Introduction to Modern Statistical Methods New York, 1939.

Rietz, Henry Lewis Mathematical Statistics Chicago, 1927.

Rietz, H. L. "Some topics in sampling theory" in Bulletin American Mathematical Society Vol. 43, pp. 209-230.

Snedecor, George W. Calculation and Interpretation of Analysis of Variance and Covariance, Ames Iowa, 1934.

Wilks, S. S. "Certain Generalizations in the Analysis of Variance" in Biometrika, Vol. 24, 1932, pp. 471-494.