Orthogonal Functions and Tests of Significance in the Analysis of Variance

By F. Yates

In many types of statistical analysis based on the method of least squares, it is necessary to test the significance of one group of effects while admitting the possible existence of other groups of effects. Thus, for example, the effect of the amount and distribution of sunshine on the yield of a crop may be under consideration, it being known that the amount and distribution of the rainfall have some effect. In such a case the simultaneous partial regression of yield on variates representing rainfall and its distribution and sunshine and its distribution must be obtained. Each regression coefficient can be tested separately by obtaining its estimated standard error in the ordinary manner, and then using the t test, but to make a comprehensive test of all the sunshine effects, it is necessary to have a combined test of the whole group of sunshine regression coefficients, making due allowance for rainfall effects.

In principle the procedure is very simple. The difference of the reduction in the sum of squares due to the regression on both sunshine and rainfall variates, and that due to the regression on the rainfall variates only, is taken. This gives the appropriate sum of squares for making the test, the ordinary analysis of variance z test being employed, with degrees of freedom corresponding to the number of independent sunshine variates.

To those familiar with the structure of the analysis of variance and the properties of orthogonal degrees of freedom, the justification of this procedure is almost obvious, but as it is of very wide applicability, and as I have had several requests for an explicit proof, it may be worth while to give a simple demonstration here. The opportunity will be taken of setting out the general properties of orthogonal functions.

If a variate y is a linear function of several independent variates $x_1, x_2, \ldots x_n$, then the efficient estimates of the regression coefficients are given by the method of least squares (*i.e.*, the ordinary method of partial regression). Instead of working with the variates $x_1, x_2, \ldots x_n$ directly, we may use linear functions of them of the type

$$\xi_1 = \lambda_1 x_1, \; \xi_2 = \mu_1 x_1 + \mu_2 x_2, \; \xi_3 = \nu_1 x_1 + \nu_2 x_2 + \nu_3 x_3, \; \ldots .$$

It can easily be shown that the partial regression equation on the ξ's, when transformed to an equation on the x's, is identical with the partial regression equation obtained by working with the x's directly.

The relations between the regression coefficients of the x's and those of the ξ's can be immediately derived from the identity

$$b_1x_1 + b_2x_2 \ldots . + b_nx_n \equiv B_1\xi_1 + B_2\xi_2 + \ldots + B_n\xi_n$$

which gives :

$$\begin{aligned} b_1 &= B_1\lambda_1 + B_2\mu_1 + B_3\nu_1 + \ldots, \\ b_2 &= B_2\mu_2 + B_3\nu_2 + \ldots, \\ &\ldots \\ b_n &= \rho_n B_n \end{aligned}$$

Note that b_n only involves B_n, b_{n-1} only involves B_n and B_{n-1}, etc.

If λ_1, μ_1, μ_2, etc., are chosen so that

$$S(\xi_1\xi_2) = 0, \quad S(\xi_1\xi_3) = 0, \quad S(\xi_2\xi_3) = 0, \text{ etc.},$$

as can clearly always be done, the ξ's are said to be *orthogonal.*

Orthogonal variates possess several important properties. Owing to the vanishing of the product sums, the normal equations each contain only one regression coefficient, being in fact :

$$B_1S(\xi_1^2) = S(\xi_1 y), \quad B_2S(\xi_2^2) = S(\xi_2 y), \quad \text{etc.}$$

The B's are therefore immediately expressible as linear functions of the y's, *and these linear functions are also mutually orthogonal.* Consequently if the y's are normally and independently distributed with unit variance about values given by a " true " regression equation, the covariances between the B's are zero, and the variances of the B's are given by

$$V(B_1) = \frac{S(\xi_1^2)}{\{S(\xi_1^2)\}^2} = \frac{1}{S(\xi_1^2)}, \text{ etc.},$$

The reduction in the total sum of squares is

$$\begin{aligned} B_1S(\xi_1y) + B_2S(\xi_2y) + \ldots \ldots + B_nS(\xi_ny) \\ = B_1^2/S(\xi_1^2) + B_2^2/S(\xi_2^2) + \ldots + B_n^2/S(\xi_n^2). \end{aligned}$$

Each component, $B_1^2/S(\xi_1^2)$, etc., is distributed independently of all the others, and any component for which the true value of the regression coefficient is zero, is distributed as is the square of a normal deviate with unit standard deviation. If the true values of all the coefficients are zero, the whole sum of squares is distributed as χ^2 with n degrees of freedom, by the definition of χ^2.

If the regression on the first k variates only is taken, this will be equivalent to a regression on the first k ξ's, since only the last $(n - k)$ ξ's contain x_{k+1} to x_n. The omission of the last k ξ's will not affect the values of the first k B's, and consequently the reduction in the sum of squares due to a regression on the first k variates only is

$$B_1^2/S(\xi_1^2) + B_2^2/S(\xi_2^2) + \ldots + B_k^2/S(\xi_k^2).$$

The difference of these two quantities is

$$B_{k+1}^2/S(\xi_{k+1}^2) + \ldots + B_n^2/S(\xi_n^2),$$

and if none of the coefficients B_{k+1} to B_n has a true value other than zero, it follows that this last sum of squares must be distributed as χ^2 with $n - k$ degrees of freedom.

We have already shown that $b_{k+1}, \ldots . b_n$ are linear functions of $B_{k+1}, \ldots . B_n$ only, so that $B_{k+1}, \ldots . B_n$ depend on $b_{k+1}, \ldots . b_n$ only, and if the true values of all these latter are zero, then so are the true values of $B_{k+1}, \ldots . B_n$. Consequently the test of significance is a valid one for deviations of the group of coefficients $b_{k+1}, \ldots b_n$ from zero.

Fisher, in *The Design of Experiments*, Section 61, has pointed out that there are a multiplicity of tests of significance based on the same null hypothesis. In the absence of any more precise information as to the particular type of deviation that may be expected, the above test, which will clearly detect deviations of any type provided they are sufficiently large, is fully appropriate. Of course if deviations of a particular type, such as that $b_{k+1}, \ldots . b_n$ all differ from zero by about the same amount, require more close investigation, more specific tests designed to detect the particular type of deviation under consideration must be employed. In particular any linear function of the b's can be tested by means of the t test. (See Fisher, *Statistical Methods for Research Workers*, Section 29.)

In the above proof the nomenclature of partial regression has been used. It should be noted, however, that the proof is equally applicable to any least square analysis involving linear functions of the independent variates. The procedure to be followed in the analysis of variance of a replicated experiment in randomized blocks, for example, can be deduced immediately by assigning one group of x's to the blocks, and another to the treatments, such that if x_1 is associated with block 1 it has the value 1 for every yield of block 1 and zero for every yield of the other blocks. The regression coefficients are thus replaced by block and treatment constants. In this case the x's for the blocks will be found to be orthogonal with the x's for the treatments, which accounts for the particularly simple form that the analysis of variance assumes. (There are certain complications arising from the introduction of redundant constants, which, however, are easily overcome. For examples of this procedure see Yates (2), (3), and Irwin (1).) When there is a set of concomitant observations, say z, however, and a regression on these is introduced (as in the customary analysis of covariance) it will be found that z is not orthogonal to the block or treatment x's, so that the test of significance has to be based on the difference between the sum of squares attributable to blocks, treatments and the regression on z, and that attributable to blocks and an (altered) regression on z only. (See Fisher, *Statistical Methods*, Section 49.1.)

References.

(1) Irwin, J. O. (1934). Independence of the Constituent Items in the Analysis of Variance. *Suppl. J. Roy. Stat. Soc.*, I, 236–251.
(2) Yates, F. (1933). The Principles of Orthogonality and Confounding in Replicated Experiments. *J. Agric. Sci.*, XXIII, 108–145.
(3) Yates, F. (1936). Incomplete Latin Squares. *J. Agric. Sci.*, XXVI, 301–315.